THERMODYNAMIC, KINETIC, AND STRUCTURAL BASIS FOR THE RELAXED DNA SEQUENCE SPECIFICITY OF "PROMISCUOUS" MUTANT *Eco*RI ENDONUCLEASES

by

Paul Joseph Sapienza

BS, James Madison University, 1998

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
2005

UNIVERSITY OF PITTSBURGH

FACULTY OF ARTS AND SCIENCES

This dissertation was presented

by

Paul Joseph Sapienza

It was defended on

January 28, 2005

and approved by

Dr. Jeffrey G. Lawrence

Dr. Craig L. Peebles

Dr. John M. Rosenberg

Dr. Gordon S. Rule

Dr. Linda Jen-Jacobson
Dissertation Director

Thermodynamic, Kinetic, and Structural Basis for the Relaxed DNA Sequence Specificity of "Promiscuous" Mutant *Eco*RI Endonucleases

Paul Joseph Sapienza, Ph.D.

University of Pittsburgh, 2005

Promiscuous mutant *Eco*RI endonucleases produce lethal to sub-lethal effects because they cleave *E. coli* DNA despite the presence of the *Eco*RI methylase. Three promiscuous mutant forms, A138T, E192K and H114Y, have been characterized with respect to their binding affinities and first-order cleavage rate constants towards the three classes of DNA sites: specific, miscognate (*Eco*RI*) and nonspecific. We have made the unanticipated and counterintuitive observations that the mutant endonucleases that exhibit relaxed specificity *in vivo* nevertheless bind more tightly than the wild-type enzyme to the specific recognition sequence *in vitro* and show even greater preference for binding to the cognate GAATTC site over miscognate sites. Binding preference for *Eco*RI* over nonspecific DNA is also improved. The mutant enzymes cleave the cognate site GAATTC at a normal rate, but cleave *Eco*RI* sites faster than does the wild-type enzyme. Thus, the mutant enzymes use two mechanisms to partially bypass the multiple fail-safe mechanisms that protect against cleavage of genomic DNA in cells carrying the wild-type *Eco*RI restriction-modification system: (a) Binding to *Eco*RI* sites is more probable than for wild-type enzyme because nonspecific DNA is less effective as a competitive inhibitor; (b) The combination of increased affinity and faster cleavage at *Eco*RI* sites makes double-strand cleavage of these sites a more probable outcome than it is for the wild-type enzyme.

The aggregate of biochemical data presented in this thesis leads us to propose a model in which the A138T miscognate DNA binding ensembles are more highly populated by complexes that are on the path to the transition state than wild-type miscognate DNA binding ensembles. The mutant accomplishes this population shift by forming 'specific-like' phosphate contacts to miscognate sites, which in turn stabilize the DNA distortion that is critical for efficient cleavage. Given that amino acid 138 lies within the enzyme's "arm", which contains amino acids that contact the DNA phosphate backbone in the specific complex, we hypothesize that the A138T mutation has an effect on the structure and/or dynamics of the "arm" protein segment, resulting in formation of functional phosphate contacts to a broader range of DNA substrates.

Acknowledgements

I thank my thesis advisor, Dr. Linda Jen-Jacobson, for her wisdom, patience, and friendship during my studies in her lab. I have benefited greatly from working with such an outstanding scientist, teacher, and person.

This thesis contains data gathered by a number of very talented undergraduates with whom I have had the privilidge of working. William H. McCoy IV was responsible for cloning and purifying the E192K enzyme, and characterizing the specific DNA binding properties of the E192K mutant. "Wild-Bill" also came up with the modification to the Deng and Nickoloff site-directed mutagenesis procedure that was instrumental in our constructing the E192K and H114Y mutations. Crystal Ann dela Torre cloned and purified the H114Y mutant, and performed the biochemistry on this enzyme. Samyukta Jana did the work characterizing the miscognate DNA interaction of the E192K enzyme.

Many ideas put forth in this work stem from data collected by and discussions with previous and current members of the Jen-Jacobson lab. I thank Michael Kurpiewski, Dr. Lisa Engler, Dr. Arabela Grigorescu and Dr. Deng Feng Cao for their legacy and their willingness to share their vast knowledge with me.

I solved the crystal structure in collaboration with Dr. John Rosenberg. I am thankful for his consultations during every step of the structure solution process. Although x-ray crystallography now represents only a fraction of John's research focus, to spend time with John at the synchrotron is to witness the passion and knowledge that he has for the field. In addition, I would like to thank Dr. Arabela Grigorescu for her help during structure refinement, Regina Kettering for her aid in data processing, and Ingrid Kuo for purifying oligos used in crystallization.

I thank my friends within and outside the Department of Biological Sciences with whom I have shared laughs, music, beer, and frisbee.

Finally, I dedicate this work to my mother, father, wife Erin, and the rest of my family whose love and support helped me achieve this goal and will be a huge part of everything I do.

TABLE OF CONTENTS

1. CHAPTER 1: INTRODUCTION 1
1.1. Levels of sequence selectivity in proteins that interact with DNA 1
1.2. Restriction endonucleases: Model System for understanding stringent discrimination of DNA binding proteins 3
1.3. Structural and energetic determinants of specificity in the *Eco*RI restriction endonuclease 3
1.3.1. Three types of DNA binding sites for the *Eco*RI restriction endonuclease 4
1.3.2. Structural view of specificity of the EcoRI restriction endonuclease 5
1.3.2.1. "Direct Readout" of the GAATTC site by the EcoRI endonuclease 5
1.3.2.2. The *Eco*RI-DNA "kink" and "indirect readout" 8
1.3.2.3. Coupling of DNA recognition to catalysis 11
1.3.2.4. An enzyme disorder to order transition is coupled to specific DNA binding 13
1.3.2.5. Structures of miscognate and non-specific complexes are required to complete the structural view of specificity 15
1.3.3. Thermodynamic and kinetic views of specificity of the *Eco*RI restriction endonuclease 18
1.3.3.1. Energetics of discrimination in binding and catalysis 18
1.3.3.2. Discrimination in vivo 24
1.3.3.3. Three different binding modes pre-transition state: adaptive and non-specific complexes 25
1.4. Attempts at changing the specificity of type II restriction endonucleases 28
1.5. Promiscuous mutants of the EcoRI endonuclease: tools to further our understanding of molecular specificity 29
2. CHAPTER 2: PROMISCUOUS *Eco*RI ENDONUCLEASE MUTANTS AND THE SPECIFIC DNA INTERACTION 32
2.1. Introduction 32
2.1.1. Specific complex probed in solution studies is a 'Pre-Transition-State' complex 32
2.1.2. Relating binding affinities to free energy changes 34
2.1.3. Relating kinetic stability of EcoRI-specific DNA complexes to their thermodynamic stabilities 35
2.1.4. Relating DNA cleavage rate constants to activation free energies 36
2.2. Results and Discussion 36
2.2.1. Design of oligonucleotide substrates 36
2.2.2. Promiscuous mutations improve binding to specific DNA 37
2.2.3. Improved mutant specific DNA binding free energy is utilized efficiently in the transition state 40
2.2.4. Thermodynamic dissection of ΔG°_{bind} of *Eco*RI specific binding by van't Hoff analysis 44

2.2.5. Structural and energetic factors that may contribute to the improved specific binding by the promiscuous enzymes 51
2.2.5.1. The A138T Mutant 51
2.2.5.2. The H114Y Mutant 52
2.2.5.3. The E192K mutant 61
3. CHAPTER 3: THE CRYSTAL STRUCTURE OF THE A138T *Eco*RI-d(TCGCGAATTCGCG) COMPLEX AT 1.95 Å RESOLUTION 65
3.1. Introduction 65
3.2. Results and Discussion 66
3.2.1. DNA numbering scheme 66
3.2.2. General features of the A138T *Eco*RI-specific DNA complex 66
3.2.3. Local effects of the A138T mutation 71
3.2.4. Structure of the A138T *Eco*RI-d(TCGCGAATTCGCG) complex provides insight into the changes in GAATTC flanking context preference by the A138T enzyme 80
3.2.5. Distal Effects of the A138T Mutation: Phosphate Contacts, the Active Site and the 'Arms' 90
3.2.6. Distal effects of the A138T mutation: pH dependence of specific binding and potential crosstalk with histidine 114 99
3.2.7. Hypotheses on the origins of the differences in ΔH° and ΔS° accompanying specific DNA binding by the A138T and wild-type enzymes 102
4. CHAPTER 4: PROMISCUOUS MUTANTS AND THE NON-COGNATE DNA INTERACTION: INSIGHTS INTO THE BASIS FOR RELAXED *IN VIVO* SPECIFICITY 110
4.1. Introduction 110
4.2. Results and Discussion 114
4.2.1. Mutant proteins maintain binding discrimination against miscognate sites 114
4.2.2. Mutant enzymes exhibit increased discrimination between non-specific and miscognate sites 117
4.2.3. Mutant proteins stabilize the transition state in miscognate complexes 121
4.2.4. Flanking sequence modulates miscognate DNA cleavage 123
4.2.5. Estimates of in vivo error rates correlate with phenotypes of mutants 128
4.2.6. Barriers and pathways in the evolution of sequence specificity 131
4.3. Conclusions 135
5. CHAPTER 5: TOWARDS A MOLECULAR UNDERSTANDING OF RELAXED SPECIFICITY EXHIBITED BY THE PROMISCUOUS MUTANTS 138
5.1. Introduction 138
5.2. Active sites of A138T-miscognate DNA complexes are more assembled than wild-type counterparts 141
5.3. Effect of flanking sequence on miscognate binding and cleavage reveals 'specific-like' properties of some A138T-miscognate complexes 151
5.4. Ethylation interference footprints reveal structural differences between wild-type and A138T-miscognate DNA complexes 156
5.5. Probes to address differences in conformation and/or dynamics of *Eco*RI arms in different complexes 172
6. CHAPTER SIX: METHODS 176
6.1. Site directed mutagenesis 176

6.2. Three “extra” amino acids at the N-terminus of *Eco*RI endonucleases used in this study ……..177
6.3. Enzyme expression purification and characterization 178
6.4. Purification of oligonucleotide substrates..... 179
6.5. DNA Concentration Determination and Duplex Formation..... 179
6.6. 5’-End Labeling of DNA 180
6.7. Determination of Equilibrium Binding Constants 180
6.7.1. Direct Equilibrium Binding 180
6.7.2. Equilibrium Competition 182
6.8. DNA Cleavage Kinetics..... 183
6.9. Kinetics of dissociation..... 184
6.10. van’t Hoff Analysis..... 185
6.11. Ethylation Interference Footprinting..... 186
6.12. Crystallization, structure determination and structural analysis of the A138T *Eco*RI-d(TCGCGAATTCGCG) complex..... 187
6.12.1. Preparation of macromolecules and crystal growth..... 187
6.12.2. Crystal cryoprotection and freezing..... 188
6.12.3. Data collection and processing 188
6.12.4. Structure Determination and Refinement 189
6.12.5. Quality checks on the final model 190
6.12.6. Structural analysis..... 190
6.13. Determination of *Eco*RI* site abundance in the Escherichia coli genome..... 191
APPENDIX A..... 192
DNA SUBSTRATES USED IN THIS WORK..... 192
APPENDIX B..... 195
BUFFER RECIPES 195
APPENDIX C..... 197
ERROR ANALYSIS 197
APPENDIX D..... 200
SFCHECK OF A138T-d(TCGCGAATTCGCG) CRYSTALLOGRAPHIC MODEL 200
APPENDIX E 205
PROCHECK OUTPUT FOR MODEL OF A138T *Eco*RI-d(TCGCGAATTCGCG) COMPLEX 205
APPENDIX F..... 211
LOCAL DNA BASE-PAIR AND BASE-STEP PARAMETERS OBTAINED FROM THE CRYSTAL STRUCTURES OF THE WILD-TYPE AND A138T *Eco*RI-d(TCGCGAATTCGCG) COMPLEXES..... 211
BIBLIOGRAPHY..... 217

LIST OF TABLES

Table 1.1 Changes in free energies of formation of EcoRI-DNA complexes ($\Delta\Delta G°_{ED}$) and transition state complexes ($\Delta\Delta G°_{I}^{\ddagger}$) for variant DNA sites [a] 19
Table 1.2 Specific and miscognate DNA cleavage rate constants 20
Table 1.3 Discrimination at the *Eco*RI endonuclease-DNA interface: Three different binding modes [a] 27
Table 2.1 Effect of flanking sequence on specific DNA binding *Eco*RI endonuclease 39
Table 2.2 Effect of *Eco*RI endonuclease mutations on specific a DNA binding and cleavage 41
Table 2.3 Thermodynamic parameters for Specific *Eco*RI binding at 25°C [a] 49
Table 2.4 Comparison of thermodynamic and kinetic constants for specific [a] binding between the wild type and H114Y enzymes 57
Table 2.5 Salt dependence parameters for *Eco*RI specific binding [a] 63
Table 3.1 Crystal parameters and data collection and refinement statistics for A138T-d(TCGCGAATTCGCG) Complex 70
Table 3.2 Comparison of hydrogen bonds at the protein-DNA interfaces of Wild-type and A138T *Eco*RI specific complexes[a] 72
Table 3.3 Effect of 7-deaza-Guanosine a substitutions within flanking DNA on EcoRI specific binding 76
Table 3.4 Effect of uracil substitutions within GAATTC flanking DNA on *Eco*RI specific binding 86
Table 3.5 Correlation between changes in Debye-Waller (B) factors at both sides of the Protein-DNA interface 94
Table 3.6 Thermodynamic parameters accompanying *Eco*RI-specific DNA binding at 25°C [a] 103
Table 4.1 Effect of *Eco*RI endonuclease mutations on non-cognate DNA binding 116
Table 4.2 Partitioning of *Eco*RI endonucleases among miscognate and non-specific DNA sites *in vivo* 120
Table 4.3 Effect of *Eco*RI endonuclease mutations on miscognate DNA cleavage 122
Table 4.4 Effect of flanking context on miscognate DNA binding and cleavage 125
Table 4.5 Predicted number of cuts in E. coli chromosomal DNA made by EcoRI endonuclease in vivo 131
Table 5.1 Effect of CaCl2 on DNA binding discrimination by the *Eco*RI endonuclease 145
Table 5.2 Effect of AAATTC flanking sequence on binding and cleavage 154

LIST OF FIGURES

Figure 1.1 Different levels of discrimination exhibited by different DNA binding proteins. 2
Figure 1.2 Three types of binding sites for the *Eco*RI restriction endonuclease 5
Figure 1.3 Direct readout and buttressing interactions at the *Eco*RI-specific DNA interface..... 7
Figure 1.4 Link between "direct" and "indirect" readout of sequence dependent information by the EcoRI endonuclease. 9
Figure 1.5 Coupling between DNA base and phosphate recognition at the *Eco*RI-specific DNA interface..... 10
Figure 1.6 A dual role for Arg145 in GAATTC binding and cleavage. 13
Figure 1.7 The "arms" of the *Eco*RI endonuclease are disordered in the absence of DNA. 15
Figure 1.8 Transition state interaction free energy change as a metric for discrimination..... 22
Figure 1.9 Phosphate contacts are important to *Eco*RI sequence discrimination. 24
Figure 1.10 Location of the "promiscuous" mutations in the *Eco*RI-TCGCGAATTCGCG complex. 31
Figure 2.1 Binding of *Eco*RI endonuclease to the specific DNA site in different flanking contexts. 38
Figure 2.2 Effect of flanking sequence on *Eco*RI specific DNA binding and cleavage. 42
Figure 2.3 Reaction coordinate diagrams showing how perturbation of the protein or DNA can affect endonuclease binding ($\Delta G°_{bind}$) and cleavage ($\Delta G°^{\ddagger}$)..... 43
Figure 2.4 Energetic components of specific and non-specific binding of *Eco*RI endonuclease. 46
Figure 2.5 Temperature dependence of the *Eco*RI-specific DNA interaction. 50
Figure 2.6 *In silico* modeling of the protein-DNA interface in the wild-type Ala138 region. 52
Figure 2.7 *In silico* modeling of the protein-DNA interface and the mutant H114Y region..... 55
Figure 2.8 Possible differences in the microscopic kinetic constants in reaction pathways of the wild-type and H114Y *Eco*RI endonucleases. 58
Figure 2.9 "Arms" of the *Eco*RI restriction endonuclease have a high abundance of charged amino-acids. 60
Figure 2.10 Salt dependence of the EcoRI-specific DNA interaction. 62
Figure 3.1 Secondary structure assignments in the wild-type and A138T *Eco*RI-d(TCGCGAATTCGCG) complexes. 68
Figure 3.2 Root mean square difference between coordinates of main chain atoms in wild-type and A138T-d(TCGCGAATTCGCG) complexes. 69
Figure 3.3 Overlay of the extended chain motifs of the wild-type and A138T-d(TCGCGAATTCGCG) complexes. 73
Figure 3.4 Stereo drawing of electron density covering the extended chain motif of the A138T-DNA complex. 73
Figure 3.5 Thr138 is involved in novel water mediated interactions with G_{-5}. 75
Figure 3.6 Molecular drawings of guanosine and 7-deaza-guanosine nucleosides. 75
Figure 3.7 Thr138 packs against G_{-5} in the A138T-specific DNA complex. 78

Figure 3.8 A138T complex lacks an ordered solvent molecule that is present at the wild-type-DNA interface. 79
Figure 3.9 Binding of *Eco*RI endonuclease to the specific DNA site in different flanking contexts. 83
Figure 3.10 Base-pair conformation parameter “opening” for DNA in wild-type and A138T-specific DNA complexes. 84
Figure 3.11 Differences in crystallographic B-factors in the structures of the wild-type and A138T specific DNA complexes. 93
Figure 3.12 Functionally important regions of the DNA backbone exhibit reduced mobility in the A138T–specific DNA complex. 94
Figure 3.13 Tight hydrogen bond between Lys113 and the scissile phosphate in the A138T-specific DNA complex. 96
Figure 3.14 The side-chain of histidine 114 is more ordered in the A138T complex than in the wild-type complex. 99
Figure 3.15 pH dependence of the *Eco*RI-specific DNA interaction in the presence and absence of calcium. 100
Figure 3.16 Water mediated hydrogen bond network couples contacts to both the GAATTC site and the flanking bases. 107
Figure 4.1 Kinetic proofreading prevents double stranded DNA cleavage by the *Eco*RI endonuclease *in vivo*. 113
Figure 4.2 Binding discrimination by the *Eco*RI endonuclease against miscognate and non-specific DNAs. 115
Figure 4.3 Differences in *Eco*RI endonuclease binding affinities between miscognate and non-specific DNAs. 118
Figure 4.4 Hierarchies of cleavage rate constants among *Eco*RI endonucleases. 123
Figure 4.5 Effect of AAATTC flanking context on cleavage rates. 124
Figure 5.1 Reaction coordinate diagrams showing how perturbations in the protein or DNA may affect endonuclease binding ($\Delta G°_{bind}$) and cleavage ($\Delta G°^{\ddagger}$). 139
Figure 5.2 Promiscuous mutant enzymes preferentially stabilize transition states for cleavage of miscognate DNA sites. 140
Figure 5.3 Ca2+ inhibits DNA cleavage by competing with Mg2+ for active site binding. 143
Figure 5.4 pH dependence of endonuclease binding as a probe for active site assembly. 144
Figure 5.5 Calcium reduces discrimination by the A138T enzyme relative to the wild-type. 146
Figure 5.6 Mg2+ affinity of active sites in *Eco*RI-miscognate DNA complexes. 150
Figure 5.7 Effect of flanking sequence on the *Eco*RI specific and miscognate DNA interactions. 155
Figure 5.8 Contacts to DNA phosphates are important in discrimination against miscognate sites. 157
Figure 5.9 Ethylation interference footprints made by the *Eco*RI endonuclease on the cgcAAATTCgcg site in the absence of Ca^{2+}. 159
Figure 5.10 Ethylation interference footprints made by the *Eco*RI endonuclease on the cgcAAATTCgcg site in the presence of Ca^{2+}. 160
Figure 5.11 Ethylation interference footprints on the cgcAAATTCgcg site embedded in a 22 base-pair double stranded deoxyoligonucleotide. 163
Figure 5.12 Ethylation interference footprints on the ggtAAATTCacc site embedded in a 22 base-pair double stranded deoxyoligonucleotide. 165

Figure 5.13 Ethylation interference footprints made by the *Eco*RI endonuclease on the ggtAAATTCacc site in the presence of Ca^{2+}..... 166
Figure 5.14 Effect of flanking sequence on AAATTC foot prints. 169
Figure 5.15 Network involving Thr138 and residues in the outer arm segment which contact the phosphate backbone. 171

1. INTRODUCTION

Protein-DNA interactions lie at the heart of many crucial biological processes, including transcription, DNA replication, and genome repair. In order to understand how DNA binding proteins achieve functional specificity and select their target sequences in the cell among other DNA sites that are chemically similar and highly abundant, an understanding of the structural and energetic rules that govern molecular recognition by DNA binding proteins is required. With the knowledge of the molecular origins of specificity associated with DNA interacting proteins, we should be able to engineer novel research tools and/or therapeutic agents that will target aberrant processes involving protein-DNA interactions.

1.1. Levels of sequence selectivity in proteins that interact with DNA

The diverse functional roles of DNA binding proteins dictate that they have different levels of DNA specificity. DNA polymerases, helicases, and histones must interact non-specifically with DNA in order to carry out their respective biological functions. This is in contrast to transcriptional activators and repressors, which must have strong selectivity for their target sequences relative to non-specific DNA in order for the few protein molecules to bind appreciably to the one or two copies of the operator present in the cell. Indeed, the *lac* repressor prefers binding to its operator DNA over non-specific DNA by 10-billion fold (Jen-Jacobson, 1997). Other proteins, such as the cI and cro repressors from the λ bacteriophage bind to a series of related operator sites (each with one or more base-pair changes) in a graduated fashion ("permissive discrimination"), although these proteins still show stringent discrimination between the set of operator sites and non-specific DNA (Harrison et al., 1990) (Figure 1.1). The graduated affinity of these proteins to the different operators is carefully tuned to their function in the "genetic switches" of the bacteriophage that encode the repressors (Albright et al., 1998).

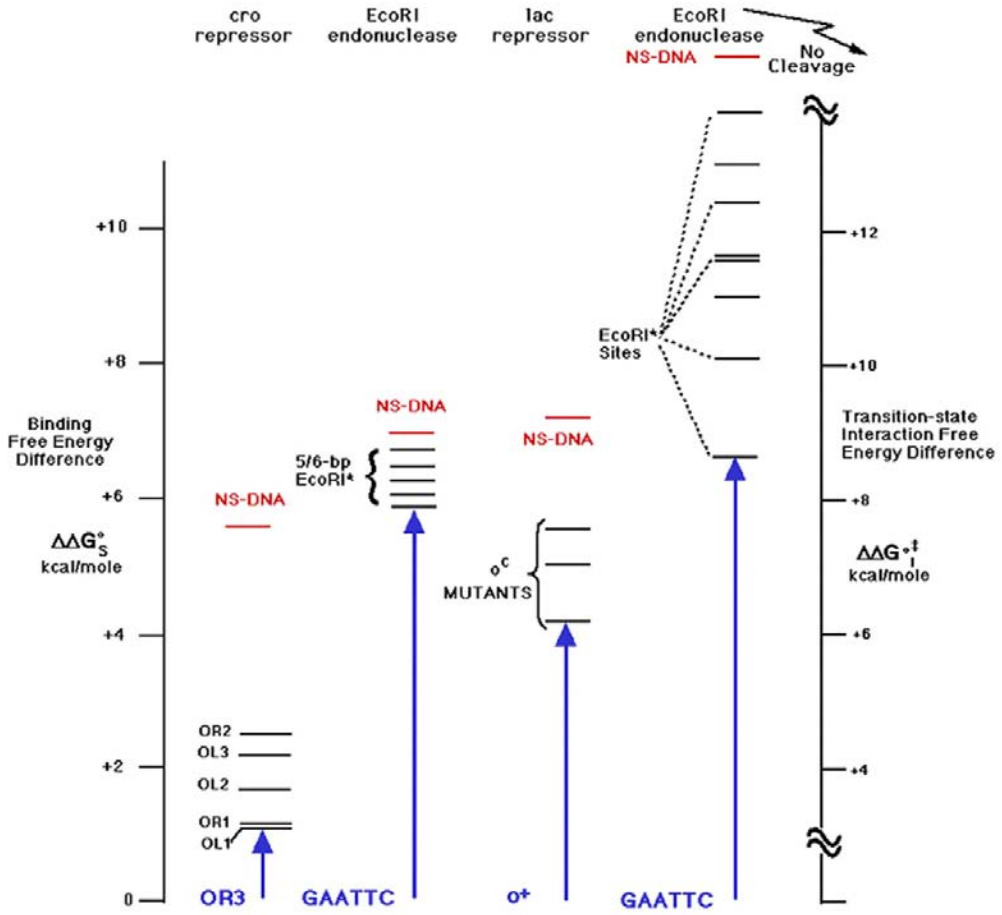


Figure 1.1 Different levels of discrimination exhibited by different DNA binding proteins.

Discrimination by the λ cro repressor, the lac repressor and the *Eco*RI restriction endonuclease against non-specific DNA (red bars) and DNA sites that differ from the cognate site by one base-pair (black bars). On the left hand axis, $\Delta\Delta G°_S = -RT\ ln\ K_{eq.,2}/K_{eq.,1}$, where $K^1_{eq,1.}$ is the equilibrium binding constant for the cognate substrate, and $K_{eq,2.}$ is the equilibrium binding constant for the non-cognate site. For each system, the cognate site is at $\Delta\Delta G°_S = 0$. The axis on the right hand side takes into account both binding and cleavage discrimination, and thus only applies to the endonuclease. See equations 1.1 – 1.4 and figure 1.8 for the definition of $\Delta\Delta G°^{\ddagger}_I$.

1.2. Restriction endonucleases: Model System for understanding stringent discrimination of DNA binding proteins

We have postulated that studying the DNA-binding proteins which exhibit the highest level of sequence discrimination will yield the largest number of rules and constraints that govern specificity. Further, we expect that the structural and energetic rules applicable to proteins interacting with DNA non-specifically or with permissive discrimination will be understood in terms of relaxation of the factors which ensure specificity in the most stringent systems. Restriction endonucleases must have the highest level of sequence discrimination in order to carry out their function of inactivating unprotected (un-methylated) foreign DNA, while concomitantly avoiding cleavage at related DNA sites within the cells' own genome. This is a matter of life and death to the cell expressing the endonuclease, as restriction of cellular DNA is potentially a lethal event. Therefore, for the type II restriction endonuclease, *Eco*RI, the second order rate constant (which takes both binding and cleavage into account) is as much as10^{11}-fold lower for cleavage of a site with a one base-pair change relative to the cognate GAATTC site (Jen-Jacobson, 1997) (Figure 1.1). This extremely high level of discrimination reflects the fact that restriction endonucleases, unlike non-catalytic DNA binding proteins, exercise discrimination not only in binding, but also in cleavage. This fact also underscores an advantage of studying the restriction enzyme over a non-catalytic counterpart because we can not only compare binding affinities for different classes of DNA sites but also the rates of cleavage of the cognate and related sites. Measuring the ability of an enzyme to discriminate against non-cognate sites at the cleavage step provides a test for functional specificity that is not available in studying transcriptional repressors or other non-catalytic DNA binding proteins.

1.3. Structural and energetic determinants of specificity in the *Eco*RI restriction endonuclease

Although many of the structural and energetic determinants of specificity in restriction endonucleases that I will describe in the following sections have broad applicability to each of

the type II restriction endonucleases studied, I will focus my introduction on principles specifically pertaining to sequence discrimination in the *Eco*RI system for two reasons: 1) The *Eco*RI restriction endonuclease is among the most well studied of all the DNA-binding proteins; thus many of the structural and energetic insights into the origins of specificity of DNA binding proteins in general and type II restriction endonucleases in particular were gleaned from examination of the *Eco*RI system. 2) The data to be presented in this thesis pertain to the *Eco*RI enzyme; thus an introduction into the idiosyncrasies of this endonuclease provides important context for the interpretation of my experimental results. In the following sections, I will outline what is known about how the molecular features of the *Eco*RI endonuclease and its complexes with different classes of DNA substrates result in the remarkable specificity exhibited by the enzyme.

1.3.1. Three types of DNA binding sites for the *Eco*RI restriction endonuclease

Type II restriction endonucleases function as homodimers that bind to and cleave within palindromic DNA recognition sites. As a consequence of this symmetric interface, the *Eco*RI enzyme cleaves at equal rates between the G and A residues in both strands of the GAATTC site, and double stranded cleavage is the likely outcome of a single binding event (Halford et al., 1983). There are nine possible miscognate sites (historically referred to as *Eco*RI* sites in the literature; miscognate site and *Eco*RI* site will be used interchangeably in this work) which differ from GAATTC by a single base-pair change. The base-pair change in these sites disrupts the rotational symmetry characteristic of the cognate site (Figure 1.2). The miscognate site asymmetry is typically translated to asymmetry in the rate constants for cleavage of the two different DNA strands, with cleavage being more rapid in the cognate half-site (Figure 1.2) (Lesser et al., 1990). Finally, non-specific DNA sites are defined as matching the cognate site in $\leq$ 4/6 base-pairs, and these sites are completely resistant to cleavage (Lesser et al., 1990).

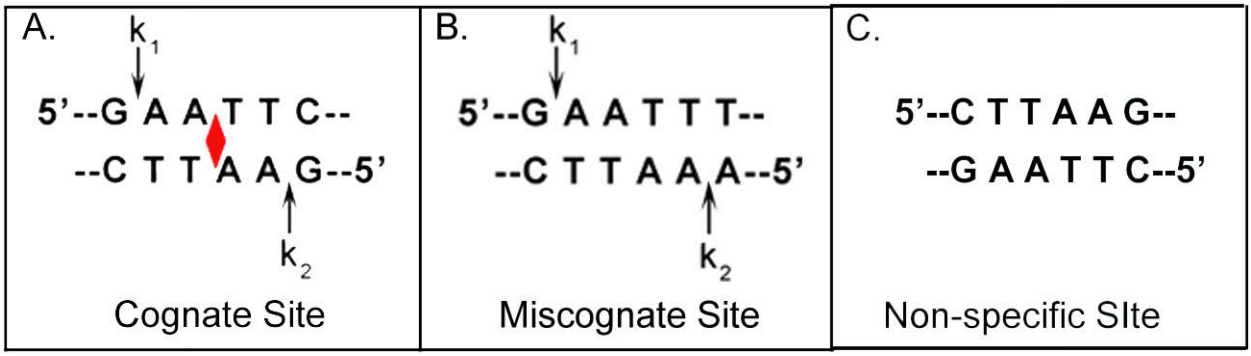


Figure 1.2 Three types of binding sites for the *Eco*RI restriction endonuclease

The cognate site, GAATTC is shown in panel A, where the red triangle represents the axis of rotational symmetry. An example of one out of nine possible asymmetric miscognate sites is shown in panel B. Note that each miscognate site contains a cognate $\frac{GAA}{CTT}$ half-site, and a miscognate half-site ($\frac{TTT}{AAA}$ in this case). The constants k_1 and k_2 refer to the rates of cleavage of the cognate and non-cognate half-sites respectively. A non-specific site is shown in panel C.

1.3.2. Structural view of specificity of the EcoRI restriction endonuclease

In the following sections, I will describe how analysis of the crystal structures of the *Eco*RI-d(TCGCGAATTCGCG) DNA complex, the free enzyme, and the free DNA provide insight into many of the factors that drive formation of the specific complex. Further, I will discuss how non-cognate structures from other site-specific DNA binding systems illuminate some general principles likely used by the *Eco*RI endonuclease and other proteins to discriminate against non-cognate DNA sites.

1.3.2.1. "Direct Readout" of the GAATTC site by the EcoRI endonuclease

In the *Eco*RI-specific DNA complex (McClarin et al., 1986, Kim et al., 1990, Grigorescu, 2003) there are 18 protein-base hydrogen bonds, among them, 14 are direct protein-base contacts, and 4 are mediated by water molecules. The complex is also stabilized by Van der Waals contacts between the protein and thymine methyl groups as well as the non-polar surfaces

of cytosine bases (Figure 1.3). These protein-base contacts comprise the "direct readout" of information (sequence dependent arrangement of functional groups) in the major groove of the DNA double helix.

It is clear that the individual base recognition contacts are not formed independently, as most amino acids involved in hydrogen bonds with DNA bases are also hydrogen bonded to each other (Figure 1.3). These buttressing interactions between DNA contacting amino-acids indicate that the contacts are formed cooperatively. For the inner AATT base pairs of the GAATTC recognition site, the DNA-base networks are robust to disruption as indicated by exhaustive structural perturbation studies using DNA base-analogues. The data show that the free energy perturbations to these networks are strictly additive (Jen-Jacobson, 1995). This suggests that if one or two of three of these contacts are deleted, the other contacts remain in place.

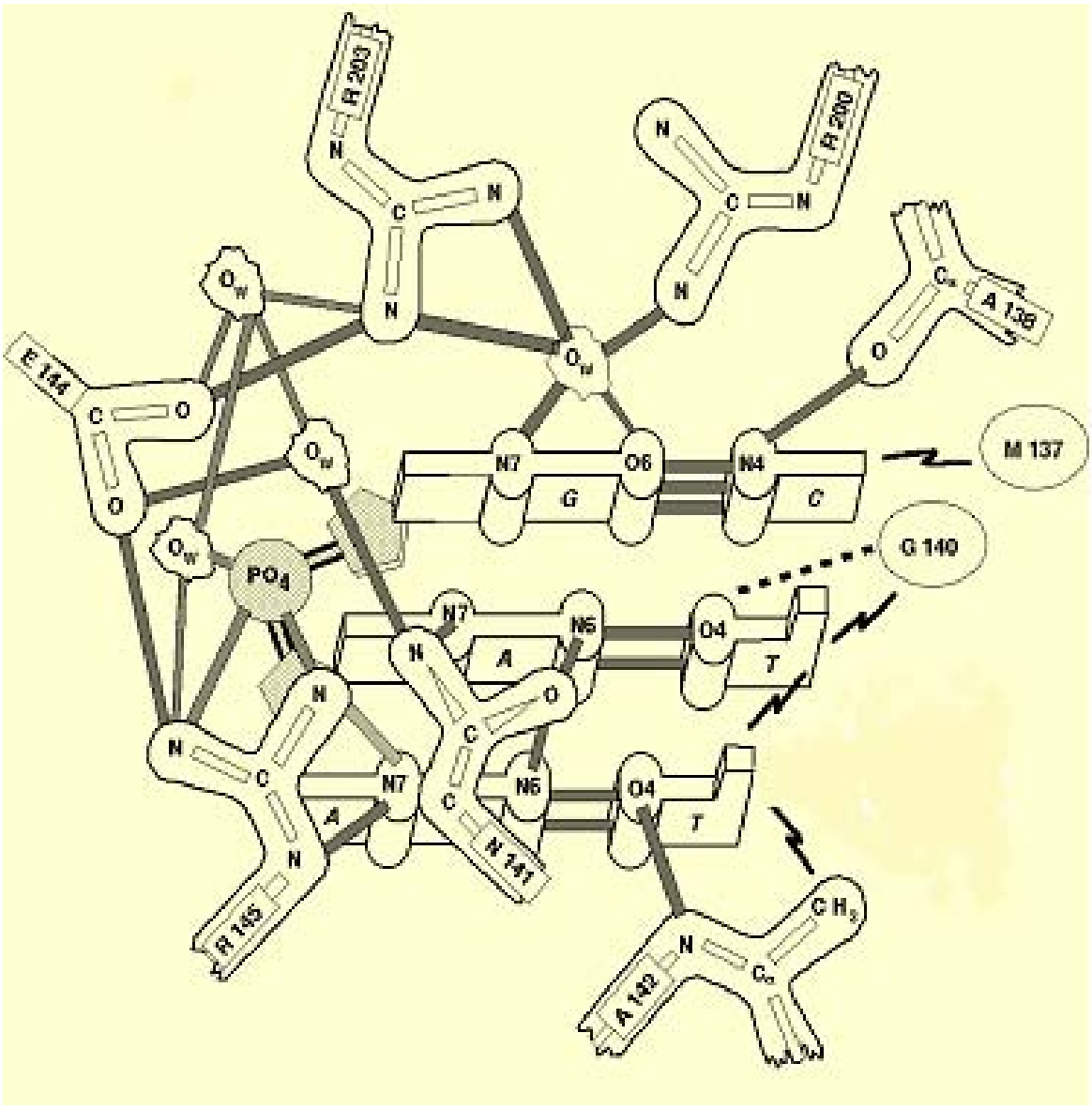


Figure 1.3 Direct readout and buttressing interactions at the *Eco*RI-specific DNA interface

Hydrogen bonds are shown as tubes, Van de Waals interactions as 'lightning', and the main-chain C_α-H—O hydrogen bond is shown as a dotted line. O_w denotes a water molecule. Figure was kindly provided by Dr. John M. Rosenberg.

1.3.2.2. The *Eco*RI-DNA "kink" and "indirect readout"

The *Eco*RI endonuclease distorts its recognition sequence upon binding (Figure 1.4) (Frederick et al., 1984). DNA distortion is a common feature in the complexes of DNA binding proteins with their target sequences. Two examples are the catabolite activator protein (CAP), which induces two localized kinks in the DNA which account for the observed DNA bending angle of 87-90 degrees (Schultz et al., 1991, Parkinson et al., 1996), and the TATA box binding protein, which induces two kinks at both ends of the DNA recognition sequence, resulting in the widening of the minor groove (Kim et al., 1993). Crystallographic and solution studies of duplex DNA suggest that the propensity of a DNA molecule to bend is dictated by its sequence (Dickerson et al., 1997, Olson et al., 1998). The term "indirect readout" was coined to describe the use of this sequence dependent DNA conformational information in site recognition by DNA binding proteins (Otwinowski et al., 1988). "Indirect readout" appears to be important in the preference of the *Eco*RI endonuclease for GAATTC based on the following rationale and observations: 1) It has been hypothesized that the energetic cost associated with achieving the kinked DNA conformation observed in the complex will be lowest for GAATTC relative to other DNA sites (Lesser et al., 1990, Lesser et al., 1993). 2) Since the major groove is wider and shallower in the *Eco*RI-bound DNA relative to the free DNA (Figure 1.4), induction of the kinked DNA conformation is crucial for the accommodation of base-recognition elements into the major groove, and hence the formation of many interactions involved in the "direct readout" of GAATTC. 3) The distorted geometry of the complexed DNA is also critical for DNA cleavage. The scissile phosphates in both strands of the GAATTC site are brought into novel positions upon *Eco*RI binding relative to the positions observed in the free DNA (Figure 1.4), such that the DNA distortion induced by *Eco*RI binding is required for the proper juxtaposition of the enzyme active sites with the scissile phosphates in both DNA strands (Grigorescu, 2003). Further, a DNA perturbation expected to create a steric conflict within the DNA kink exacts a very small binding penalty, yet results in a large reduction in the DNA cleavage rate (Kurpiewski et al., 2004). These observations also demonstrate that "direct" and "indirect" readouts of information are not mutually exclusive but rather intertwined determinants of specificity.

A further example of the interrelatedness of the "direct" and "indirect" readout components of the specific DNA recognition mechanism of the *Eco*RI endonuclease is

appreciated from an analysis of protein-phosphate contacts in the specific complex. In the structure with the cognate DNA site, the *Eco*RI endonuclease makes hydrogen bonded and ionic contacts to 14 phosphates (7 in each DNA strand). Of these 14 phosphates, six (three on each DNA strand at **p**N**p**GAA**p**TTC) have been described as crucial phosphate "clamps" (Lesser et al., 1990). These "clamp" phosphates are individually contacted by multiple protein moieties with high geometric precision (Lesser et al., 1992, Kurpiewski et al., 1996). Because the clamp phosphates are brought into novel positions as a result of the DNA kink imposed by enzyme binding, interactions between the enzyme and "clamp" phosphates are believed to stabilize the distorted DNA conformation, and the DNA distortion is in turn necessary for the insertion of the base recognition elements into the major groove. Figure 1.5 illustrates an example of the coupling between phosphate recognition and direct readout of the sequence specific DNA functional groups, where a hydrogen bond network involving a residue making a contact to the first base of the recognition site and the clamp phosphate at N**p**NGAATTC is shown.

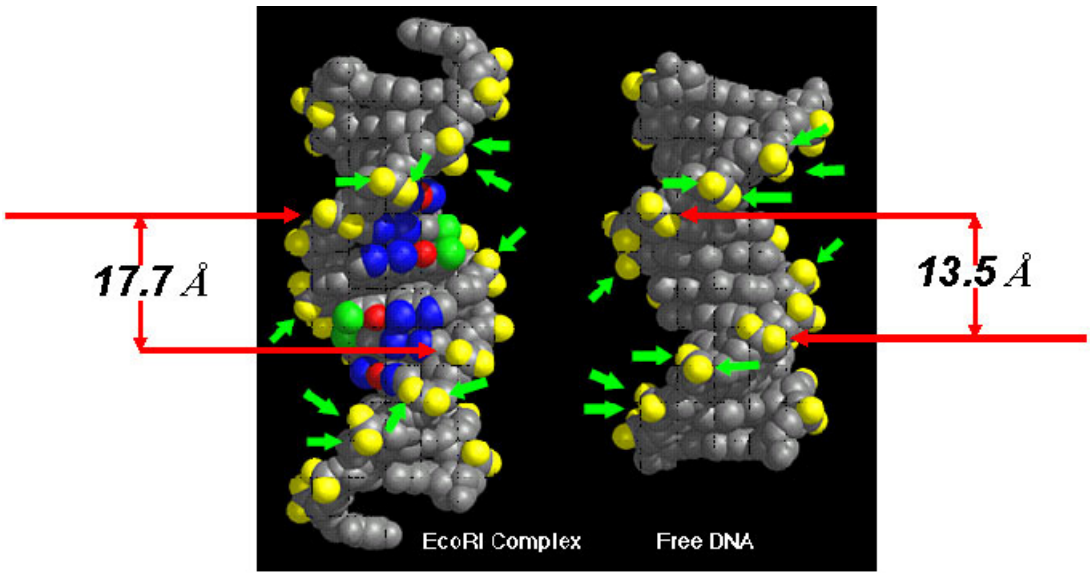


Figure 1.4 Link between "direct" and "indirect" readout of sequence dependent information by the EcoRI endonuclease.

The structures of the free DNA (pdbid 355D), and the DNA from the *Eco*RI-specific DNA complex (Grigorescu, 2003) are compared. The two scissile phosphates are highlighted by the red arrows with the inter-scissile phosphate distances shown. Phosphates that are contacted by the enzyme are highlighted by green arrows. The contacted phosphates and scissile phsopahtes adopt novel positions upon DNA binding demonstrating that direct and indirect readout by the EcoRI endonuclease are linked. This figure was made by Dr. Linda Jen-Jacobson, and the analysis was performed by Dr. Linda Jen-Jacobson and Dr. Arabela Grigorescu.

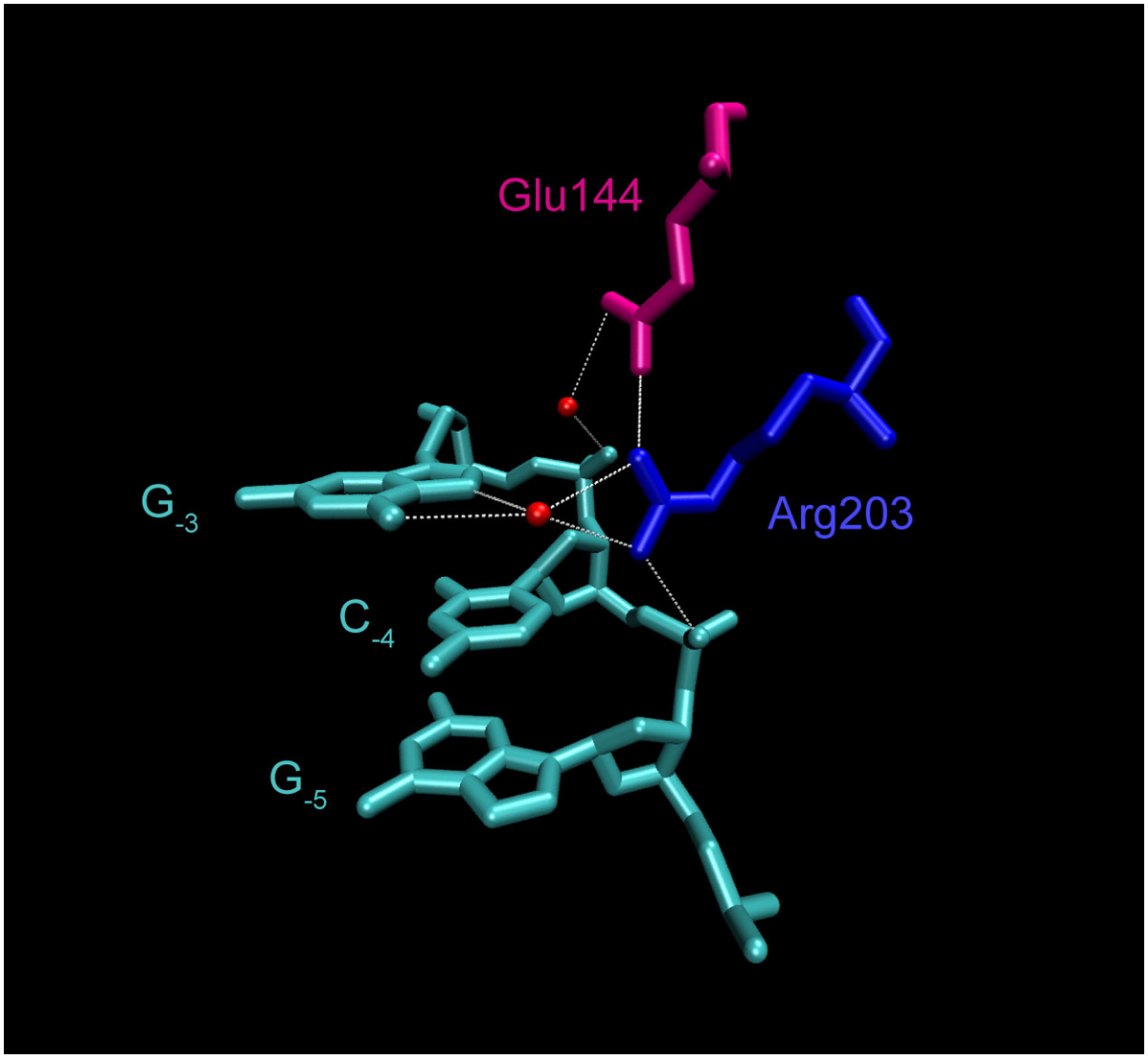


Figure 1.5 Coupling between DNA base and phosphate recognition in the specific complex

Hydrogen bond network involving residues contacting two "clamp" phosphates (N_{-5}**p**N_{-4}**p**G_{-3}AATTC) and the G base of the GAATTC recognition site. Note that Arg203 not only makes water mediated contacts with the first base of the recognition site (**G**AATTC), but also makes hydrogen-bonds with a key "clamp" phosphate upstream of the site, as well as a buttressing interaction with Glu144, which itself makes a water mediated contact to a "clamp" phosphate. Hydrogen bonds are depicted as white dotted lines in the figure.

1.3.2.3. Coupling of DNA recognition to catalysis

In order for restriction endonucleases to carry out their biological function of inactivating foreign DNA molecules, tight binding to specific recognition sites within invading DNA must be coupled to cleavage of *both* DNA strands, as single stranded nicks are repaired by DNA ligase *in vivo* (Heitman et al., 1989). The enzyme must also avoid cleaving miscognate sites since these sites are numerous within the bacterial chromosome and unprotected by the companion methylase. The design of homodimeric enzymes which recognize symmetrical target sequences is an effective strategy for meeting these functional requirements. Examination of the *Eco*RI-specific DNA interface explains how recognition of GAATTC is coupled to DNA cleavage, and how mechanical communication between the two enzyme subunits will lead to cleavage of both strands of the symmetrical GAATTC site as well as to aid in discrimination against cleavage of asymmetric miscognate sites.

The amino acid side chains, Asp91, Glu111, and K113 have been shown to be critical for catalysis by the *Eco*RI enzyme (King et al., 1989, Grabowski et al., 1995); and hence constitute the enzyme's active site. The catalytic amino-acids are contiguous with residues that make contacts to the clamp phosphates at N**p**NGAATTC (Ser87), NN**p**GAATTC (Lys89), and NNGAA**p**TTC (Gly116). These backbone interactions therefore anchor the active site amino acids in place with geometric precision in the specific complex, providing an example of coupling between DNA recognition and active site assembly.

The Arg145 amino acid also plays a dual role in recognition and catalysis (Jen-Jacobson, 1997, Grigorescu, 2003). The side-chain of Arg145 donates a hydrogen bond to the scissile phosphate (Figure 1.6A). Mutagenesis experiments suggested that Arg145 is important for catalysis (Wolfes et al., 1986). Indeed, interaction of this positively charged amino acid with the scissile phosphate may stabilize the phosphorane intermediate during catalysis. Arg145 is also involved in the direct readout of the recognition site as Arg145 is hydrogen bonded to the N7 group of the inner adenine of GAATTC (Figure 1.6A). Interestingly, the inner adenine is located at the center of the *Eco*RI-DNA kink [the roll angle between the central ApT base step is -54° (Kim, 1994) compared with 4° in the structure of the free B-DNA of the same sequence (Shui et al., 1998)], thus the interaction between Arg145 and the inner adenine plays a role in "reading" DNA distortion in the complex. Further, it has been postulated that by its involvement in a

"cross-talk ring" (Kurpiewski et al., 2004), Arg145 is also involved in mechanically coupling the two different scissile phosphates, which themselves are about 20 Å apart (Figure 1.6A). The ring is a structural feature in which the guanidine group of Arg145 of each subunit is hydrogen bonded to Glu144 of the other subunit, with the other two sides of the ring being composed of the peptide bonds linking residues 144 and 145 of the same subunit. The crosstalk ring is further linked to DNA recognition as the methylene group of Arg145 is packed against the main-chain carbonyls of Asn141 and Ala142, two amino acids whose side-chains contact GAATTC. The crosstalk ring is also formed in the specific complex of the *Mun*I restriction endonuclease (Deibert et al., 1999) (recognition site CAATTG), which distorts the central AATT module of its recognition site in a similar manner to *Eco*RI (Figure 1.6B). This observation supports the suggestion that the ring reads/stabilizes the kinked conformation of the DNA. Taken together, Arg145 and the crosstalk ring provide structural basis for reading DNA distortion, coupling DNA recognition to catalysis, and providing a mechanical link between the catalytic centers of the two DNA half-sites.

A structural perturbation (substitution of a non-bridging scissile phosphoryl oxygen with a methyl group) made in one strand of a DNA molecule containing the specific recognition sequence that is expected to have a sublte effect on the geometry of the crosstalk ring affected the cleavage rate of not only the strand containing the perturbation, but also reduced the cleavage rate of the unmodified strand (Kurpiewski et al., 2004). If the effect of this minor perturbation in one DNA half-site is transmitted to the unmodified half-site, one can imagine that a more severe DNA perturbation (such a single base-pair change) would have the same effect on the cleavage rate for the unmodified DNA half-site. Indeed, the rates of cleavage of miscognate DNAs are reduced in both the modified and unmodified half-sites (see section 1.3.3.1). Perturbation of the crosstalk ring is likely partly responsible for preventing double stranded cleavage of sites closely related to GAATTC by transmitting the locally acute effects of the base-pair change in the modified half-site to the unmodified site.

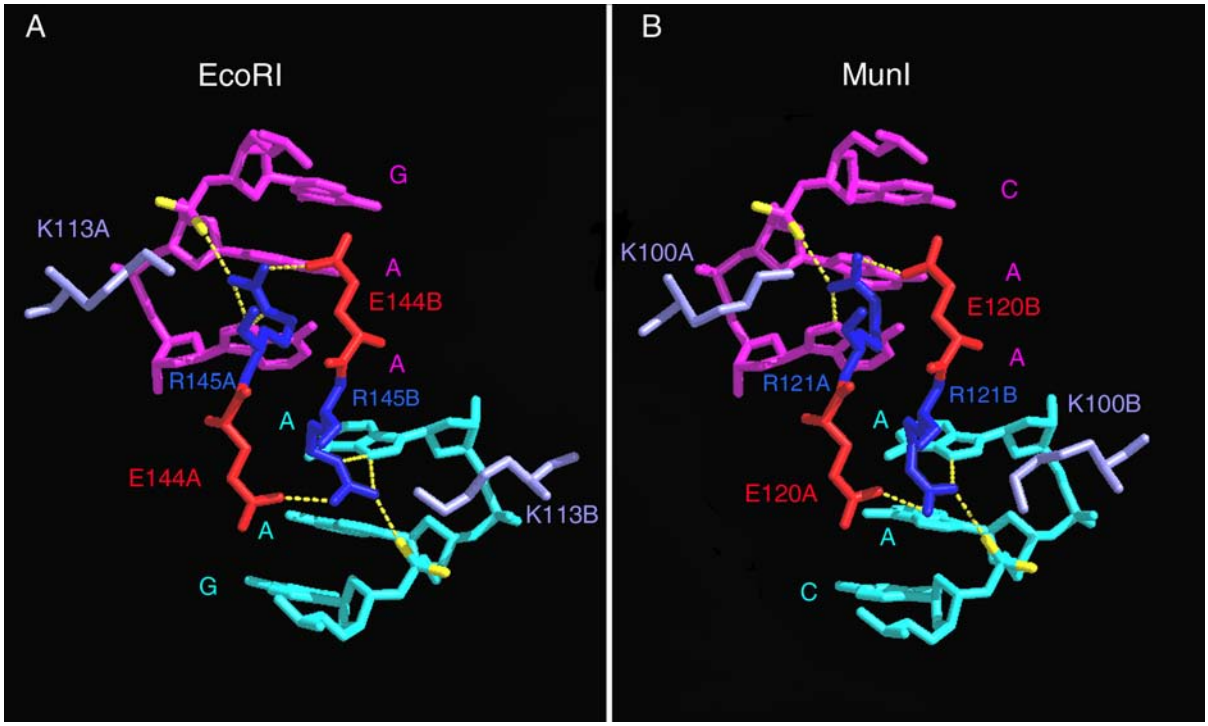


Figure 1.6 A dual role for Arg145 in GAATTC binding and cleavage.

Panel A) Arg145 is shown making hydrogen bonds (yellow dashed lines) with the scissile phosphate, and the central adenine of GAATTC. The *Eco*RI "crosstalk ring" is composed of Glu144 and Arg145 from both subunits of the homodimer (subunit designations A & B), and results in communication between the two active sites in the complex. Panel B) Near identical structure of the "crosstalk ring" in the complex between the *Mun*I restriction and its recognition sequence, CAATTC. Figure taken from (Kurpiewski et al., 2004).

1.3.2.4. An enzyme disorder to order transition is coupled to specific DNA binding

The structure of the *Eco*RI endonuclease in complex with specific DNA revealed two major protuberances emanating from the main body of the enzyme (McClarin et al., 1986), (Kim et al., 1990). These outgrowths which encircle the DNA in the specific complex were termed the "arms" of the enzyme (Figure 1.7). Upon inspection of the topology of the complex, Rosenberg and colleagues immediately appreciated that: "DNA binding must be associated with a conformational change of the protein because the arms encircle the DNA to such an extent that it is unlikely that the DNA could enter the active site in the absence of some movement" (McClarin et al., 1986). Indeed, the arms are disordered in the crystal structure of the apo-enzyme (Grigorescu, 2003). Grigorescu proposed that that the intermolecular contacts between the arms and the specific recognition site (as well as contacts between the arms and the DNA backbone

within and surrounding GAATTT) stabilize the structure of the arms observed in the *Eco*RI-specific DNA complex (Grigorescu, 2003). The mechanism of this conformational change is not known.

Coupling between the local folding of protein elements involved in DNA recognition to DNA binding is a common theme among DNA binding proteins. Other site-specific binding proteins that undergo coupled folding transitions include GCN4 (Weiss et al., 1990), the retinoid x receptor (Holmbeck et al., 1998), LEF-1 (Love et al., 2004), *lac* repressor (Spolar et al., 1994), and the *Eco*RV restriction endonuclease (Winkler et al., 1993, Perona et al., 1997) among others. Coupling of local folding of the recognition elements to specific binding would contribute to the specificity of protein-DNA interactions if only the specific recognition site stabilizes the folded conformation. The same hypothesis (framed in alternative terms) states that only in the presence of the favorable enthalpy gained from intermolecular contacts between the protein and the specific recognition site can the penalty in configurational entropy associated with the folding transition be offset (Freire, 1998, Uversky, 2002). Grigorescu suggested that if this folding transition is only triggered by the specific recognition sequence, the arms would play an important role in discrimination by the *Eco*RI endonuclease (Grigorescu, 2003). However it remains unclear if and how the *Eco*RI endonuclease folding transition is perturbed upon binding to miscognate and non-specific sites. Thermodynamic data presented in section 1.3.3.3 support the idea that the conformational change involving the *Eco*RI arms is not coupled to non-specific complex formation.

Interestingly, many key *side-chains* involved in GAATTC recognition are also disordered in the apo-enzyme (Grigorescu, 2003). Among the disordered side-chains are: Lys89 (contacts the **p**GAATTC clamp phosphate), Lys113 (electrostatic contact to scissile phosphate), His114 (electrostatic contact to GA**p**ATTC phosphate), Asn141 (hydrogen bonds to both adenines in GAATTC), Arg145 (hydrogen bonds to scissile phosphate and inner adenine, constituent of "crosstalk ring"-see section 1.3.2.4), and Arg200 (water mediated interactions to guanosine of the recognition site). The role of these amino acids in recognizing the distorted geometry of the DNA bases and DNA backbone highlights the critical role of DNA distortion in the induced fit between the protein and specific recognition sequence, and how perturbations which may prevent the DNA from attaining this detailed conformation (such as single base-pair changes) will lead to discrimination against such DNA sites.

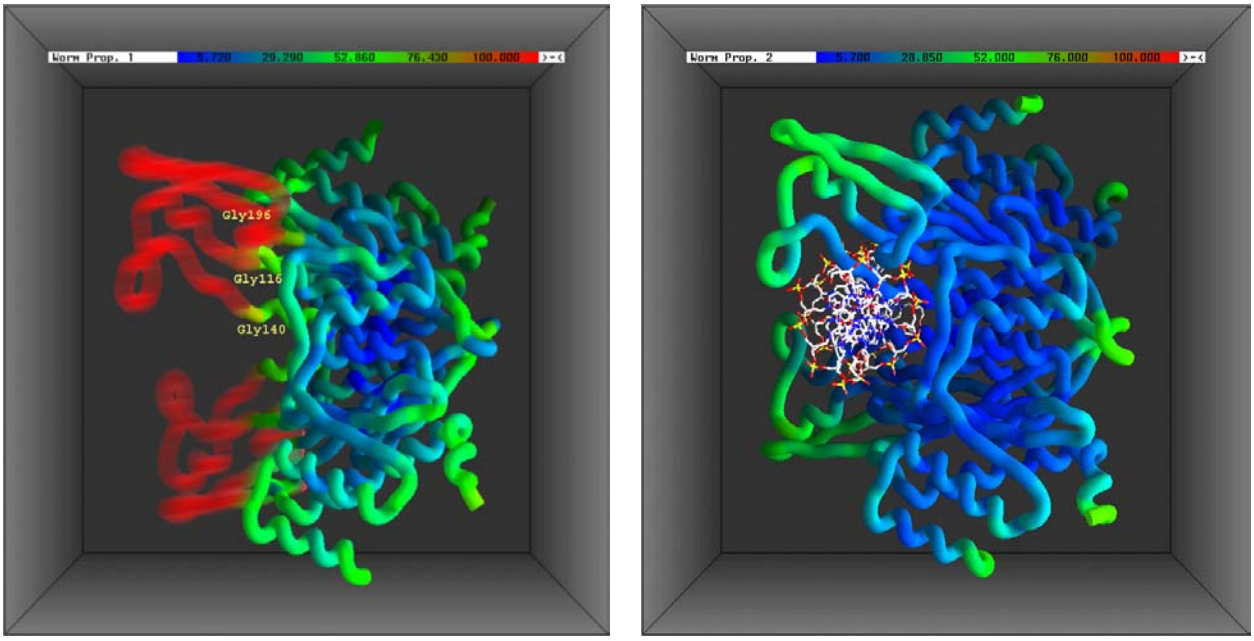


Figure 1.7 The "arms" of the *Eco*RI endonuclease are disordered in the absence of DNA.

Qualitative model depicting the distribution of structural stability in the unbound *Eco*RI dimer (A), and in the specific DNA complex (B). Both structures are described in (Grigorescu, 2003). The backbone of the protein is color-coded according to atomic crystallographic B-factors; Dark blue B=20 $Å^2$ (highly stable regions), green B=50 $Å^2$ (intermediate stability), red B=100 $Å^2$ (these regions constitute the "arms" of the endonuclease, and are disordered in the absence of DNA, and are thus blurred in panel A. The B-factors of the structures were re-scaled such that the atoms that form the hydrophobic core of the main domain have identical B-factors. Comparative analysis was performed by and figure was kindly provided by Dr. Arabela Grigorescu.

1.3.2.5. Structures of miscognate and non-specific complexes are required to complete the structural view of specificity

There are currently high resolution structures of 14 different type II restriction endonucleases bound to their cognate recognition sites available in the Protein Databank (Berman et al., 2000). While these structural snapshots provide valuable insight into the factors that stabilize the specific complexes, structures of miscognate and non-specific complexes are required to complete a structural view of specificity. There is no restriction endonuclease for

which structures of the free enzyme, as well as the specific, miscognate, and non-specific complexes are known.

There are only two structures of restriction endonucleases in complex with non-cognate DNA sites. A 3.0 Å resolution structure of the *Eco*RV restriction endonuclease in complex with the non-specific DNA, CGAGCTCG is available (Winkler et al., 1993). However, this complex is somewhat artificial in that the enzyme is bound to two DNA octamers arranged end to end, with the octamer-octamer junction positioned across from the active site of the enzyme. Therefore, the "scissile" phosphate is absent from this complex. A high resolution (1.9 Å) structure of the *Bam*HI restriction endonuclease (recognition sequence GGATCC) in complex with a DNA containing the miscognate site, GAATCC (Viadiu et al., 2000) is the other example, however, there is reason to believe that this snapshot does not represent a true miscognate complex either. First, the 11-base-pair oligonucleotide observed in the complex is too short to allow for the formation of all the phosphate contacts directly observed in the crystal structure of the specific complex (Newman et al., 1994), or inferred to exist from ethylation interference patterns of miscognate complexes (Engler, 1998). Further, the crystals were grown at the non-physiological pH of 4.8, meaning that the ionization states of amino acids, and hence the electrostatic properties of the protein and DNA do not reflect those of the physiological complex. Rosenberg and colleagues solved the preliminary crystal structure (3.5 Å resolution) of the *Eco*RI endonuclease in complex with a 17 base-pair oligonucleotide containing the GACTTC miscognate site (Wilkosz, 1993), although publication of this structure awaits further refinement of the model. Taken together, these structures are either incomplete, represent snapshots of the non-specific binding mode or artifacts of the solution conditions used during crystallization. Thus, a structure of a true restriction endonuclease miscognate complex remains elusive.

Aside from the complexes of restriction enzymes discussed above, there are only three additional structures of non-cognate structures available among all other site-specific DNA binding proteins: the glucocorticoid receptor DNA binding domain (Luisi et al., 1991), a monomer of the cro repressor bound to a consensus half operator (Albright et al., 1998), and the *lac* repressor bound to a non-specific DNA site (Kalodimos et al., 2004). There are features common to each of these non-cognate complexes that shed light on possible general mechanisms of discrimination by specific DNA binding proteins. In each of the non-cognate complexes, the association between the protein and DNA appears to be loose, as the total solvent accessible

surface area (ASA) buried in the non-cognate complexes is much smaller than in the specific complexes. For example, the *lac* repressor, in binding to the cognate operator *O1* results in 4900 $Å^2$ of buried surface area compared to 3500 $Å^2$ buried surface area in the non-specific complex (Kalodimos et al., 2004). A consequence of this reduction in buried (non-polar) surface area would be a reduction in the favorable energetic contribution from the hydrophobic effect accompanying formation of the non-specific complex relative to the specific one. In contrast to the elaborate networks of protein-base and buttressing hydrogen bonds that exist at the interfaces of specific complexes, there tend to be few if any protein base contacts in the non-cognate structures. In addition, while the DNA phosphate backbone is contacted in the non-specific complexes, the contacts tend to be ionic contacts which are much less geometrically constrained compared to the hydrogen bonds involving main-chain and short polar side-chains that are observed in specific complexes. In contrast to the dramatic DNA distortions seen in the cognate DNA complexes with the *Eco*RV enzyme (Winkler et al., 1993) and the *lac* repressor (Kalodimos et al., 2004), the DNAs in the corresponding non-cognate complexes exhibit little or no distortion. Given the importance of DNA distortion in the recognition and cleavage mechanisms of the *Eco*RI endonuclease, a failure to achieve the kinked DNA conformation in miscognate sites would drastically reduce binding affinity as well as the rates of DNA cleavage, and hence play a large role in sequence discrimination in the *Eco*RI system.

The structures of the operator DNA (Kalodimos et al., 2001), (Kalodimos et al., 2004) and non-specific DNA (Kalodimos et al., 2004) complexes with the *lac* headpiece were solved by nuclear magnetic resonance (NMR) spectroscopy, which allows for a comparison of the dynamic as well as structural properties of the two complexes. The non-specific complex was found to be quite flexible on the biological time scale indicating that the binding funnel for the non-specific complex ensemble is wide and rugged, where inter-conversion among many microstates is likely. This is in contrast to the *lac*-operator complex, where the specifically bound protein exhibits significantly reduced mobility relative to the mobility of the free and non-specifically bound protein. These experimental findings are consistent with the molecular dynamics simulations of the *Eco*RV-DNA complexes, where motions of the protein and DNA are greatly increased in the non-specific complex relative to the motions in the specific complex (Duan et al., 1996). These results suggest that non-specific complexes are relatively loose and dynamic compared to the specific counterparts.

1.3.3. Thermodynamic and kinetic views of specificity of the *Eco*RI restriction endonuclease

Analysis of the structures of the free macromolecules and the *Eco*RI-specific DNA complex has been a powerful tool in advancing our understanding of the specificity mechanisms of the *Eco*RI restriction endonuclease. However, structure gazing often raises as many questions as it answers. Indeed, the availability of structures has led to many hypotheses which were tested by measuring properties of protein-DNA complexes in solution. By probing the consequences of various DNA structural perturbations (Jen-Jacobson, 1995, Jen-Jacobson, 1997) (from subtle removal of a single functional group to a complete base-pair change) on DNA binding and cleavage, correlations between structure and function can be made. This approach was successful in quantifying discrimination of the *Eco*RI endonuclease against non-cognate DNAs and in gaining an appreciation of which properties of the system are important for its sequence selectivity *in vivo*. Finally, determining the thermodynamic signatures of specific, non-specific, and miscognate complexes allowed correlations to be made between the energetics and the available structural data, and informed us about features of the non-specific and miscognate complexes for which crystal structures are not yet available.

1.3.3.1. Energetics of discrimination in binding and catalysis

Lesser et al. measured the equilibrium binding affinities (Table 1.1) and the first order cleavage rate constants (Table 1.2) for each of the nine possible *Eco*RI miscognate sites (Lesser et al., 1990). Binding to *EcoRI* miscognate sites entails significant energetic penalties relative to the interaction with GAATTC ($\Delta\Delta G^\circ$ = 4 to 6 kcal/mol). Interestingly, binding affinity for the 'inverted' site, CTTAAG, where every base-pair is changed relative to GAATTC is comparable to miscognate DNA binding affinity (Table 1.1). These penalties are much more severe than those associated with binding to DNA sites containing more subtle perturbations such as the disruption of a single hydrogen bond or Van der Waals contact at the interface ($\Delta\Delta G^\circ$ = 1 to 2 kcal/mol) (Lesser et al., 1990, Jen-Jacobson, 1995, Jen-Jacobson, 1997). Miscognate sites (for

example $\frac{AAATTC}{TTTAAG}$) each contain a normal $\frac{GAA}{CTT}$ half site, and an altered half site ($\frac{AAA}{TTT}$ in this example).

Table 1.1 Changes in free energies of formation of EcoRI-DNA complexes ($\Delta\Delta G°_{ED}$) and transition state complexes ($\Delta\Delta G°_{I}^{\ddagger}$) for variant DNA sites [a]

Site	K_A * (M^{-1})	$\Delta\Delta G°_{ED}$ § (kcal/mol)	k_1 (sec^{-1})	Relative $(k_1 \times K_A)$ ‖ $(M\text{-}sec)^{-1}$	$\Delta\Delta G°_{I}^{\ddagger}$ (kcal/mol)
GAATT C CTTAA↑G	$3.3(\pm1.6)\times10^{11}$	0	$3.4(\pm0.7)\times10^{-1}$	1	0
		First-base substitutions			
TAATT C ATTAA↑G	$3.1(\pm1.6)\times10^{8}$	4.1±0.5	$5.0(\pm0.3)\times10^{-3}$	1.5×10^{-5}	6.6±0.5
AAATT C TTTAA↑G	$1.8(\pm1.0)\times10^{8}$	4.4±0.5	$8.4(\pm1.7)\times10^{-4}$	1.4×10^{-6}	8.0±0.5
CAATT C GTTAA↑G	$5.3(\pm2.2)\times10^{7}$	5.2±0.4	$2.0(\pm0.7)\times10^{-4}$	1.0×10^{-7}	9.5±0.5
		Second-base substitutions			
GCATT C CGTAA↑G	$4.4(\pm1.7)\times10^{7}$	5.3±0.4	$6.7(\pm0.5)\times10^{-5}$	2.7×10^{-8}	10.3±0.4
GGATT C CCTAA↑G	$2.6(\pm1.7)\times10^{7}$	5.6±0.5	$6.2(\pm1.5)\times10^{-6}$	1.5×10^{-9}	12.0±0.6
GTATT C CATAA↑G	$1.6(\pm0.8)\times10^{8}$	4.5±0.5	~2 ×10^{-7}	2.9×10^{-10}	~13
		Third-base substitutions			
GACTT C CTGAA↑G	$5.9(\pm0.6)\times10^{7}$	5.1±0.3	$4.5(\pm0.6)\times10^{-4}$	2.5×10^{-7}	9.0±0.3
GAGTT C CTCAA↑G	$1.1(\pm0.9)\times10^{8}$	4.7±0.6	$8.9(\pm0.3)\times10^{-5}$	8.9×10^{-8}	9.6±0.3
GATTT C CTAAA↑G	$3.4(\pm0.7)\times10^{7}$	5.4±0.4	$2.0(\pm0.5)\times10^{-6}$	6.2×10^{-10}	12.5±0.4
		Double substitutions¶			
GGATCC	$9.7(\pm4.8)\times10^{7}$	4.8±0.5	No cleavage		
GAGCTC	$4.3(\pm2.0)\times10^{7}$	5.3±0.4	No cleavage		
AAATTA	$2.0(\pm1.0)\times10^{7}$	5.7±0.5	No cleavage		
		Inverted site			
CTTAAG	$4.7(\pm2.6)\times10^{7}$	5.2±0.5	No cleavage		

[a] Table taken from (Lesser et al., 1990)

[*] Equilibrium association constants, K_A were measured in Binding Buffer (Appendix X) plus 95mM NaCl (pH 7.3).

[§] $\Delta\Delta G°_{ED} = -RT \ln (K_{A,\ non\text{-}cognate}/K_{A,\ cognate})$

[‖] Measures discrimination relative to GAATTC

The first order rate constants for cleavage of both the modified and unmodified half sites of miscognate DNAs are dramatically reduced relative to GAATTC (Table 1.2). Thus the effects of the base-pair change are manifested not only in the site containing the perturbation, but also in the distal, cognate half-site. This is in contrast to more subtle DNA perturbations where minor (if any) reductions in cleavage rates are observed in the modified half-sites, and cleavage rates of the cognate half-sites are un-effected. Sites with two or more base-pair changes compared to GAATTC (defined earlier as non-specific DNA) are completely resistant to cleavage.

Table 1.2 Specific and miscognate DNA cleavage rate constants

Site	k_1 (sec^{-1})	k_2 (sec^{-1})	k_3 (sec^{-1})	k_4 (sec^{-1})
GAATTC	$3.4(\pm0.7)\times10^{-1}$	$3.2(\pm0.8)\times10^{-1}$	$2.5(\pm0.9)\times10^{-1}$	$2.2(\pm0.6)\times10^{-1}$
TAATTC	$5.0(\pm0.3)\times10^{-3}$	$2.2(\pm1.5)\times10^{-4}$	$1.4(\pm1.6)\times10^{-5}$	$1.6(\pm1.1)\times10^{-3}$
AAATTC	$8.4(\pm1.7)\times10^{-4}$	$1.5(\pm0.3)\times10^{-4}$	$2.8(\pm0.6)\times10^{-5}$	$3.8(\pm1.7)\times10^{-5}$
CAATTC	$2.0(\pm0.7)\times10^{-4}$	$1.1(\pm0.2)\times10^{-5}$	*	*
GCATTC	$6.7(\pm0.5)\times10^{-5}$	$6.3(\pm1.9)\times10^{-6}$	*	*
GGATTC	$6.2(\pm1.5)\times10^{-6}$	Nil[§]	Nil	NMF[‖]
GTATTC	$\sim2\times10^{-7}$	Nil	Nil	NMF
GACTTC	$4.5(\pm0.6)\times10^{-4}$	$7.2(\pm0.7)\times10^{-4}$	$1.0(\pm1.1)\times10^{-5}$	$1.1(\pm0.9)\times10^{-4}$
GAGTTC	$8.9(\pm0.3)\times10^{-5}$	$9.3(\pm5.9)\times10^{-6}$	*	*
GATTTC	$2.0(\pm0.5)\times10^{-6}$	Nil	Nil	NMF

Table taken from (Lesser et al., 1990)
Naming convention for rate constants describing *miscognate* DNA cleavage are as follows: k_1 is the rate associated with nicking the cognate half-site, k_2 the rate constant for nicking the non-cognate half-site, and k_3 and k_4 are the rate constants for making a cut in the second strand provided that a nick in the cognate and non-cognate half-sites respectively had already occurred.
[§] No product observed
[‖] No meaningful figure; for the G**G**ATTC, G**T**ATTC, and GA**T**TTC sites, no cleavage in the variant half-site was observed. Thus the substrate for the reaction described by k_4 never appears.

The reaction in which binding of free endonuclease to DNA leads to cleavage of both DNA strands is a complex series of steps including reaction intermediates, product release, and even dissociation from a nicked substrate. In addition, the binding constant (K_A) and cleavage rate constant (k_c) may be affected by non-productive binding (complexes that will never reach the transition state) (Fersht, 1985). A metric that avoids the influence of the above obscuring factors is the transition state interaction free energy change, $\Delta G^{\circ}_{I}{}^{\ddagger}$ (Lesser et al., 1990), which is related to the probability of going from the free macromolecules to the activated complex where the first DNA strand is cleaved (Figure 1.8). The transition state interaction free energy change is defined as:

$$\Delta G^{\circ}{}_{I}{}^{\ddagger} = \Delta G^{\circ}{}_{Bind} + \Delta G^{\circ\ddagger} \qquad (1.1)$$

where

$$\Delta G^{\circ}{}_{Bind} = -RT \ln K_A \qquad (1.2)$$

and

$$\Delta G^{\circ\ddagger} = RT \ln\left(\frac{kT}{h}\right) - RT \ln k_c \qquad (1.3)$$

where k_c is the first order cleavage rate constant for the first bond breaking step, k is the Boltzman constant, h is the Planck constant, R is the ideal gas constant, and T the absolute temperature. We can therefore measure the equilibrium binding constants for DNA sites and the rate of cognate half-site cleavage (the cognate half-site is a constant in every miscognate site), which permits us to calculate the transition state interaction discrimination energy between any two substrates (designated by 1 and 2 subscripts below):

$$\Delta\Delta G^{\circ}{}_{I}{}^{\ddagger} = -RT \ln\left[\frac{(k_{c1} * K_{A1})}{(k_{c2} * K_{A2})}\right] \qquad (1.4)$$

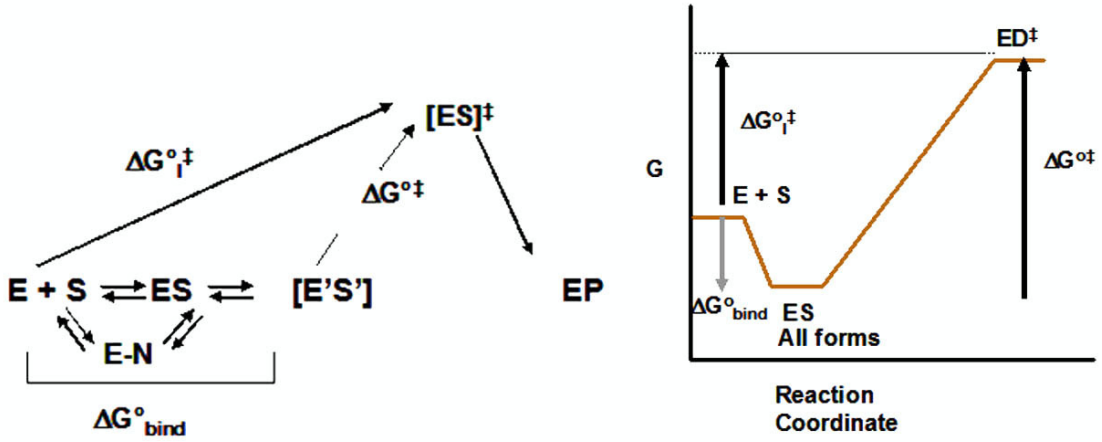


Figure 1.8 Transition state interaction free energy change as a metric for discrimination.

The transition state interaction free energy ($\Delta G^{\circ}_{I}{}^{\ddagger}$) is the probability of forming the activated complex from the free enzyme (E) and the substrate (S), and takes into account both assembly of the enzyme-substrate complex (ΔG°_{bind}), and its activation to the transition state ($\Delta G^{\circ\ddagger}$). See equations 1.1 to 1.4 in the text (section 1.3.3.1) for equations describing $\Delta G^{\circ}_{I}{}^{\ddagger}$.

Single base-pair substitutions result in large unfavorable transition state interaction energies ($\Delta\Delta G^{\circ}_{I}{}^{\ddagger}$ = +6.6 to +13 kcal/mol) relative to GAATTC, reflective of the severe penalties in both miscognate DNA binding and cleavage (Table 1.1). This is in contrast to more subtle perturbations to the GAATTC site (e.g. base analog substitutions that remove a single hydrogen bond at the interface), for which transition state interaction free energy penalties are small ($\Delta\Delta G^{\circ}_{I}{}^{\ddagger}$ = 0 to +2.5 kcal/mol) (Jen-Jacobson, 1995, Jen-Jacobson, 1997), thus discrimination against miscognate sites cannot be explained based solely on disruption of protein base contacts, as a base-pair change would be expected to remove at most, two hydrogen bonds relative to the specific complex, and the isolated disruption of two hydrogen bonds would be predicted to cost about 2 to 3 kcal in transition state interaction free energy. Based on this deduction, Lesser et al. proposed that stringent transition state interaction energy discrimination against miscognate DNAs was due to three interdependent factors: disruption of protein-base contacts ($\Delta\Delta G_{Base}$), disruption of protein-phosphate contacts ($\Delta\Delta G_{Phos}$), and $\Delta\Delta G_{Reorg}$, which includes differences in conformational changes, entropic factors, and the cost of driving variant DNA sites into the kinked conformation (Lesser et al., 1990).

Ethylation interference foot printing (Siebenlist et al., 1980), a probe which tests which phosphates are important for binding, demonstrated a structural corollary to the observed energetic discrimination against miscognate sites. First, ethylation interference patterns for GAATTC, and an isosteric site (G in GAATTC was substituted with 7-deaza-guanosine) were indistinguishable, consistent with the idea that mild perturbations affect only $\Delta\Delta G_{Base}$ (Figure 1.9A&D) (Lesser et al., 1990). Footprints on miscognate sites showed remarkable differences in phosphate contacts compared to the phosphate contacts in the specific complex. Taking the AAATTC miscognate site as an illustrative example, all three phosphate clamps are absent between the enzyme and the strand containing the non-cognate half site (Figure 1.9B). The enzyme adapts toward the cognate half-site where two out of three phosphate clamps are conserved. However the strengths of the N**p**N**p**GAATTC clamps are reduced relative to the specific complex and the central GAA**p**TTC clamp is almost entirely absent in both half-sites of the miscognate complex. There is no localized footprint on non-specific DNA (Lesser et al., 1990). These observations provided direct evidence that hydrogen bonding to the DNA bases is not the only means for sequence discrimination as was originally suggested (Seeman et al., 1976), and that particular phosphate contacts do not serve in merely contributing to the non-specific component of the interaction free energy, but rather play an important role in specificity. Failure to form clamp phosphates also implies that the energy required to achieve the kinked DNA conformation would be higher in miscognate complexes, as disruption of these contacts would lead to a reduction to the stabilization of the distorted DNA conformation, meaning $\Delta\Delta G_{Reorg}$ also contributes to the overall discrimination against miscognate sites.

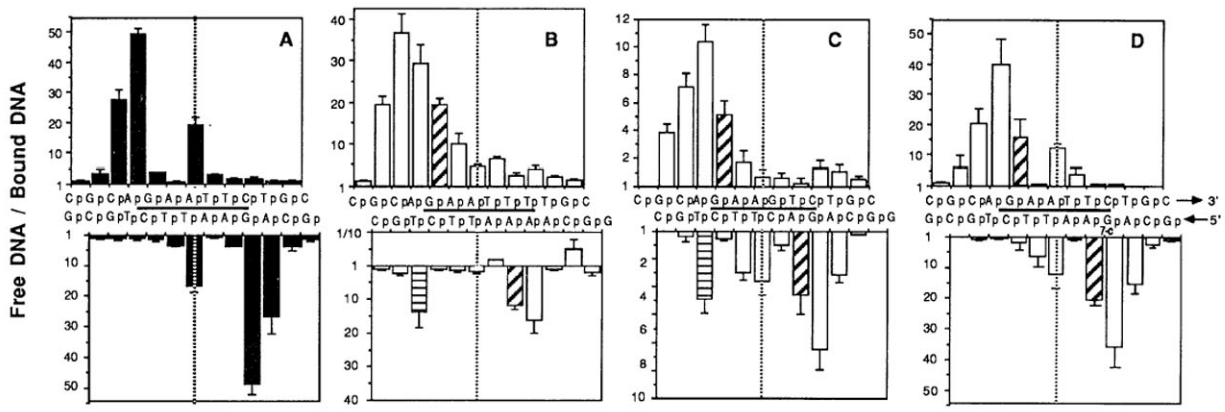


Figure 1.9 Phosphate contacts are important to *Eco*RI sequence discrimination.

Ethylation interference footprints of the *Eco*RI endonuclease binding to the specific site (A), the miscognate sites AAATTC (B) and GACTTC (C), and the specific site with 7-deaza-guanosine substituted for **G**AATTC (D). In panels C-D, the bottom figure represents the footprint on the strand containing the modified half-site. Figure is taken from (Lesser et al., 1990)

1.3.3.2. Discrimination in vivo

The *in vitro* measurements of binding and cleavage discrimination of the *Eco*RI enzyme provided a quantitative understanding of how restriction enzymes achieve discrimination *in vivo*. Lesser et al. proposed that type II restriction endonucleases have evolved multiple fail-safes in order to avoid genomic DNA cleavage *in vivo* (Lesser et al., 1990). The companion methylase provides the first level of protection as the probability of cleavage of the methylated cognate site is severely reduced relative to the unmodified site (Jen-Jacobson et al., 1996). In addition, there is stringent discrimination against binding to miscognate sites and the structural adaptations in such miscognate complexes further reduce the probability of formation of complexes with the precise juxtaposition of enzyme and DNA elements that are required for catalysis. In the event that a nick occurs in one strand (most likely the cognate half-site) of a miscognate site, cleavage of the second DNA strand is extremely unlikely since the dissociation rate constant for miscognate complexes is $>10^5$-fold faster than the rate of non-cognate strand cleavage (David Lesser, unpublished). Thus, potentially lethal double stranded cuts are avoided by this 'kinetic proofreading' mechanism, and the single strand nicks can be sealed by DNA ligase (Heitman et

al., 1989). Finally, binding affinities for non-specific (≤ 4/6 base-pair match, not cleaved) and miscognate DNAs are comparable. Since non-specific DNA sites outnumber *Eco*RI* sites in the *E. coli* genome by a factor of ~450, non-specific DNA is expected to effectively compete with *Eco*RI* sites for endonuclease binding *in vivo*.

1.3.3.3. Three different binding modes: pre-transition state, adaptive and non-specific complexes

The thermodynamic profile of the *Eco*RI interaction with specific, miscognate and non-specific sites delineate three classes of binding modes. I have already discussed an outline of the three binding modes based on binding affinities, site localization, as well as cleavage rates and rate symmetry for cleavage of the two half-sites (Table 1.3). Additional indicators lend support to the existence of three distinct binding modes: Ca^{2+} binds to the active sites of type restriction endonuclease complexes in a manner similar to the biological co-factor, Mg^{2+}, but does promote DNA cleavage (Horton et al., 1998, Viadiu et al., 1998, Horton et al., 2000, Etzkorn et al., 2004). Specific binding is strongly promoted by the presence of Ca^{2+}, or by the protonation of the active site residues that provide the oxygen ligands for metal binding. Compared to specific binding, addition of Ca^{2+} results in an intermediate enhancement of miscognate DNA binding affinity, and miscognate DNA binding is less strongly dependent on pH (Engler et al., 1997, Engler, 1998). For non-specific binding there is little or no pH dependence, and addition of Ca^{2+} in the binding reaction has no effect (Table 1.3). The effects on binding using the Ca^{2+} and pH probes indicate that the active sites of miscognate complexes are less well assembled than their specific counterparts, and the active site of the non-specific complex is not assembled at all.

Perhaps the most informative probe of differences in the three binding modes is the heat capacity change ($\Delta C°_p$) accompanying complex formation. The thermodynamic signature of the *Eco*RI-specific DNA complex is a large and negative heat capacity change for the binding reaction ($\Delta C°_p$= -2.6 kcal/mol K) (Jen-Jacobson, 2000, Grigorescu, 2003) (Table 1.3). A large and negative heat capacity change is the thermodynamic signature for the specific binding mode of other restriction endonucleases (Jen-Jacobson, 2000) and many other highly specific DNA binding proteins such as *Cro* (Takeda et al., 1992), *Trp* (Ladbury et al., 1994), *Lac* (Frank et al.,

1997), and λcI (Merabet et al., 1995) repressors among others. All of the features that lead to a tight complementary interface (release of water from non-polar surfaces, trapping of waters that mediate protein-DNA interactions, and the restriction of the vibrational motions of interfacial protein and DNA groups) also contribute to a negative $\Delta C°_p$ for the binding reaction (Sturtevant, 1977). This is in contrast to the *Eco*RI-non-specific DNA binding reaction where the $\Delta C°_p$ is ~0. Studies by Dengfeng Cao demonstrated that a significant contributor to the differences in heat capacity change between the specific and non-specific DNA binding reactions is a severe decrease in the amount of water released from the free macromolecules upon non-specific complex formation as compared to specific binding (Cao, 2001) (Table 1.3). A zero heat capacity change for the non-specific DNA binding reaction also suggests that folding of the arms of the *EcoRI* endonuclease (section 1.3.2.4) which is coupled to specific binding (Grigorescu, 2003), is not coupled to non-specific DNA binding. First, the coupled folding reaction likely buries a significant amount of non-polar surface area, which would lead to a negative (non-zero) heat capacity change for non-specific complex formation if the folding transition did occur. Second, theoretical studies suggest that any mandatory coupling of equilibria in a binding process will contribute to a *negative* heat capacity change for that overall binding process (Eftink et al., 1983). Finally, a zero heat capacity change for non-specific complex formation indicates a lack of vibrational restriction in the non-specific complex. These ideas argue that the *Eco*RI-non-specific DNA complex is looser, and lacks the surface complementarity of the specific complex in part because the requisite conformational changes in the protein and DNA are not stabilized in the non-specific complex.

Binding of the *Eco*RI endonuclease to miscognate DNAs is accompanied by heat capacity changes that are intermediate between the values obtained for specific and non-specific complex formation (Jen-Jacobson, 2000, Grigorescu, 2003) (Table 1.3). The implication is that the magnitude of $\Delta C°_p$ for miscognate complex formation is much smaller than the $\Delta C°_p$ for specific binding because the factors that contribute to a negative $\Delta C°_p$ (release of water from non-polar surface, immobilization of interfacial water, restriction of vibrational modes) are all reduced in the "adaptive" miscognate complex. That $\Delta C°_p$ values for miscognate binding are intermediate between non-specific ($\Delta C°_p = 0$) and specific binding ($\Delta C°_p = -2.6$ kcal/mol K) argues that miscognate and non-specific complexes are structurally distinct, although binding affinities for the two classes of sites are similar.

Table 1.3 Discrimination at the *Eco*RI endonuclease-DNA interface: Three different binding modes [a]

	Specific Complex	Adaptive Miscognate Complex	Non-specific Complex
DNA Site	GAATTC	One incorrect base-pair	≥ 2 incorrect base-pairs
Specificity Ratio [b]	1	1.3×10^5	2.5×10^6
Site Localization [c] (Footprinting)	Strong	Moderate and asymmetric	None
pH effect [d]	Strong	Weak	Very weak
Ca^{2+} effect [e]	Strong	Weak	None
Salt dependence [f]	Steep	Intermediate	Very Steep
Heat capacity change ($\Delta C°_p$) [g]	Large and negative -2.6 kcal/mol K	Intermediate -0.1 to 1.0 kcal/mol K	Zero
Number of Waters released [h]	430	280-380	50
DNA conformation [i]	Symmetric Distortion	Asymmetric distortion?	None?
Cleavage of DNA strands [j]	Symmetric rates	Reduced and asymmetric rates	None

[a] Trends pertain qualitatively to the *Eco*RI, *Bam*HI and *Eco*RV restriction endonucleases. Exact numbers are for the *Eco*RI endonuclease unless otherwise noted.

[b] Specificit y Ratio $= \frac{K_{Specific}}{K_{Non-cognate}}$ from (Jen-Jacobson, 1997)

[c] (Lesser et al., 1990)

[d] Data not available for *Eco*RI, data are from the *Eco*RV system (Engler et al., 1997)

[e] (Kurpiewski et al., 2004) and this work

[f] (David R. Lesser, unpublished)

[g] (Jen-Jacobson, 2000, Grigorescu, 2003)

[h] Number of waters released from free macromolecules upon complex formation (Cao, 2001)

[i] These are predictions; there is not direct evidence for these claims

[j] (Lesser et al., 1990)

1.4. Attempts at changing the specificity of type II restriction endonucleases

The study of restriction endonucleases is motivated by interests that are both academic and applied in nature. On one hand restriction endonucleases were chosen as model systems to understand the rules governing specificity of DNA binding proteins because the enzymes are among the most highly specific, and are also extremely diverse in their recognition sequences. Restriction endonucleases with their hundreds of different sequence specificities represent fertile ground for the illumination of many unique specificity principles. On the other hand, the identification, cloning, and characterization of restriction endonucleases were critical in the development of recombinant DNA technologies. Indeed many studies of restriction endonucleases are carried out with the hope that the studies will bear fruit leading to a wider array of applications in laboratory and medicinal science.

Although there are a total of 588 commercially available type II restriction endonucleases representing 211 distinct specificities (Roberts et al., 2003), it remains a goal to be able to change the specificity of existing enzymes. The principle motivation for achieving this goal is the expectation that novel applications will follow. For example, most restriction endonucleases recognize sites that are 4 or 6 base-pairs in length. These sites appear too frequently within large genomes to be useful in their manipulation. There are some rare cutting (recognition sites of 8 or 10 bp) restriction endonucleases, but these enzymes are themselves rare. Thus, it would be extremely useful to be able to either extend the length of the recognition site of an enzyme that normally cleaves a 6 base-pair recognition site, or change the specificity of an existing enzyme that recognizes an eight or ten base-pair recognition sequence. One can also imagine that it would be useful to change the specificity of existing endonucleases from extremophilic bacteria. A wider range of specificities conferred by heat tolerant restriction endonuclease for example, would facilitate the development of novel laboratory technologies.

Attempts to change the specificity of an existing restriction endonuclease have been uniformly unsuccessful. Exhaustive mutation of the amino acids directly involved in DNA recognition has not yielded enzyme variants with altered specificity, but rather loss of function mutants that have lost the ability to cleave the cognate site with no concomitant gain in miscognate DNA cleavage activity (Alves et al., 1989, Needels et al., 1989, Heitman et al., 1990, Osuna et al., 1991, Fritz et al., 1998, Ivanenko et al., 1998, Dorner et al., 1999, Zhu et al., 2003).

These failures outline the difficulty of the problem, and underscore the idea that recognition is not as simple as being defined by the sum of one-to-one interactions between protein side chains and DNA bases. The principles of specificity described in the previous sections of this introduction also support this idea, and suggest that changing the specificity of the *Eco*RI restriction endonuclease from GAATTC to some other site would require at least three types of amino acid changes: (a) those affecting direct readout of the bases in the target sequence; (b) changes stabilizing sequence dependent features of the complex that will be required for efficient catalysis of the new recognition sequence; (c) changes that will couple conformational changes to recognition of the new sequence. These classes of residues are no more than partially overlapping, so that three or more substitutions are likely to be required before a true novel specificity is generated. Further, it is impossible to predict in general how the enzyme might behave at various intermediate points in a multi-step sequence of mutations. Prior mutational inactivation of the endonuclease catalytic center (Rimseliene et al., 2003) or a parallel change in the specificity of the methylase might be required for safety. This set of requirements is certainly not unique to the *Eco*RI system, as existing information on other type II endonucleases suggests the same kinds of constraints.

1.5. Promiscuous mutants of the EcoRI endonuclease: tools to further our understanding of molecular specificity

As an alternative approach to site-directed mutagenesis, Heitman and Model (Heitman et al., 1990) screened for relaxed-specificity point mutations of *Eco*RI endonuclease through the ability to induce the SOS DNA-damage response in *E. coli* despite co-expression of the *Eco*RI methylase. The amino acid changes identified in this screen were A138T, A138V, E192K, H114Y and Y193H. Semi-quantitative experiments with the purified 'promiscuous' mutant enzymes showed that they make double-strand cuts at *Eco*RI* sites in plasmid and bacteriophage DNA more readily than does the wild-type enzyme. This preliminary characterization also suggested that the promiscuous enzymes retained wild-type levels of activity towards the canonical GAATTC recognition site (Heitman and Model, 1990). These observations

demonstrate that the mutants represent cases of relaxed specificity, and support the idea that single amino-acid changes are insufficient to *change* the specificity of a restriction endonuclease.

This thesis is concerned with the thermodynamic, kinetic, and structural characterization of the promiscuous mutants of the *Eco*RI endonuclease. Our motivation for studying these mutant enzymes and their complexes with specific, miscognate and nonspecific DNAs is as follows: 1) Since the mutants enzymes were identified based their ability to induce the SOS response in *E. coli* cells, characterization of these mutant enzymes will likely illuminate principles that result in the stringent *in vivo* specificity of restriction enzymes that is required for their function. 2) Studying the promiscuous mutants may provide additional knowledge of the molecular features of the *Eco*RI restriction enzyme that lead to its remarkable specificity. For example, each of the amino-acid residues implicated by the Heitman et al. screen lies within the arms of the *Eco*RI endonuclease (Glu192, Tyr193), or at the junction between the arms and the main domain of the enzyme (Figure 1.10). Therefore, studying these mutants might provide insight into the molecular details of the disorder to order transition of the arms (section 1.3.2.4), and the role of this structural transition in the mechanism of specificity. 3) Understanding how specificity is relaxed in a highly specific system such as *Eco*RI could help us understand the ways in which non-specific and less stringent DNA binding proteins recognize DNA. 4) Finally, it is our experience that rules gleaned from studying the *Eco*RI endonuclease are usually applicable to other restriction endonucleases and site-specific DNA binding proteins.

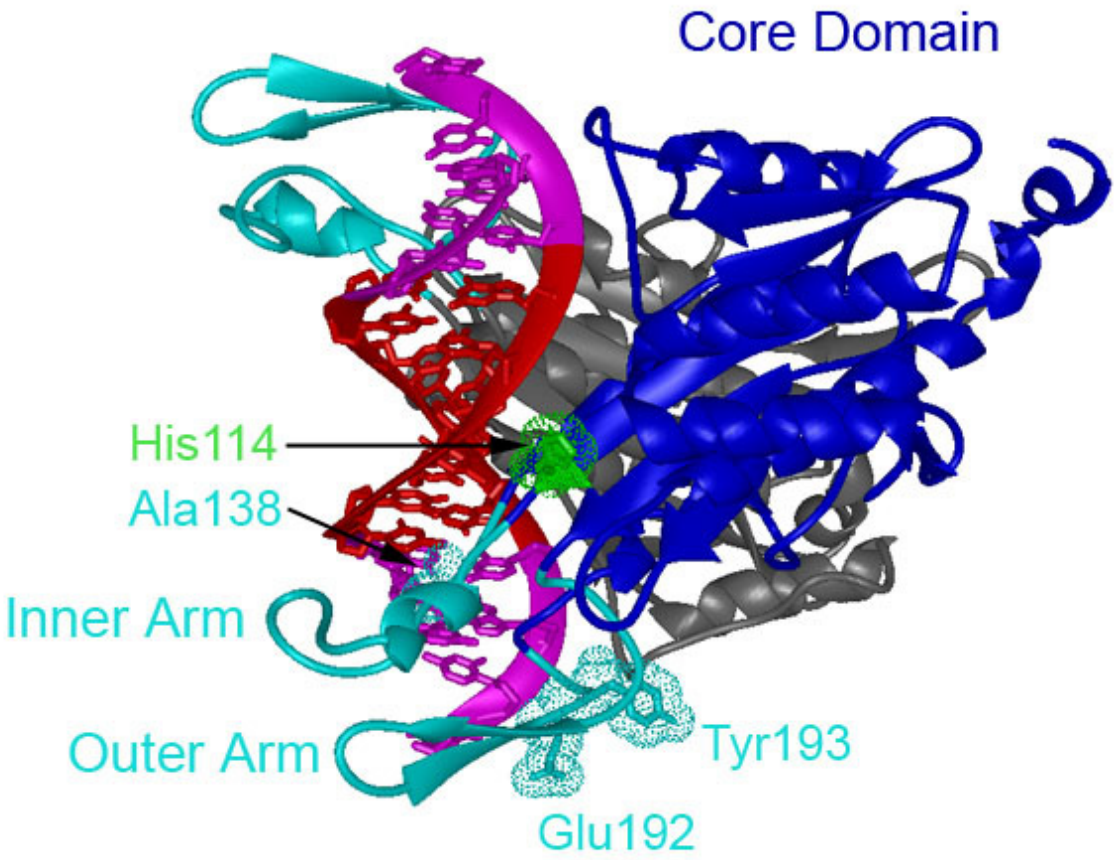


Figure 1.10 Location of the "promiscuous" mutations in the *Eco*RI-TCGCGAATTCGCG complex.

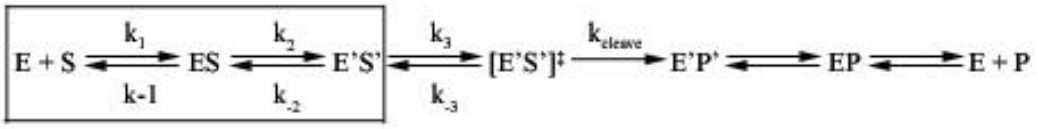
The GAATTC site is in red, and the flanking DNA regions in magenta. Side chains implicated in relaxed specificity are highlighted by van der Waals spheres in one of the two symmetrical subunits (other subunit in gray). Each of the mutations is located in or near the inner or outer arms (cyan), which undergo a disorder to order transition upon specific DNA binding (Grigorescu, 2003).

2. PROMISCUOUS *Eco*RI ENDONUCLEASE MUTANTS AND THE SPECIFIC DNA INTERACTION

2.1. Introduction

2.1.1. Specific complex probed in solution studies is a 'Pre-Transition-State' complex

To investigate the *Eco*RI-specific DNA binding interaction, we measure the equilibrium association constants for purified *Eco*RI endonucleases binding to a 24 base-pair oligonucleotide containing the GAATTC site. Cleavage is prevented by omitting the metal co-factor, thus restricting the reaction to the boxed portion of scheme 2.1 below.



2.1

Formation of the specific complex proceeds through an intermediate, ES, believed to be the complex that undergoes linear diffusion along the DNA, and having structural and energetic features of a 'non-specific' complex (Jack et al., 1982, Terry et al., 1985). The rate of formation of this collision complex has been shown to be diffusion controlled (Jack et al., 1982). When the conformations of the free macromolecules are compared with the protein and DNA conformations observed in the crystal structure of the specific complex, it is clear that conformational changes have occurred in both the DNA and enzyme molecules (Shui et al., 1998, Grigorescu, 2003). Indeed, even binding to a specific recognition site embedded in a short oligonucleotide containing relatively few non-specific sites proceeds through an intermediate as demonstrated by the observation that the slope of the salt dependence of the apparent association

rate (d log $k_a^{app.}$/d log [salt]) is greater than 1 (Lohman et al., 1978, Record et al., 1991, Jen-Jacobson, 1997), which suggests that k_2 in scheme 2.1 above represents a rate limiting isomerization step. Several lines of evidence indicate that our equilibrium binding measurements are not scoring the collision complex, but rather a pre-transition state complex, (where the protein and DNA molecules have undergone the requisite conformational changes required to achieve DNA cleavage) and it is this pre-transition state complex that is visualized in the crystal structure solved in the absence of divalent metal:

1) It has been shown that many subtle perturbations to the DNA side of the *Eco*RI-specific DNA interface (i.e. base analogs that remove single intermolecular hydrogen bonds) exact binding penalties compared to the interactions with the unmodified DNA, yet no further penalties are incurred in the transition state (Cleavage rates of the unmodified and modified DNA sites are the same), suggesting the same interactions exist in both the ground and transition states (Jen-Jacobson, 1995, Jen-Jacobson, 1997).

2) The magnitudes and additivities of the binding penalties associated with a large number of DNA perturbations to the GAATTC site have been interpretable based on the contacts observed in the crystal structure.

3) The structure of a post-reactive complex (Mn^{2+} soaked into the lattice and cleavage occurred '*in crystallo*', pdbid 1QPS) retains all features of the metal-free complex with the exceptions of breakage of the scissile bond and the rotations of active sight residues that participate in coordination of the Mn^{2+} ion.

4) Computational studies in which the crystallographic coordinates of the metal-free complex were used as the starting point in molecular dynamics simulations of the specific complex with the computationally added Mg^{2+} co-factor revealed no-significant changes in the protein-DNA interface beyond local side-chain and solvent rearrangements within the active site (Kurpiewski et al., 2004).

Taken together these observations suggest that we can make structural perturbations to the system, measure the effects of the perturbations on DNA binding and cleavage, and interpret the measurements in solution in the context of the crystal structure of the metal free complex. These structure function studies have resulted in a greater understanding of the system than would have been possible based on inspection of the crystallographic coordinates alone.

2.1.2. Relating binding affinities to free energy changes

To quantify binding affinity, we determine the equilibrium association constant, K_A, calculated from the common expression in equation 2.2:

$$K_A = \frac{[ES]}{[E][S]} \qquad 2.2$$

The observed binding constant is an apparent binding constant with contributions from both specific binding and binding to non-specific sites within the deoxyribonucleotide (von Hippel et al., 1974, Clore et al., 1982). This observed binding constant can be defined as:

$$K_{Obs.} = K_{Sp.} + N_{NS} * K_{NS} \qquad 2.3$$

where $K_{Sp.}$ is the binding constant for specific complex formation, K_{NS} the binding constant for non-specific complex formation, and N_{NS} the number of non-specific sites within the oligomer. For specific DNA, $K_{Obs.}$ is essentially equal to $K_{Sp.,}$ since $K_{Sp.} >> K_{NS}$. The stability of the complex can then be described in terms of the standard Gibbs free energy change:

$$\Delta G^\circ = -RT \ln K_{Obs.} \qquad 2.4$$

We can also use a thermodynamic parameter to express the difference in binding affinities between a mutant and the wild-type enzyme for example, by taking the difference between the standard Gibbs free energy changes for the formation of each complex (equation 2.5). This equation can also be used to express the difference in binding free energy for one enzyme binding to different DNA sites.

$$\Delta\Delta G^\circ = (\Delta G^\circ_{Mut} - \Delta G^\circ_{WT}) = -RT \ln \frac{K_{Mut}}{K_{WT}}.$$

Or : 2.5

$$\Delta\Delta G^\circ = (\Delta G^\circ_{Site1} - \Delta G^\circ_{Site2}) = -RT \ln \frac{K_{Site1}}{K_{Site2}}.$$

2.1.3. Relating kinetic stability of EcoRI-specific DNA complexes to their thermodynamic stabilities

For the assessment of the effects of structural perturbations, measurements of the kinetics of association and dissociation of protein-DNA complexes are complementary to measurements of equilibrium stability. The dissociation rate constant, $k_{d,obs.}$ can be related to a free energy of activation for complex dissociation, $\Delta G^\circ_d{}^\ddagger$, by the equation:

$$\Delta G^\circ_d{}^\ddagger = -RT \ln \frac{k_{d,obs}}{A} \qquad 2.6$$

where *A* is a factor that specifies the frequency of decomposition of the hypothetical transition state intermediate. There is currently no accepted way to determine the value of the frequency factor for non-covalent transition states. It may be reasonable to assume for minor structural perturbations that the values of *A* are the same for the modified and unmodified complexes, such that the difference in activation free energies for complex dissociation can be expressed as (Jen-Jacobson, 1995):

$$\Delta\Delta G^\circ_d{}^\ddagger = \Delta G^\circ_d{}^\ddagger(unmodified) - \Delta G^\circ_d{}^\ddagger(modified) = -RT \ln \frac{k_{d,obs.}(mod)}{k_{d,obs.}(unmod)} \qquad 2.7$$

This assumption has been experimentally verified in studies where either the *Eco*RI restriction enzyme or its DNA recognition site is perturbed. In such cases, a change in the equilibrium

association constant, $K_{A,obs.}$ is accounted for by a change in $k_{d,obs}$, such that $\Delta\Delta G°_{bind} = -\Delta\Delta G°_{d}^{\ddagger}$ (Jen-Jacobson, 1995). This implies that differences in the kinetic stability of *Eco*RI specific complexes can be equated to differences in thermodynamic stability. The data also imply that these perturbations do not affect the rate limiting association step, $k_{a,calc}$, where $k_{a,calc} = K_{A,obs}/k_{d,obs}$.

2.1.4. Relating DNA cleavage rate constants to activation free energies

To measure differences in activation free energy, we determine the first order rate constant for DNA cleavage, k_{cleave} (see Materials and Methods, section 6.8 for details). The difference in activation free energy for DNA cleavage between the wild-type and mutant enzymes for example is calculated as:

$$\Delta\Delta G^{\circ\ddagger} = -RT \ln (k_{cleave}^{Mutant}/k_{cleave}^{Wild\text{-}type}) \qquad 2.8$$

2.2. Results and Discussion

2.2.1. Design of oligonucleotide substrates

All experiments utilized variants of a synthetic 24 base-pair oligonucleotide that carries the cognate GAATTC site off-center, so that cleavage in each of the strands results in distinguishable products. The parent sequence was GGGCGGGXXXGAATTCYYYGGCGC (recognition site underlined) in which XXX/YYY denote the flanking triplets which were varied in some experiments. The flanking context of the specific recognition site has been shown to dramatically affect binding affinity of the *Eco*RI restriction endonuclease (Jen-Jacobson, 1997) and other restriction enzymes (Engler et al., 1997, Engler et al., 2001). The base-pair triplet is the dominant modulator of specific DNA binding affinity for the *Eco*RI (David Chi, Michael

Kurpiewski, Arabela Grigorescu and Linda Jen-Jacobson, *unpublished*), and *Bam*HI (Daniel Lai, Lisa Engler and Linda Jen-Jacobson, *unpublished*) endonucleases, as changing the length of DNAs beyond 12 base-pairs (recognition site in the center), or the identity of bases further than three base-pairs away from the recognition site has no effect on binding. Thus, to probe the effect of flanking context on *Eco*RI endonuclease DNA binding and cleavage, we utilize DNA substrates in which the GAATTC site is flanked by symmetrical base-pair triplets. For a complete description of the oligonucleotides used in this work, see Appendix A.

2.2.2. Promiscuous mutations improve binding to specific DNA

A naïve hypothesis for 'promiscuous' specificity of the mutant proteins is that they have reduced binding affinity for the specific recognition site. On the contrary, equilibrium binding experiments in the absence of divalent metals (Table 2.1) show that the binding of mutant enzymes to the GAATTC recognition site is more favorable than wild-type binding by 10- to 520-fold ($\Delta\Delta G°_{bind}$= -1.3 to -3.7 kcal/mol), depending on the sequence surrounding GAATTC (Table 2.1). Even this simple measure, however, must take into account the sequence context surrounding the recognition site, because previous data have shown (Jen-Jacobson, 1997) that the three base pairs abutting the specific recognition sequence modulate the free energy of binding ($\Delta G°_{bind}$) of the wild-type protein over a range of 3.7 kcal/mol, corresponding to a 560-fold difference in affinities. To determine if the promiscuous mutant proteins retain the wild-type preferences for particular flanking contexts, as well as the quantitative distinctions among these contexts, we measured binding with a series of oligonucleotides containing different sequences flanking the GAATTC site. Among the flanking triplets tested were those found to be the 'best' (CGC triplet) and 'worst' (AAA triplet) for the wild-type enzyme (M.R. Kurpiewski & L. Jen-Jacobson, *unpublished*). As shown in Table 2.1, the wild-type and mutant endonucleases have generally similar ranges of affinities for specific DNA sites in different contexts. However, while the order of specific flanking sequence preference is the same for the wild-type, H114Y and E192K enzymes, the flanking sequence hierarchy for the A138T mutant has some rearrangements (Figure 2.1). Notably, the wild-type enzyme prefers the best triplet (CGC) over GGT by 12-fold, but the A138T enzyme does not discriminate between CGC and GGT. In addition, wild-type enzyme incurs significant penalties for binding to GAT or CCC flanking

triplets (K_A is reduced 56-fold and 230-fold, respectively) whereas the A138T enzyme incurs smaller penalties for the same triplets (K_A reduced 4-fold and 6-fold, respectively). I discuss the possible structural and energetic origins of the altered flanking sequence preference exhibited by the A138T enzyme in Chapter 3. Nevertheless, the significant fact is that each of the three mutant enzymes shows improved binding to GAATTC in any given flanking context.

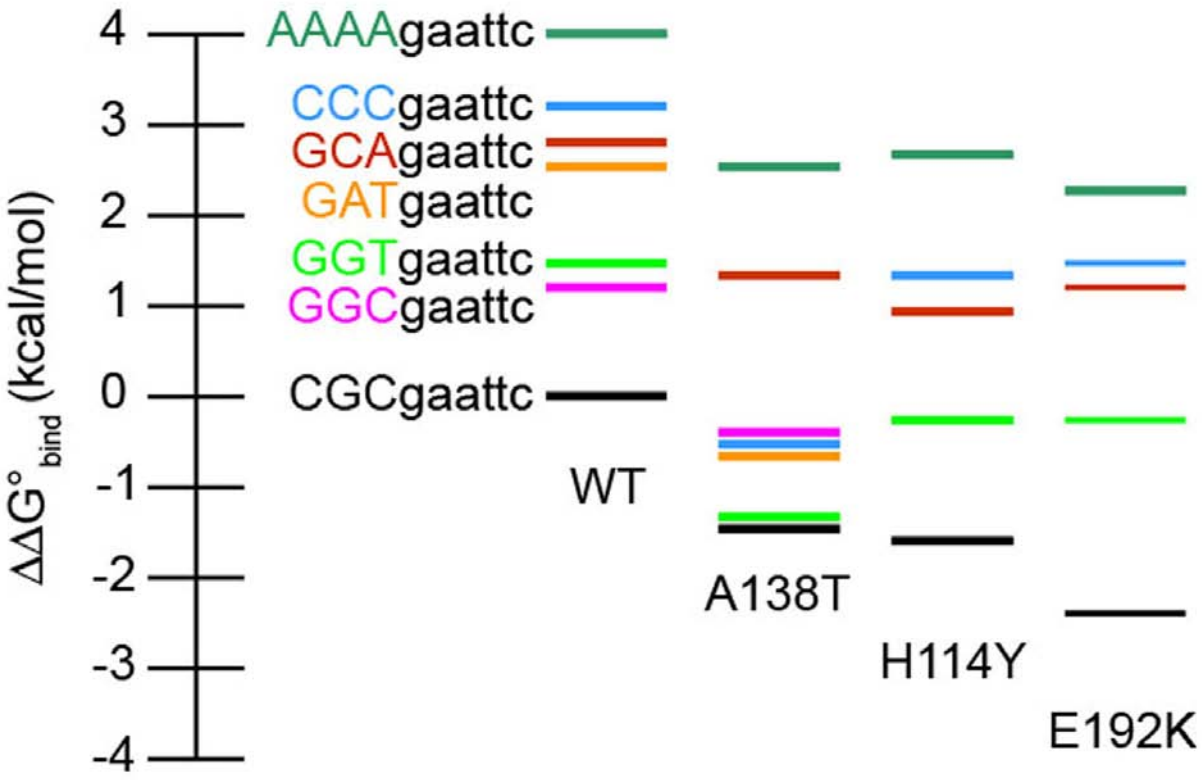


Figure 2.1 Binding of *Eco*RI endonuclease to the specific DNA site in different flanking contexts.

Scale at left has its origin ($\Delta\Delta G°_{bind}$ = 0) at the wild-type interaction with its most favored specific sequence. Similar flanking triplets have bars of the same color in each enzyme's column. Note that only the three bases 5' of the recognition site are shown, however, the three flanking sequences are symmetrical in both strands (see Table 2.1 notes or Appendix A for exact sequences). The wild-type, H114Y and E192K enzymes have the same hierarchy of flanking preferences, but the A138T mutation results in the modification of this hierarchy.

Table 2.1 Effect of flanking sequence on specific DNA binding *Eco*RI endonuclease

Specific Sequence [a]	K_A [b] (M^{-1})	Fold Increase $K_{A\ Mut.}/K_{A\ WT}$	$\Delta\Delta G°_{Bind\ (Context)}$ [c] (kcal/mol)	$\Delta\Delta G°_{Bind\ (Mut.)}$ [d] (kcal/mol)
Wild-type				
CGCgaattcGCG	9.6 (±0.3)x10^9	-	0	0
GGCgaattcGCC	1.3 (±0.4)x10^9	-	1.2 ±0.2	0
GGTgaattcACC	7.9 (±0.9)x10^8	-	1.5 ±0.1	0
GATgaattcATC	1.7 (±0.1)x10^8	-	2.4 ±0.1	0
GCAgaattcTGC	8.1 (±2.0)x10^7	-	2.8 ±0.1	0
CCCgaattcGGG	4.2 (±0.5)x10^7	-	3.2 ±0.1	0
AAAAgaattcTTTT	1.1 (±0.1)x10^7	-	4.0 ±0.1	0
A138T				
CGCgaattcGCG	1.3 (±0.1)x10^{11}	14	0	-1.5 ±0.1
GGTgaattcACC	1.2 (±0.1)x10^{11}	150	0.1 ±0.1	-2.9 ±0.1
GATgaattcATC	3.6 (±0.5)x10^{10}	210	0.7 ±0.1	-3.1 ±0.1
CCCgaattcGGG	2.2 (±0.1)x10^{10}	520	1.0 ±0.1	-3.7 ±0.1
GGCgaattcGCC	1.8 (±0.1)x10^{10}	14	1.1 ±0.1	-1.5 ±0.1
GCAgaattcTGC	1.1 (±0.1)x10^9	14	2.8 ±0.1	-1.5 ±0.1
AAAAgaattcTTTT	1.1 (±0.2)x10^8	10	4.0 ±0.1	-1.3 ±0.1
H114Y				
CGCgaattcGCG	1.6 (±0.1)x10^{11}	17	0	-1.7 ±0.1
GGTgaattcACC	1.4 (±0.1)x10^{10}	18	1.4 ±0.1	-1.7 ±0.1
GCAgaattcTGC	2.0 (±0.1)x10^9	25	2.6 ±0.1	-1.9 ±0.2
CCCgaattcGGG	1.0 (±0.1)x10^9	24	3.0 ±0.1	-1.9 ±0.1
AAAAgaattcTTTT	1.1 (±0.1)x10^8	10	4.3 ±0.1	-1.4 ±0.1
E192K				
CGCgaattcGCG	4.5 (±0.1)x10^{11}	47	0	-2.3 ±0.1
GGTgaattcACC	1.4 (±0.2)x10^{10}	18	2.0 ±0.1	-1.7 ±0.1
GCAgaattcTGC	1.4 (±0.1)x10^9	17	3.4 ±0.1	-1.7 ±0.2
CCCgaattcGGG	7.4 (±0.6)x10^8	18	3.7 ±0.1	-1.7 ±0.1
AAAAgaattcTTTT	2.1 (±0.1)x10^8	19	4.5 ±0.1	-1.7 ±0.1

[a] The specific recognition site and its immediate flanking regions are shown. The entire sequence is5'-GGGCGGGXXXGAATTCYYYGGCGC, where XXX and YYY denote the flanking triplets. The flanking AAAA tract is the only case where the fourth base away from the GAATTC site has an influence. Sequences are listed in order of best to worst flanking contexts for each enzyme. Note that some triplets that were tested with the wild-type and A138T endonucleases were not tested with the H114Y and E192K endonucleases.

[b] Equilibrium association constants were measured in binding buffer plus 0.22 M KCl (pH 7.3, 21°C); means ± std. dev. of ≥ 3 determinations for each sequence.

[c] For each *enzyme*, the difference in the observed standard binding free energy between the best specific flanking sequence (reference) and each additional flanking sequence is calculated as:

$$\Delta\Delta G°_{bind\ (Context)} = -RT \ln (K_A^{Flank}/K_A^{Best\ Flank}) \text{ at } 294 \text{ K.}$$

[d] For each *flanking sequence*, the difference in the observed standard binding free energy between the wild-type (reference) and mutant enzyme is calculated as:

$$\Delta\Delta G°_{bind\ (Mutant)} = -RT \ln (K_A^{Mutant}/K_A^{Wild\text{-}type}) \text{ at } 294 \text{ K.}$$

2.2.3. Improved mutant specific DNA binding free energy is utilized efficiently in the transition state

The classical view of catalysis holds that the catalytic power of an enzyme is strictly derived from stabilization of the transition state. Tight binding by an enzyme to its substrate was regarded as anti-catalytic since a deeper free energy well for the ES complex would create a larger activation free energy barrier, $\Delta G^{\circ\ddagger}$ between the ES and $ES^{\ddagger}$ complexes, thereby reducing the rate of reaction (Schowen, 1978). Thus, specificity (tight binding to the correct substrate) and increased catalytic power appeared to be contradictory objectives in enzyme design. A resolution of this contradiction was provided by Albery and Knowles in their 'uniform binding' postulate, which states that tight binding *and* catalytic efficiency are attainable if enzyme-substrate complexes contain interactions that contribute equally to the stability of the ground and transition states (Albery et al., 1976). It was previously shown that the specific DNA cleavage rate is insensitive to flanking sequence (Jen-Jacobson, 1997) (Figure 2.2A); this is also true for the promiscuous A138T protein (Figure 2.2B). Figure 2.3 contains reaction coordinate diagrams showing the effects of protein or DNA perturbations on binding and cleavage. The perturbation may result in less favorable ($\Delta\Delta G^{\circ}_{Bind} > 0$), or more favorable ($\Delta\Delta G^{\circ}_{Bind} < 0$) binding, depending on the modification. In either case, the free energy of activation ($\Delta G^{\circ\ddagger}$) may increase, decrease or remain the same. Modifying the specific DNA flanking sequence to a less than optimal triplet is representative of the case where the free energies of the E'S' complex and the transition state are raised by similar amounts (case depicted in Figure 2.3B). A structural interpretation of these data is that changing the GAATTC flanking sequence affects the free energy of complex formation, yet the final conformations of the different complexes are very similar, such that the rates of DNA cleavage are unaffected.

We asked if the relaxed if the promiscuous mutations represent cases of uniform binding mutations (Albery and Knowles, 1976), such that the improved binding free energy is efficiently utilized in the stabilization of the transition state. Single turnover cleavage reactions were conducted by equilibrating protein and DNA (without Mg^{2+}) at concentrations such that all DNA is protein bound, then initiating reactions with Mg^{2+}. As substrate, we used the specific GAATTC site embedded in the CGC flanking context, which is most preferred in binding (Table 2.1). Data in Table 2.2 show that the H114Y and E192K mutant enzymes cleave both half-sites

of the specific substrate at the same rates as the wild-type enzyme, within experimental error. Therefore the H114Y and E192K amino-acid changes represent the first reported cases of uniform binding mutations, which lower the free energy of the ES and $ES^{\ddagger}$ complexes to the same extent (case depicted in Figure 2.3E). By contrast, the A138T enzyme exhibits a small (~50%) but significant increase (compared to wild-type enzyme) in the rates of GAATTC cleavage in both DNA half-sites (case depicted in Figure 2.3F). This suggests that the A138T enzyme stabilizes the transition-complex better than does the wild-type enzyme.

Table 2.2 Effect of *Eco*RI endonuclease mutations on specific a DNA binding and cleavage

Enzyme	K_A [a] (M^{-1})	$\Delta\Delta G°_{bind}$ [b] (kcal/mol)	$k_{1,cleave}$ [c] (s-1)	$\Delta\Delta G°_1{}^{\ddagger}$ [d] (kcal/mol	$k_{2,cleave}$ [c] (s^{-1})	$\Delta\Delta G°_2{}^{\ddagger}$ [d] (kcal/mol
Wild-type	9.6 (± 0.3)x10^9	0	0.27 ±0.01	0	0.29 ±0.04	0
A138T	1.3(± 0.1)x10^{11}	-1.5 ±0.1	0.42 ±0.04	-0.3 ±0.1	0.41 ±0.02	-0.2 ±0.1
H114Y	1.6(± 0.1)x10^{11}	-1.6 ±0.1	0.30 ±0.02	-0.1 ±0.1	0.31 ±0.02	-0.1 ±0.1
E192K	4.5(± 0.1)x10^{11}	-2.3 ±0.1	0.23 ±0.00	0.1 ±0.1	0.26 ±0.00	0.1 ±0.1

[a] Sequence of double stranded 24 base-pair oligonucleotide is 5'-GGGCGGGCGCGAATTCGCGGGCGC (recognition site underlined). Equilibrium association constants were measured in binding buffer plus 0.22 M KCl (pH 7.3, 21°C); means ± std. dev. of ≥ 3 determinations for each enzyme.

[b] Difference in the observed standard binding free energy between the reference wild-type enzyme (line 1) and each mutant; calculated as

$$\Delta\Delta G°_{bind} = -RT \ln (K_A{}^{Mutant}/K_A{}^{Wild\text{-}type}) \text{ at } 294 \text{ K.}$$

[c] Means ± SD of ≥ 3 determinations; first-order strand scission constants were obtained under single-turnover conditions using 8 mM Mg2+ to initiate the reaction; k_2 represents cleavage of strand indicated in footnote (a); k_1 represents cleavage of the complementary strand.

[e] Difference in activation free energy for cleavage between the wild-type enzyme (line 1) and each promiscuous mutant; calculated as

$$\Delta\Delta G°_1{}^{\ddagger} = -RT \ln (k_1{}^{Mutant}/k_1{}^{Wild\text{-}type}) \text{ or } \Delta\Delta G°_2{}^{\ddagger} = -RT \ln (k_2{}^{Mutant}/k_2{}^{Wild\text{-}type})$$

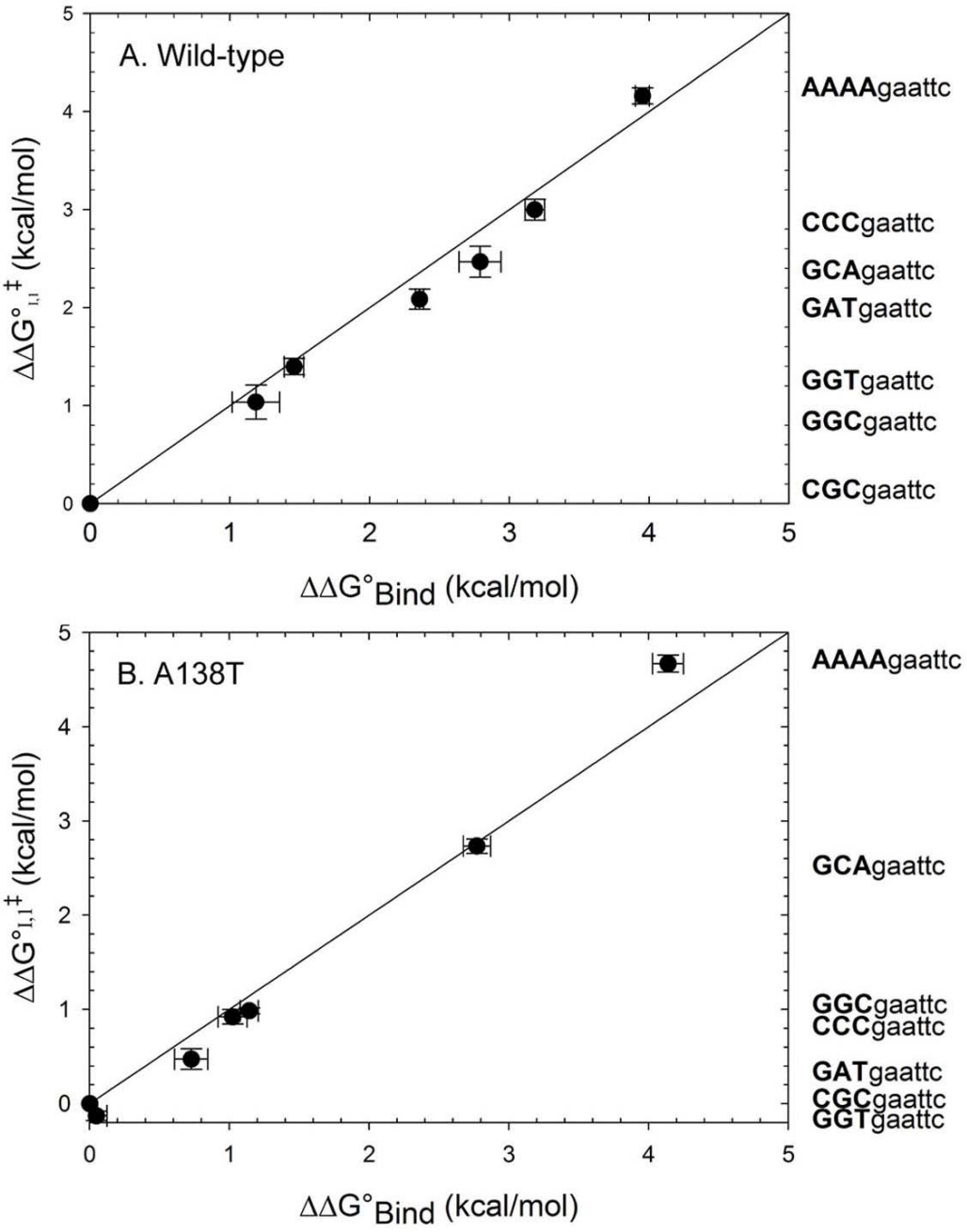


Figure 2.2 Effect of flanking sequence on *Eco*RI specific DNA binding and cleavage.

Data for the wild-type and A138T enzymes are in panels A and B respectively. All values are given relative to the interaction with the specific site most preferred in binding (CGC flanking context). DNA substrates are 24mers (full sequence in Appendix A) The truncated sequence is aligned with its data point and placed to the right of the graph. The transition state interaction free energy (ordinate) is given by $\Delta\Delta G_I^{o\ddagger} = \Delta\Delta G^{\circ}_{Bind} + \Delta\Delta G^{o\ddagger}$. The line of unit slope represents $\Delta\Delta G_I^{o\ddagger} = \Delta\Delta G^{\circ}_{Bind}$, that is $\Delta\Delta G^{o\ddagger} = 0$. Only the sequence with flanking AAAA has $\Delta\Delta G^{o\ddagger}$ significantly greater than 0.

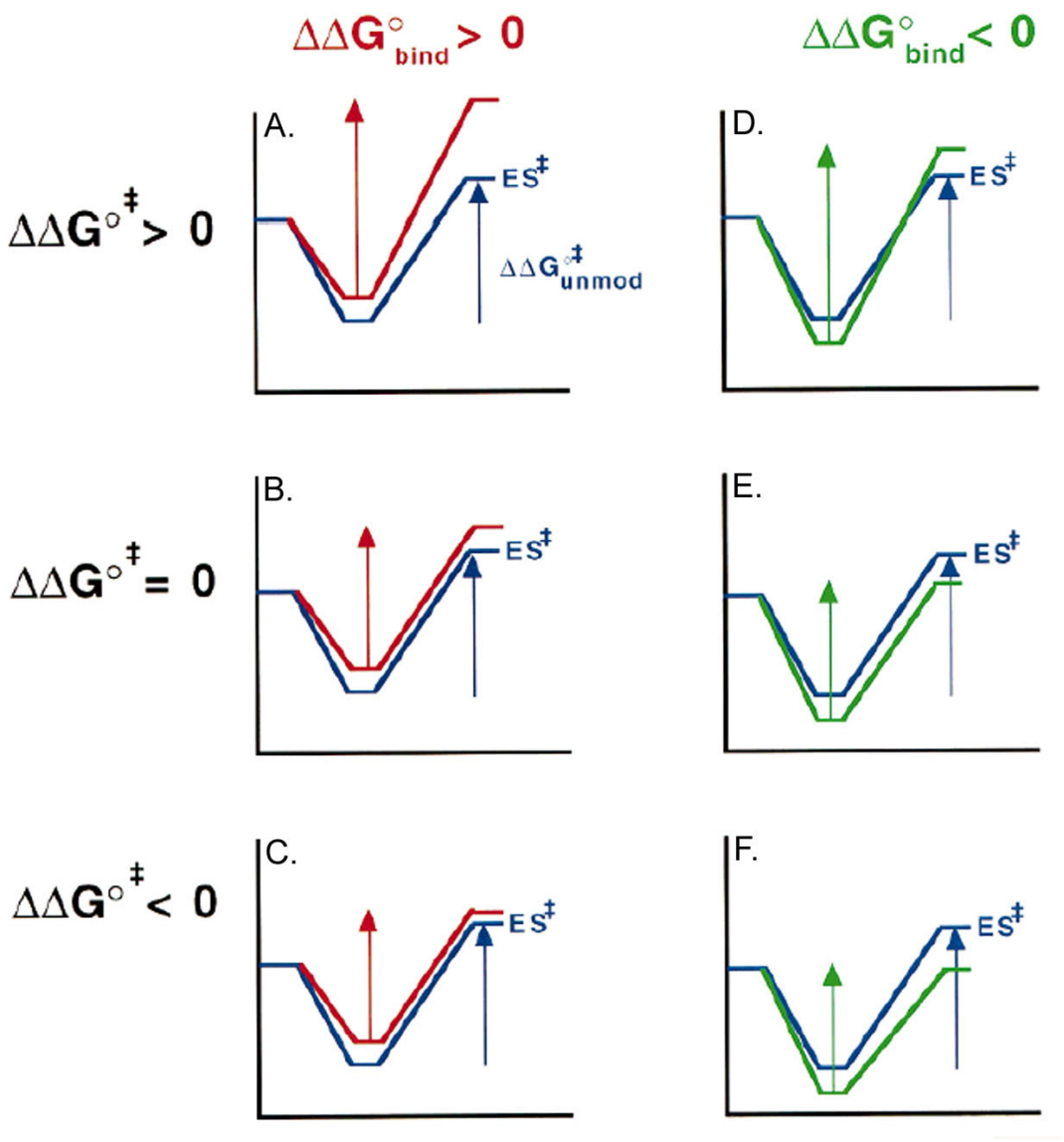


Figure 2.3 Reaction coordinate diagrams showing how perturbation of the protein or DNA can affect endonuclease binding ($\Delta G°_{bind}$) and cleavage ($\Delta G°^{‡}$).

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, ($\Delta G°^{‡}$) for the unmodified substrate is indicated by the blue arrow. Reaction profiles are in red for modifications that inhibit binding ($\Delta\Delta G°_{Bind} > 0$) and in green for modifications that improve binding ($\Delta\Delta G°_{Bind} < 0$). Activation free energies for modified substrates are indicated by corresponding red and green arrows. Figure taken directly from (Jen-Jacobson, 1997).

2.2.4. Thermodynamic dissection of $\Delta G°_{bind}$ of *Eco*RI specific binding by van't Hoff analysis

The standard Gibbs free energy change (ΔG°) for a molecular association process can be expressed as a sum of the enthalpic and entropic components by the standard equation:

$$\Delta G° = \Delta H° - T\Delta S° \qquad 2.9$$

where ΔH° and ΔS° are the standard enthalpy and entropy changes in the system associated with complex formation respectively. While the free energy change tells us about the stability of the complex, a dissection of the free energy change into its constituent changes in enthalpy and entropy, coupled with structural perturbation of both sides of the protein-DNA interface can inform us about the molecular features of the system that contribute to the observed free energy change.

What are the molecular origins of favorable and or unfavorable entropy and enthalpy changes associated with the formation of a protein-DNA complex? Favorable changes in enthalpy of a macromolecular association reaction derive from the formation of intermolecular contacts (hydrogen bonds, Van der Waals contacts, and salt bridges). In turn, these intermolecular contacts results in the reduction of rotational, vibrational and configurational degrees of freedom of the interacting groups in the complex relative to the free molecules; these processes make an unfavorable contribution ΔS° for the binding reaction (Doig et al., 1995, Dunitz, 1995). Protein-DNA complexes are not formed in a vacuum, but rather in aqueous solution containing dissolved ions. The differential interactions of water and ions with the free and bound molecular species are expected to have profound effects on the thermodynamics of complex formation. Release of bound solvent and ions from molecular surfaces will have a favorable effect on the ΔS° of binding. In fact, a thermodynamic driving force for many protein DNA interactions derives from the release of bound water from non-polar surfaces upon binding (the hydrophobic effect) (Ha et al., 1989, Spolar and Record, 1994). However, the desolvation of polar groups will have an unfavorable effect on ΔH° of binding. Trapping solvent or ions at the interface contributes unfavorably to the ΔS° of a binding reaction, but favorably to ΔH°

provided the trapped solvent molecules bridge interaction between the protein and DNA (Connelly, 1997) It has also been shown that significant changes in the conformations of both the protein and DNA molecules are often coupled to binding. Distortion of the DNA substrate will result in unfavorable $\Delta H°$ for binding primarily due to disruption of stacking interactions between the DNA bases (Jen-Jacobson, 2000). Coupled folding of the protein will contribute favorably to the enthalpy change as a result of forming the intra-molecular non-covalent interactions. The favorable contribution to the enthalpy change associated with coupled folding is opposed by a loss of configurational entropy in the system. Taken together, these ideas suggest that the net favorable free energy change for the formation of a protein-DNA complex is the net of large opposing forces (Figure 2.4)

The observed enthalpy and entropy changes in binding derive from changes in the states of the macromolecules, small ions, and water in the system; these also affect the heat capacity of the system. The change in the heat capacity change of the system at constant pressure is defined as the derivative of the enthalpy change with temperature:

$$\Delta C°_p = \left(\frac{\partial H°}{\partial T} \right)_p \qquad 2.10$$

The major sources of heat capacity changes associated with macromolecular interactions have been proposed to be: 1) Solvent reorganization due to the release of water from non-polar and polar surfaces of the free macromolecules (Spolar and Record, 1994); 2) Changes in the hydrogen bonding networks and electrostatic interactions present in the system (Cooper, 2000); 3) Stiffening of the "soft" vibrational modes (500 cm^{-1} or lower) of the system (Sturtevant, 1977, Ladbury et al., 1994, Jen-Jacobson, 2000); and 4) Coupling of equilibria to binding (i.e. folding, protonation) (Eftink et al., 1983). Thermodynamic studies on a number of DNA binding proteins such as *Eco*RI (Ha et al., 1989, Jen-Jacobson, 2000), *Bam*HI (Engler, 1998, Jen-Jacobson, 2000), EcoRV (L.E. Engler, *unpublished*), cro repressor (Takeda et al., 1992), λ repressor (Merabet and Ackers, 1995), and trp repressor (Ladbury et al., 1994), show that a large, negative heat capacity change is a thermodynamic signature associated with formation of a specific protein-DNA complex. The hydrophobic effect, restriction of vibrational modes, and

formation of hydrogen bonded networks (sometimes including trapped solvent molecules) contribute to the observed negative heat capacity change and are all indicative of the formation of a tight, complementary interface. Thus the heat capacity change, $\Delta C°_p$, represents a third thermodynamic observable (in addition to $\Delta H°$ and $\Delta S°$) that informs us about the changes in the molecular features of the interacting species when a protein-DNA complex is formed in an aqueous milieu. The ability to relate changes in these thermodynamic parameters to structural changes that result from a perturbation to the protein and/or DNA will bring us one step closer to rationally engineering changes in the structure of macromolecules that will achieve predictable functions.

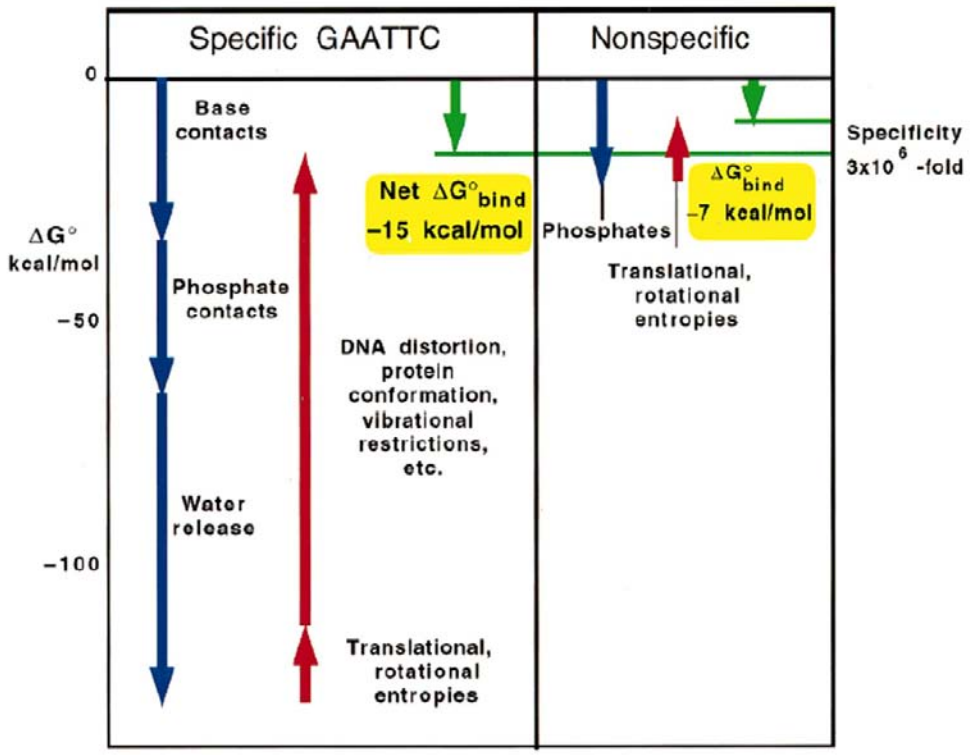


Figure 2.4 Energetic components of specific and non-specific binding of *Eco*RI endonuclease.

Favorable contributions are represented by blue arrows, unfavorable contributions by red, and the net $\Delta G°_{bind}$ by green. The magnitudes of the components of the free energy change are estimates, and may vary from the true values by ±50% (Jen-Jacobson, 1997). The division into separate components should not be taken to mean that these contributions are independent of each other; they are not. Figure taken directly from (Jen-Jacobson, 1997).

Thermodynamic studies on the *Eco*RI endonuclease to date have involved perturbations to the DNA side of the protein-DNA interface (Jen-Jacobson, 1995, Jen-Jacobson, 1997, Jen-Jacobson, 2000, Grigorescu, 2003). In this work, the structural perturbations involve the mutations in *Eco*RI restriction endonuclease that result in relaxed specificity (Heitman and Model, 1990), marking the first thermodynamic analysis of the effects of perturbations to the protein side of the *Eco*RI-DNA interface. To do this, I will use van't Hoff analysis to ask how the 'promiscuous' mutations affect the enthalpy, entropy and heat capacity changes associated with the formation of mutant *Eco*RI-specific DNA complexes.

A non-linear fit of the dependence of the equilibrium association constant, $K_{obs.}$ on temperature to the van't Hoff equation (2.11) gives the heat capacity change of the reaction, $\Delta C°_p$, and two characteristic temperatures, T_H, and T_S, where $\Delta H°=0$ and $\Delta S°=0$ respectively.

$$\Delta G° = \Delta H° - T\Delta S° = \Delta C°_p \left[(T - T_H) - T \ln\left(\frac{T}{T_S}\right) \right] \qquad 2.11$$

Using the values of $\Delta C°_p$, T_H, and T_s, one can calculate the enthalpy and entropy changes at any temperature from the following equations:

$$\ln K_{obs.} = \left(\frac{\Delta C°_p}{R}\right)\left[\left(\frac{T_H}{T}\right) - \ln\left(\frac{T_S}{T}\right) - 1\right] \qquad 2.12$$

$$\Delta H° = \Delta C°_p (T - T_H) \qquad 2.13$$

$$\Delta S° = \Delta C°_p \ln\left(\frac{T}{T_S}\right) \qquad 2.14$$

Plots of ln Kobs. vs. temperature for the *Eco*RI specific DNA binding interactions are shown in figure 2.5. My observed value of the heat capacity change for the wild-type interaction (-1.8 kcal/mol K, Table 2.3) is different from the values reported in the literature (Ha et al., 1989, Jen-Jacobson, 2000, Grigorescu, 2003). These differences are explained as follows: The studies

by Ha et al. (Ha et al., 1989) examined the temperature dependence of *Eco*RI binding to pBR322 DNA, which contains many non-specific sites, and whose GAATTC site is abutted by asymmetric flanking triplets in the two DNA strands, with the identities of the two flanking triplets being different from the CGC flanking context used in my and the remaining studies. The $\Delta C°p$ value of -2.5 kcal/mol K reported by Jen-Jacobson and colleagues (Jen-Jacobson, 2000) was obtained at a lower salt concentration (0.18 M NaCl) than the salt concentrations in which I made my measurements (0.27 M KCl). It has been shown that the salt concentration can have a dramatic effect on the both the enthalpy and heat capacity changes associated with DNA binding (Kozlov et al., 1998, Kozlov et al., 2000, Holbrook et al., 2001, Poon et al., 2004), with the magnitude of the heat capacity change decreasing strongly with increasing salt concentration in each of these cases. Grigorescu found a heat capacity change of -2.6 kcal/mol K for specific DNA binding (Grigorescu, 2003). The conditions (0.4M KOAc, 20mM BTP, pH7.3) for these measurements were also different than the conditions in which I made my measurements (0.27M KCl, 10mM Cacodylate, pH 7.3). In this case, differences in both the salt concentration and the identity of the anion (L. E. Engler and L. Jen-Jacobson, unpublished) can explain the smaller magnitude of $\Delta C°_p$ that I have observed.

In comparing my findings for the heat capacity changes associated with formation of wild-type and mutant specific complexes, the values for wild-type, A138T and E192K binding are the same within experimental error (Table 2.3). However, the H114Y enzyme specific binding reaction is accompanied by a small, but significant ($\Delta\Delta C°_p$ of -0.3 kcal/mol K) increase in the magnitude of the heat capacity change relative to wild-type EcoRI binding (Table 2.3).

Implicit in the differences in specific binding free energy changes observed between wild-type and mutants, is that the constituent enthalpy and entropy changes must differ as well. Formation of the A138T, H114Y and E192K specific complexes is attended by a less favorable enthalpy change than wild-type binding by 2 to 5 kcal/mol at 25°C (Table 2.3). The reduction in the magnitude of the favorable enthalpy change associated with mutant binding is more than compensated for by an increase in the favorable entropy change for mutant binding of 3.5 to 7.4 kcal/mol compared to wild-type binding at 25°C, resulting in the more favorable change in specific DNA binding free energy observed for the mutants (Table 2.3).

Table 2.3 Thermodynamic parameters for Specific *Eco*RI binding at 25°C [a]

	$\Delta C°_p$ (kcal/mol K)	T_H	T_S	$\Delta G°$ (kcal/mol)	$\Delta H°$ (kcal/mol)	$T\Delta S°$ (kcal/mol)
WT	-1.84 ± 0.04	289.8 ± 0.1	296.9 ± 0.1	-13.1	-15.4 ± 0.4	-2.3 ± 0.2
A138T	-1.90 ± 0.04	291.4 ± 0.1	299.1 ± 0.2	-14.6	-12.8 ± 0.3	1.8 ± 0.3
E192K	-1.79 ± 0.08	292.4 ± 0.2	301.0 ± 0.3	-15.4	-10.3 ± 0.5	5.1 ± 0.6
H114Y	-2.13 ± 0.05	291.8 ± 0.1	298.7 ± 0.1	-14.7	-13.5 ± 0.4	1.2 ± 0.3
	$\Delta\Delta C°_p$ [b] (kcal/mol K)			$\Delta\Delta G$ (kcal/mol)	$\Delta\Delta H$ (kcal/mol)	$T\Delta\Delta S$) (kcal/mol)
WT	0			0	0	0
A138T	-0.06 ± 0.06			-1.5	2.6 ± 0.5	4.1 ± 0.4
E192K	0.05 ± 0.09			-2.3	5.1 ± 0.6	7.4 ± 0.6
H114Y	-0.29 ± 0.06			-1.6	1.9 ± 0.6	3.5 ± 0.4

[a] Equilibrium association constants were measured as a function of temperature and fit to the van't Hoff (equation 2.12, see text) in order to obtain $\Delta C°_p$, T_H and T_S. These values were then entered into equations 2.13 and 2.14 in order to calculate the $\Delta H°$ and $\Delta S°$ values for the binding reactions at 25°C. Binding constants were measured using the oligonucleotide 5' GGGCGGGCGCGAATTCGCGGGCGC in 10mM cacodylate, 0.27M KCl, pH 7.3.
[b] All $\Delta\Delta$ parameters are referenced to the wild-type.

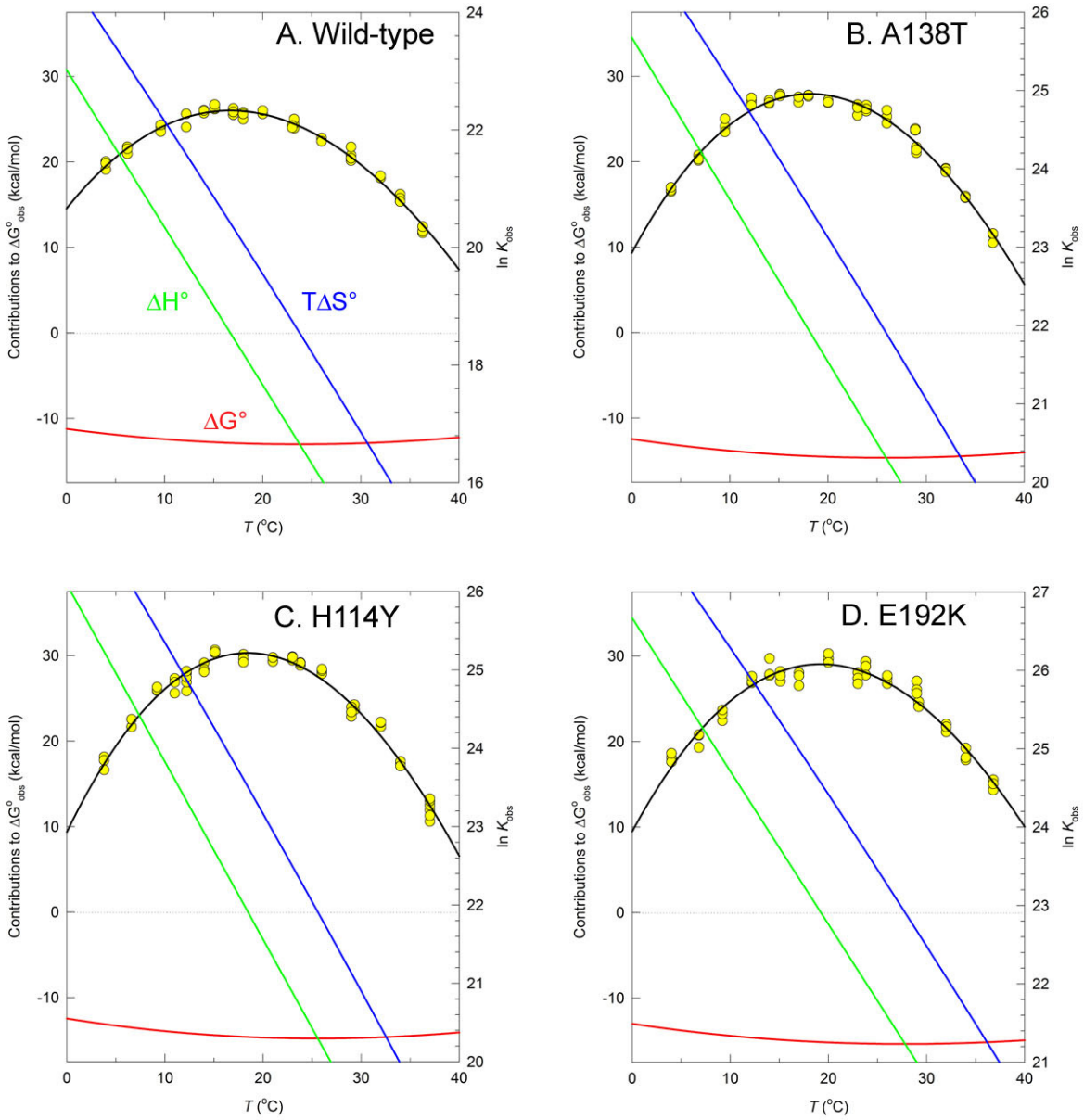


Figure 2.5 Temperature dependence of the *Eco*RI-specific DNA interaction.

The equilibrium association constant for binding to the oligonucleotide containing the specific recognition sequence, 5'-GGGCGGGCGCGAATTCGCGGGCGC was measured as a function of temperature in 10mM Cacodylate, 0.27M KCl, pH 7.3. The raw data points for $K_{obs.}$ are the yellow circles. The black line represents the fit of the temperature dependence of $K_{obs.}$ to the van't Hoff equation (eq. 2.12 see text). The enthalpy (green line) and entropy changes (blue line) were calculated with equations 2.13 and 2.14 respectively. The experimental ΔG° value (red line) is calculated from $\Delta G°_{bind} = -RT\ lnK_{obs.}$

2.2.5. Structural and energetic factors that may contribute to the improved specific binding by the promiscuous enzymes

I have shown that tighter specific binding is common to each of the relaxed specificity mutant enzymes, and that the promiscuous mutant enzymes share a common 'thermodynamic signature'; that is the tighter binding affinity relative to the wild-type enzyme is driven by a favorable increase in the entropy change for the binding reactions. The finding that mutants bind more tightly to the specific recognition site is surprising since an increase in binding affinity for the specific site is inconsistent with the traditional notion of relaxed specificity where a reduction in the binding specificity ratio ($K_{specific}/K_{non\text{-}specific}$) is expected. Further, two of the three residues implicated in relaxed specificity, Ala138 and His114, reside in DNA recognition elements of the protein which have been shown to be critical to the activity of the enzyme (Yanofsky et al., 1987, Grabowski et al., 1995). Thus, tinkering with these regions would not be expected to confer a gain in favorable binding free energy. In the sections that follow, I will discuss each of the mutants in turn, and present hypotheses to explain the possible origins of the more favorable specific binding and differences in the thermodynamic parameters exhibited by the mutant enzymes.

2.2.5.1. The A138T Mutant

Of the amino acids implicated in relaxed specificity, Ala138 is the only residue that makes a hydrogen bond with a DNA functional group within the GAATTC recognition sequence; the main chain carbonyl oxygen pf Ala138 accepts a hydrogen bond from the 4-amino group of the cytosine (Figure 2.6). Alanine 138 is located within the 'extended chain motif' of the *Eco*RI endonuclease. This motif contains many contiguous residues that are responsible for the direct readout of the GAATTC site. The extended chain motif has been shown to be a hot-spot for the isolation of null mutations in the *Eco*RI endonuclease, demonstrating that this region of the molecule is critical for its activity (Yanofsky et al., 1987). Thus, perturbation of the extended chain motif would not be predicted *a priori* to confer an increase in DNA binding affinity on the molecule. I will defer discussion of the possible structural basis for improved

A138T binding and the differences in ΔH° and ΔS° until Chapter 3 where the structure of the A138T specific complex is presented.

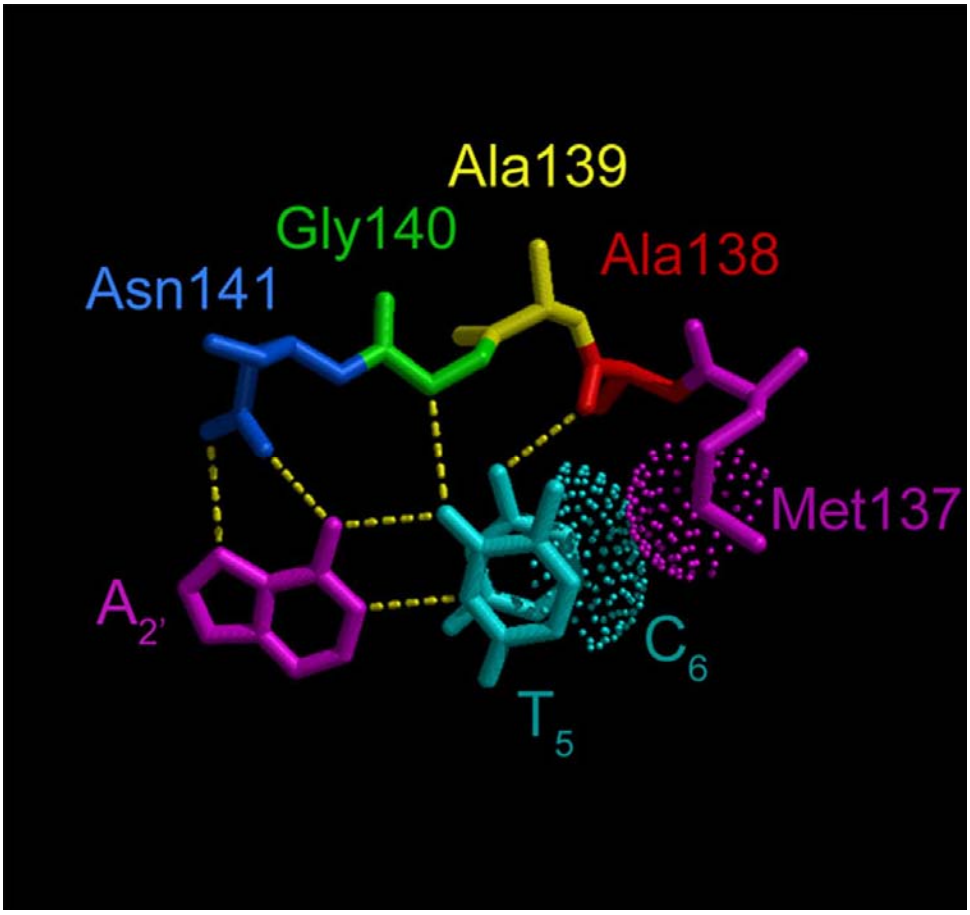


Figure 2.6 *In silico* modeling of the protein-DNA interface in the wild-type Ala138 region.
Coordinates are from a further refined version (Grigorescu, 2003) of PDB entry 1CKQ. Contacts between the *Eco*RI endonuclease extended chain motif and the GAATTC recognition site are shown. Hydrogen bonds are depicted as yellow dotted lines, and van der Waals contacts by dotted spheres.

2.2.5.2. The H114Y Mutant

In the *Eco*RI specific complex, His114 is positioned to make a Coulombic interaction with one of the non-bridging oxygens of the GApATTC phosphate (Figure 2.7A). One hypothesis is that the H114Y substitution might replace the Coulombic interaction with a more

favorable hydrogen bond between the tyrosine hydroxyl and the DNA phosphate. To test this hypothesis, I replaced His114 with a tyrosine in the structure of the wild-type-specific DNA complex *in silico*, while varying the side-chain rotation. The first simulation, (Figure 2.7A), in which I retained the dihedral angles characteristic of the canonical amino acid, showed that the geometry does not support the formation of a novel hydrogen bond, because the tyrosine hydroxyl in 6.8 Å from the nearest DNA functional group. Next, structural models were created in which His114 was replaced by each of the three most common tyrosine rotamers observed in high-resolution crystal structures (Kleywegt et al., 1998). The first of the rotamers (48% of tyrosines in high-resolution structures) places the tyrosine in position to pack against the Glu170 side-chain, but no new protein-DNA contacts are made (Figure 2.7B). The second rotamer (32% of tyrosines in high-resolution structures) positions the side-chain such that a hydrogen bond might form between the tyrosine hydroxyl and the main-chain carbonyl of Gly116 (Figure 2.7C), the main-chain amide of which makes a functionally important contact to the GAApTTC phosphate (Lesser et al., 1992). The third rotamer (14% of tyrosines in high resolution structures) would result in a severe steric clash between Tyr114 and the DNA backbone, unless there were significant structural adaptations in the protein (Figure 2.7D).

It should be emphasized that these molecular modeling studies represent only a first step in probing the structural effects of the H114Y substitution. It would be more desirable to follow *in silico* mutagenesis with molecular dynamics simulations of the complex in order to gain an understanding of the effects of the mutation on the structure and dynamics of the complex. It is also significant that other amino acids besides tyrosine can be substituted for His114 to confer a relaxed specificity phenotype. A saturation mutagenesis study of residue 114 revealed that mutations to phenylalanine, serine and threonine were also promiscuous (Flores et al., 1995). Taken together with the simulations in Figure 2.7, these observations strongly imply that the improved binding of the A138T and H114Y enzymes to the specific GAATTC site is not the result of new direct protein-base contacts.

In her structural characterization of the unbound *Eco*RI endonuclease, Grigorescu noted that three out of the four residues that delineate the boundaries between the ordered main domain and the disordered arms are glycines which make interactions with the DNA (Grigorescu, 2003). She hypothesized that these interactions are partially responsible for stabilizing the folded conformation of the arms that is induced upon specific DNA binding (Grigorescu, 2003).

Interestingly, it is one of these glycine residues, Gly116, which potentially forms a hydrogen bond with the side-chain Tyr114 in one of the modeled complexes (Figure 2.7C). This residue is disordered in the free enzyme, but adopts a fixed conformation in the complex. The penalty in configurational entropy for restricting the motion of a glycine residue is greater than the penalty for restricting the motion of other side-chains (Pace et al., 1998), since glycine is free to inhabit a greater area of main-chain dihedral angle space than any other amino-acid. If the putative interaction between Tyr114 and Gly116 in the complex also exists in the free enzyme, the motion of Gly116 would already be restricted in the unbound state. The result would be a more favorable $\Delta S°$ for mutant binding relative to wild-type binding, which is consistent with my thermodynamic results (Table 2.3).

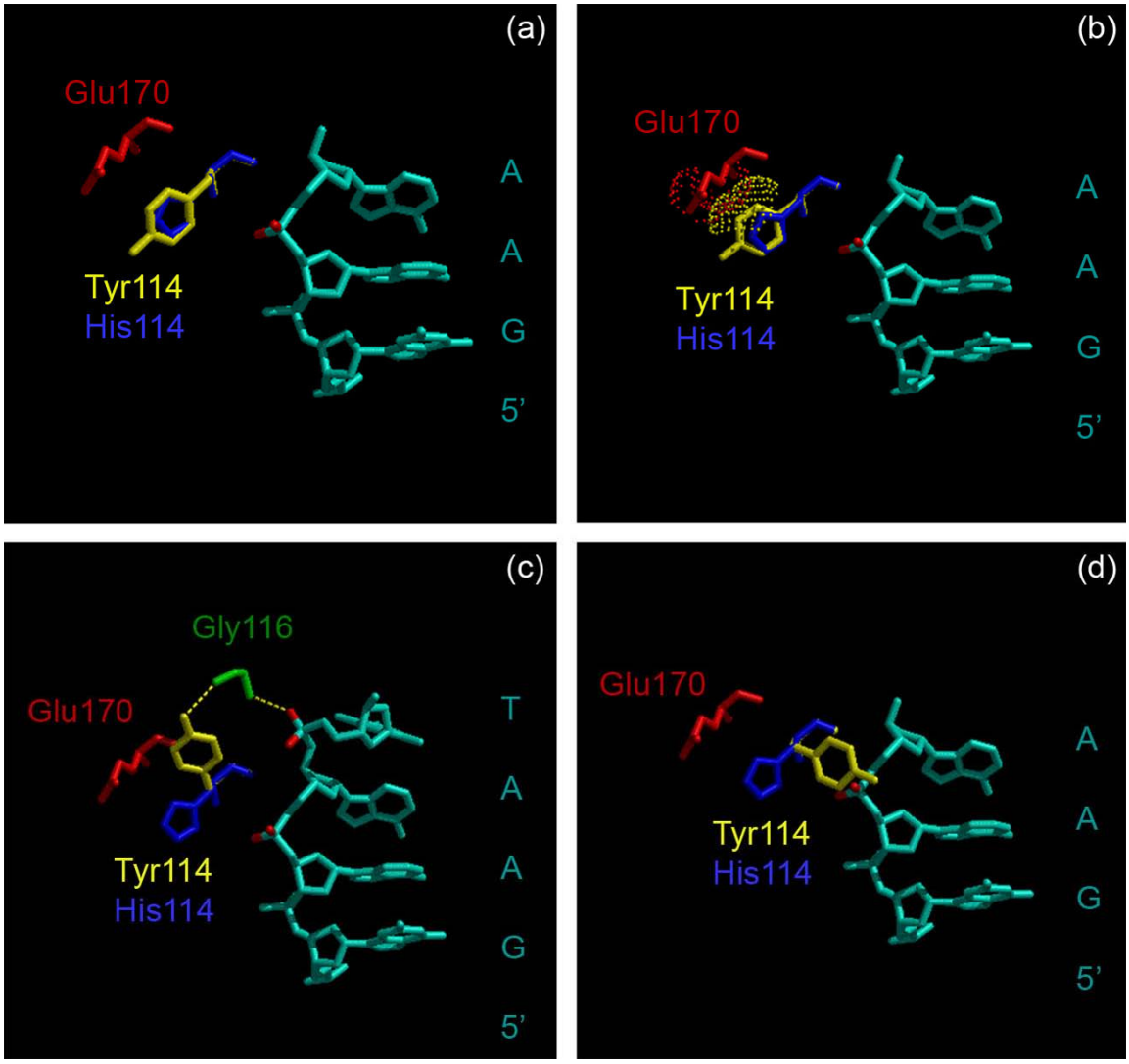


Figure 2.7 *In silico* modeling of the protein-DNA interface and the mutant H114Y region.

Coordinates are from a further refined version (Grigorescu, 2003) of PDB entry 1CKQ.

(a) In the wild-type structure, His114 (blue) is in position to make Coulombic interactions with the phosphate at GApATTC and the Glu170 side-chain. In this panel, His114 was changed to tyrosine (yellow) while conserving the main-chain and side-chain dihedral angles. This conformation of Tyr114 would not result in the formation of any new protein-DNA contacts.

(b) His114 (blue) is changed to tyrosine (yellow), adopting the rotamer that is observed for 48% of tyrosines in high-resolution x-ray structures. Tyr114 is too far (7.3 Å) from the DNA to make significant contacts.

(c) His114 (blue) is changed to tyrosine (yellow), adopting the rotamer that is observed for 32% of tyrosines in high-resolution x-ray structures. In this conformation, Tyr114 is at appropriate distance and angle to make a novel hydrogen bond with the carbonyl of Gly116, the amide of which in turn contacts the GAApTTC phosphate.

(d) His114 (blue) is changed to tyrosine (yellow), adopting the rotamer that is observed for 14%of tyrosines in high-resolution x-ray structures. This produces a severe steric clash between the Tyr114 side-chain and the DNA backbone.

Studies of the H114Y enzyme were conducted as a collaboration between Crystal Ann dela Torre and myself. Some of the analysis that follows is adapted from the paper Crystal prepared as a Beckman scholar. In our characterization of the specific binding affinity of the H114Y enzyme, we made the observation that the mutant protein is much slower to reach equilibrium than the wild-type. In filter binding assays, we find a maximum value of the K_A for specific binding after 18' of equilibration for the wild-type-specific DNA binding reaction (0.22M KCl, pH 7.3, 21°C). When we prolong the duration of incubation, we obtain the same K_A value, indicating that we are indeed scoring the amount of protein-DNA complex present at equilibrium if we filter the reaction at 18'. In a similar analysis of the H114Y enzyme, we find that the K_A value for mutant specific binding reaches a maximum only after *five hours* of incubation (0.22M KCl, pH7.3, 21°C). This prompted us to ask how the DNA dissociation and DNA association rates are affected by the H114Y mutation. Experimental measurements of the dissociate rates (k_d) of the of the wild-type and H114Y specific DNA complexes are shown in Table 2.4. Whereas the H114Y enzyme binds more tightly to specific DNA than the wild-type enzyme by 17-fold ($\Delta\Delta G°_{bind}$ = -1.6 kcal/mol), the rate of dissociation of the mutant complex is 80-fold slower than the wild-type complex dissociation rate ($-\Delta\Delta G°_d{}^{\ddagger}$ = -2.6 kcal/mol). By inserting the values of the equilibrium association constant and the dissociation rate constant into the equation,

$$K_A = \frac{k_a}{k_d} \qquad 2.15$$

we calculate that the H114Y mutant has a 5-fold reduction in the apparent DNA association rate, k_a, compared to the wild-type protein, consistent with the increased time required for the equilibration of the mutant binding reaction (Table 2.4). The H114Y mutation therefore represents the only perturbation of the *Eco*RI endonuclease (including the A138T and E192K promiscuous mutants) or the GAATTC recognition site that affects the rate limiting step of DNA of association such that $\Delta\Delta G°_{bind} \neq -\Delta\Delta G°_d{}^{\ddagger}$ (Jen-Jacobson, 1995).

Table 2.4 Comparison of thermodynamic and kinetic constants for specific [a] binding between the wild type and H114Y enzymes

Enzyme	K_A (M^{-1})	k_d (sec^{-1})	k_a (sec^{-1})
Wild type *Eco*RI	$9.6 \pm 0.3 \times 10^{9}$	$7.4 \pm 1.0 \times 10^{-3}$	$7.1 \pm 0.1 \times 10^{7}$
H114Y *Eco*RI	$1.6 \pm 0.1 \times 10^{11}$	$9.3 \pm 2.0 \times 10^{-5}$	$1.5 \pm 0.2 \times 10^{7}$

[a] Sequence of the double stranded oligonucleotide is 5' GGGCGGGCGCGAATTCGCGGGCGC. The *Eco*RI specific recognition sequence is underlined. Comparison of the wild type and H114Y mutant binding interactions with the specific site at 0.22M KCl, pH 7.3, 21°C. The K_A and k_d values shown were experimentally determined. The k_a values shown were calculated from measurements of K_A and k_d using the relationship $K_A = k_a / k_d$.

To gain insight into the origin of the increased time required to reach H114Y-specific DNA binding equilibrium, a useful conceptual framework involves the comparison of possible kinetic schemes for the formation of wild type and H114Y specific complexes. For wild-type *Eco*RI, formation of the specific complex proceeds via the collision complex ES, generally viewed to have the structural and energetic features of a "nonspecific" complex (Figure 2.8A). The rate of formation of this initial collision complex is diffusion controlled (Jack et al., 1982). The specific complex E'S', visualized in the crystal structure, shows that both the protein and DNA conformations are dramatically different from the free molecules, and it is the formation of this E'S' complex that we are scoring in our filter binding assays (see section 2.1.1). Upon addition of the essential Mg^{2+} cofactor, the E'S' complex proceeds to the activated or transition state $[E'S']^{\ddagger}$ to produce the cleaved DNA product (P) via the enzyme-product intermediates (E'P' and EP). For wild type enzyme, it has been established that the conformational transition from the nonspecific (ES) complex to the pre-transition state E'S' complex is the rate-limiting step for the formation of the specific complex (Halford and Johnson, 1983). The equilibrium association constant (K_A) determined in the binding reactions describes the overall reaction for specific complex formation from the free molecules (E+S to E'S'). (Note: $K_{app} = k_{a,app} / k_{d,app} = K_{A,1}K_{A,2}$; where $K_{A,1}=k_{a1} / k_{d1}$ and $K_{A,2}= k_{a2} / k_{d2}$).

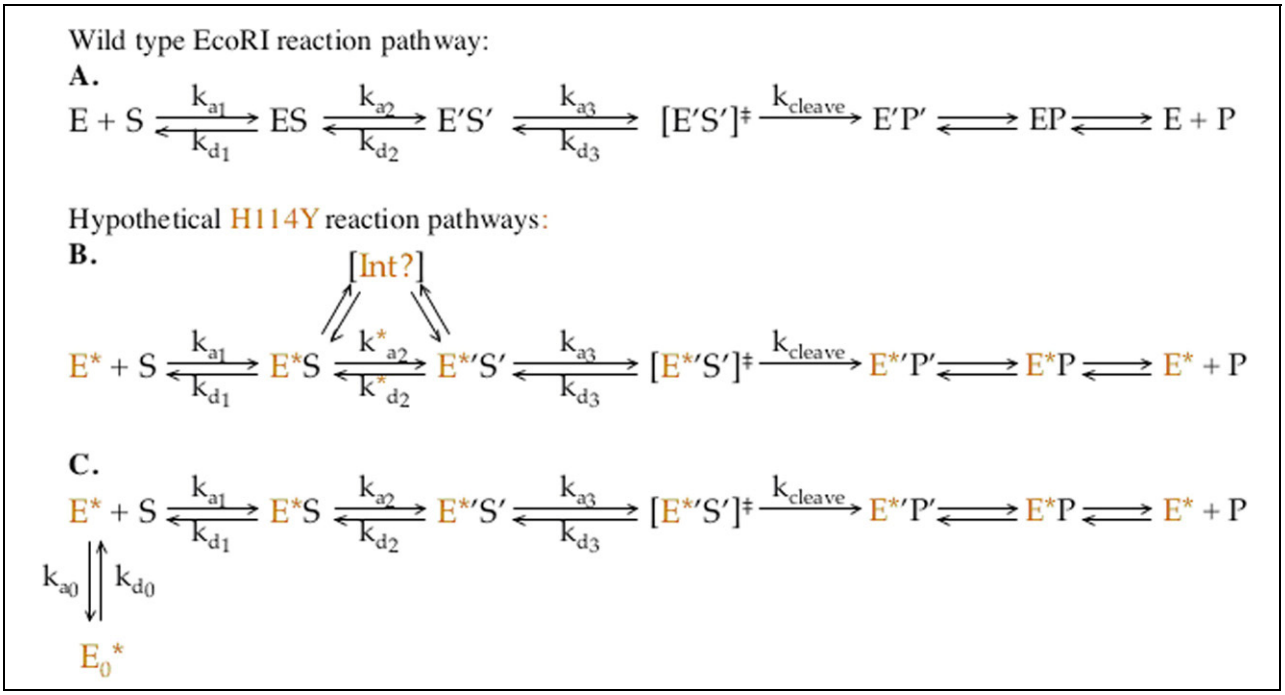


Figure 2.8 Possible differences in the microscopic kinetic constants in reaction pathways of the wild-type and H114Y *Eco*RI endonucleases.

A) Reaction pathway for the wild-type *Eco*RI endonuclease. B) Possible reaction pathway for H114Y *Eco*RI endonuclease (in orange), where the initial collision complex E*S proceeds through a different intermediate on the path towards the pre-transition state complex, E*'S'. C) Alternative pathway for H114Y *Eco*RI, where the free enzyme E_0* undergoes a mandatory isomerization prior to formation of the collision complex. The models depicted in schemes B and C can explain the reduction in the apparent association rate constant for H114Y specific DNA binding (Table 2.4).

The proposed kinetic schemes for the H114Y mutant are shown in Figures 2.8B and 2.8C. Which intrinsic constants might change to account for the observations of slower apparent dissociation and association rates (Table 2.4)? We presume that k_{a1} and k_{d1} remain the same for the mutant because they reflect the diffusion-controlled step for formation of the collision complex from free protein and DNA. One simple possibility is that changes have occurred in the intrinsic kinetic constants k_{a2} and k_{d2} constants for the rate limiting conversion (isomerization) to E*'S'. There is also the possibility that an additional intermediate is involved in the rate limiting conversion of E*S (nonspecific complex) to E*'S' (Figure 2.8B). A second possibility (Figure

2.8C) is that the *free* H114Y must undergo a slow isomerization step (k_{ao}, k_{do}) prior to binding DNA.

How might we view the above isomerizations at the structural/molecular level? All the promiscuous mutations reside either in the arm (which is disordered in the free enzyme) or at the junction between the arm and the primary core recognition domain. Cognate site binding induces folding (ordering) of the arms, which embraces the DNA (visualized in the structure of the co-crystalline complex, Figure 1.10). The coupled folding of the arms is the only known conformational change that occurs in the protein upon binding as revealed from a comparison of the crystal structures of the free and bound enzymes (Grigorescu, 2003). Thus it is an attractive hypothesis to suggest that the isomerization of the nonspecific complex to a form in which the specific DNA is enfolded by the arms is the rate limiting step for binding and that the rate of this step is slower for the H114Y mutant than for wild type. There is the possibility that an improperly folded intermediate complex must be slowly converted in order to observe maximum affinity of the specific complex, but once the specific complex is formed the unwrapping of the arm takes place much more slowly for H114Y than wild type enzyme, with the consequence that the lifetime of the H114Y cognate complex is much longer. Alternatively the H114Y mutation produces a partially but improperly folded free enzyme that must isomerize even before forming the initial collision complex (Figure 2.8C). Interestingly, the time required to reach equilibrium is dependent on salt concentration. That is, at higher salt concentrations (0.3M KCl, pH 7.3, 21°C), the H114Y-cognate DNA interaction achieves equilibrium after ~20 minutes rather than the 5 hours at 0.22M KCl. That is, differences in the intrinsic kinetic k_{a2}, and k_{d2} constants for wild-type and mutant specific complex formation are no longer significant at higher salt concentrations or that the higher salt disrupts an improperly formed interaction in the free enzyme (E_o*; Figure 2.8C). Indeed, the inner and outer arms are highly charged, with many salt-links formed between oppositely charged amino acids in the specific complex (Figure 2.9). The two models in figure 2.8B&C can potentially be distinguished by rigorous transient-state kinetic measurements of apparent association ($k_{a,app}$) and dissociation ($k_{d,app}$) rate constants or by experimental approaches that can characterize the dynamics of the arms in the free and bound proteins.

We anticipate that the H114Y *Eco*RI complexes with miscognate DNA sites will also have longer lifetimes than the wild-type counterparts. Longer miscognate complex lifetimes

would allow more time for the promiscuous H114Y mutant to make double stranded cuts at these sites. Thus it will be of interest to ask if the structure and/or dynamics of the arms of the free enzyme, the specific complex or miscognate complexes differ between the wild-type and H114Y *Eco*RI enzymes. In chapter 5, I describe experimental approaches to probe the structure and dynamic properties of the arms of the free and bound *Eco*RI endonuclease.

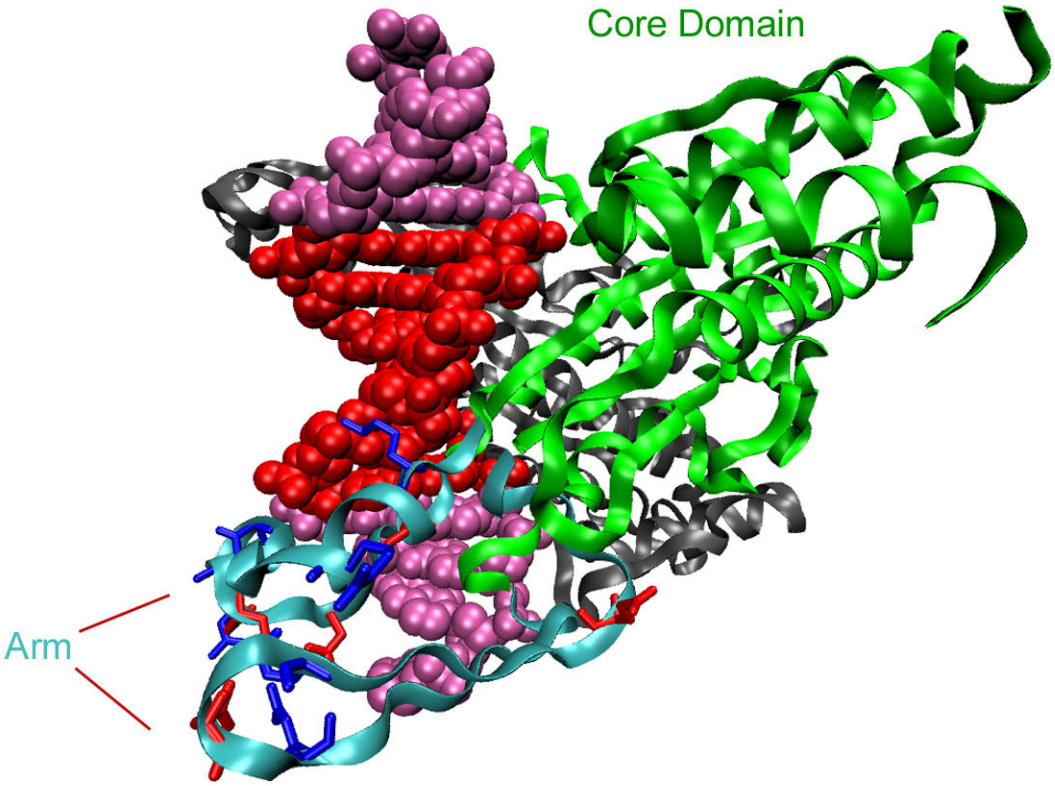


Figure 2.9 "Arms" of the *Eco*RI restriction endonuclease have a high abundance of charged amino-acids.

The *Eco*RI-d(TCGCGAATTCGCG) complex (Grigorescu, 2003) is shown with the GAATTC site in red, and the flanking DNA regions in magenta. The core domain of one enzyme subunit is in green, with the arm of the same subunit in cyan. The complete second enzyme subunit is in grey. Acidic (Asp and Glu) and basic (Lys and Arg) amino acids within the arm are colored red and blue respectively.

2.2.5.3. The E192K mutant

In the structure of the wild-type specific complex, Glu192 is located within the arm, 9 Å from the DNA backbone, across from the DNA base-pairs flanking the GAATTC site. This distance is too great for Glu192 to make a contact with the DNA in the canonical complex. E192K is however the only promiscuous mutant for which tighter binding can be rationalized, where a gain in favorable electrostatic binding free energy may derive from both the removal of a negative charge and the addition of a positive charge proximal to the negatively charged DNA backbone (Heitman and Model, 1990). Interestingly, the Y193H mutant of the *Eco*RI endonuclease also exhibits relaxed specificity, lending strength to the argument that the placement of a positive charge in this region is sufficient for the relaxed specificity phenotypes of both the E192K and Y193H mutants. Mutagenesis studies with the *Bam*HI endonuclease (Xu et al., 1991), λ-repressor (Nelson et al., 1985) and lexA repressor (Oertel-Buchheit et al., 1992) have also shown that substitution of a basic residue for an acidic residue at the protein-DNA interface leads to more favorable binding. Studies which used the non-linear Poisson-Boltzman equation to calculate differences in the electrostatic binding free energies between the wild-type and E34K mutant λcI repressor (Misra et al., 1994), predicted an improvement in binding and steeper salt dependence of binding for the E34K mutant, despite the fact that Glu34 is greater than 5 Å away from the DNA backbone of the operator. These results suggest that long range electrostatic effects on DNA binding affinity can be significant.

To test the idea that the E192K mutant affects the electrostatics of complex formation, we (W. H. McCoy IV, S. V Jana, and P. Sapienza) measured the salt dependence of specific DNA binding for the wild-type and three relaxed specificity endonucleases. The plots of log K_A vs. log [KCl] for specific DNA binding are presented in Figure 2.10 and the slopes from linear fits to the data are reported in Table 2.5. It is clear that plots of log K_A vs. log [KCl] are linear for all of the endonucleases, and that the slopes are similar for the wild-type, A138T and H114Y enzymes. However, specific binding depends more strongly on salt concentration for the E192K enzyme (slope of -13.7) compared to the wild-type protein (slope of -11.8).

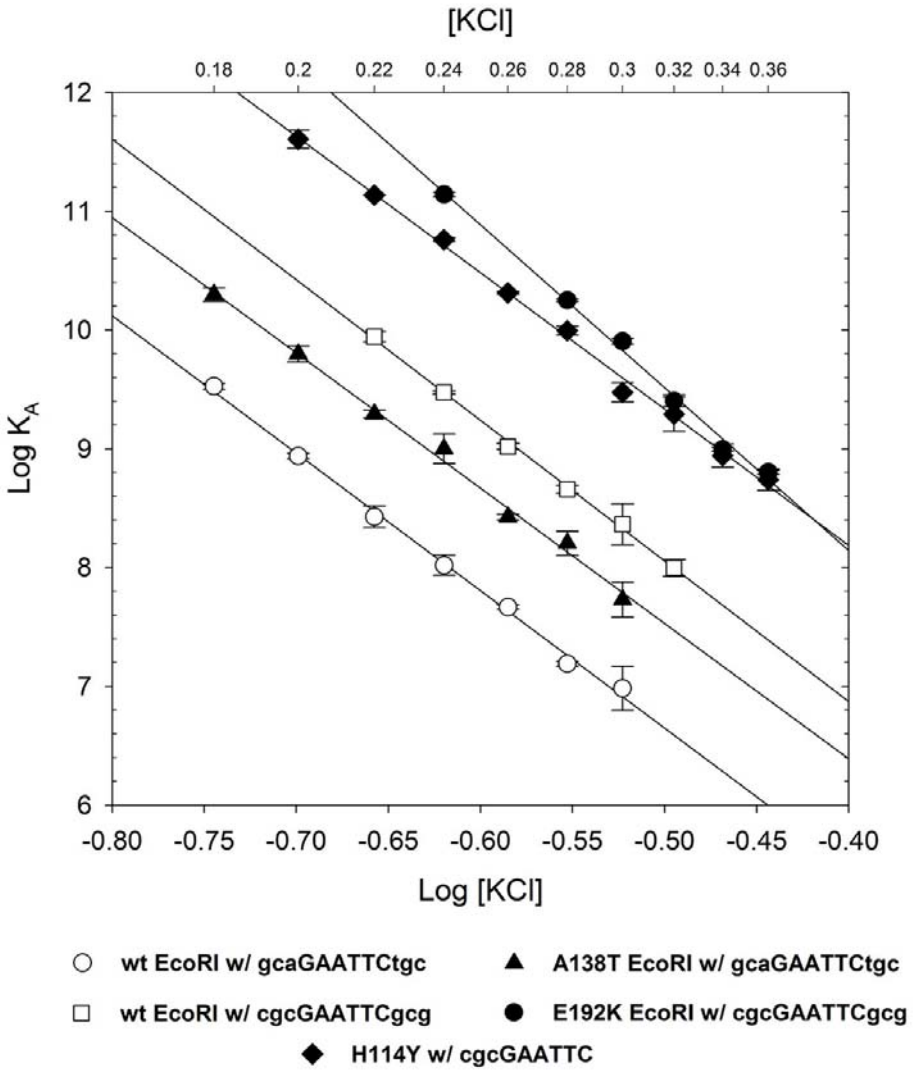


Figure 2.10 Salt dependence of the EcoRI-specific DNA interaction.

Equilibrium association constants (K_A) were measured in binding buffer (pH 7.3, 21°) 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 Table 2.5. DNA substrates are 24-base-pair oligonucleotides containing the GAATTC site embed in the symmetrical GCA or CGC contexts (see key above).

Table 2.5 Salt dependence parameters for *Eco*RI specific binding [a]

Enzyme/Sequence	Slope [b]	R^2	Salt Range
WT/ cgcGAATTCgcg	-11.8 ± 0.2	0.998	0.22-0.32M KCl
E192K/ cgcGAATTCgcg	-13.7 ± 0.5	0.995	0.24-0.36M KCl
H114Y/ cgcGAATTCgcg	-11.5 ± 0.2	0.997	0.20-0.36M KCl
Wt/ gcaGAATTCtgc	-11.6 ± 0.3	0.997	0.18-0.30M KCl
A138T/ gcaGAATTCtgc	-11.4 ± 0.4	0.995	0.18-0.30M KCl

[a] Equilibrium association constants were determined for the 24 base-pair oligonucleotide substrates (only 12 bases are shown in the table) in binding buffer, pH 7.3, 21°C under varying [KCl].
[b] The slopes are reported from a linear regression analysis of log K_A vs. log [KCl] plots (see Figure 2.9).

Salt dependent effects on protein-DNA interactions are typically interpreted within the framework of one of two competing models, the first based on the counter ion-condensation (CC) theory proposed by Manning (Manning, 1977, Manning, 1978), and the other based on the idea that the non-linear Poisson Boltzman equation can be used to solve for the electrostatic potential over the surface of a macromolecule (Sharp et al., 1995).

Counter ion condensation theory models the DNA as a polyanionic rod surrounded by counter ions composed of two distinct populations. A layer of cations is 'condensed' on the DNA, thereby neutralizing the charges on the phosphate backbone. This condensed population does not depend on the solution salt concentration. The second layer is a salt dependent Debye-Hückel layer which is treated as a classical ion atmosphere. According to this model, the salt effects on nucleic acid equilibria are derived from the differential entropies of cation release from DNA at different salt-concentrations, with cation release being more favorable at low salt concentrations. The number of thermodynamically bound cations released (or salt-links formed) upon protein binding can be determined from a plot of log KA vs. log [Salt] according to equation 2.16:

$$Slope = m'*\psi \qquad 2.16$$

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 salt dependence

slopes (Table 2.5) for wild-type and E192K EcoRI specific binding enzymes we calculate (Eq. 2.16) that two additional cations are predicted to be released from the DNA upon binding to the E192K mutant protein compared to the number of cations released upon wild-type complex formation. This result is satisfying since the E192K mutation would place two additional positive charges at the protein DNA interface relative to the wild-type system, one from each enzyme monomer. However, Lys192 would probably not be close enough to the DNA phosphate backbone make a salt-link that would displace a bound cation from a DNA phosphate in the absence of severe conformational changes in the complex. We find such conformational changes unlikely since the outer arm in which residue 192 is situated packs against the inner arm, which contains many of the residues making sequence specific contacts to major groove functional groups. The observation that the E192K mutant maintains its binding affinity for and cleavage activity towards the specific site argues against any major structural adaptations involving the protein arms.

The Poisson Boltzman (PB) model differs from the CC model in two key aspects: Whereas CC theory treats the ion atmosphere as two discreet layers, PB theory models the ion atmosphere around DNA as a single, continuously distributed population. In addition, the PB model relates ion distribution to the detailed molecular structures of both binding partners of protein-DNA interactions. This feature of the PB model is superior to the CC model since the latter considers only the structure of the DNA binding partner, and the treatment of DNA as a polyanionic rod is rather primitive. The differences in the two theories are manifest in differences in the thermodynamic descriptions of salt-dependent effects on protein-DNA equilibria. CC theory ascribes the salt observed dependence of protein-ligand interactions to the entropy of cation release. The PB model views salt effects as the summation of salt-dependent influences on ion-molecule, ion-solvent and ion-ion interactions in the free and bound molecular species. Although the PB model is more sophisticated, we have not performed the calculations necessary to relate differences in salt-dependence slope to differences in the thermodynamic parameters. Under the framework of PB theory, we are presently limited to the conclusion that the E192K and wild-type complexes differ with respect to their global and or local electrostatic interactions. The E192K mutant does represent a case that could be used as a benchmark for verification of emerging methods in electrostatics calculations.

3. THE CRYSTAL STRUCTURE OF THE A138T *Eco*RI-d(TCGCGAATTCGCG) COMPLEX AT 1.95 Å RESOLUTION

3.1. Introduction

We are ultimately interested in how the promiscuous mutations affect the miscognate DNA binding mode of the enzyme, resulting in the relaxation of specificity (see Chapters 4&5). However, *Eco*RI miscognate complexes have thus far proven refractory to high resolution structural characterization (Wilkosz, 1993). We feel that structural analysis of a promiscuous *Eco*RI-specific complex represents an interesting molecular recognition problem that may provide structural correlates to our thermodynamic measurements.

In Chapter 2, I demonstrated that the promiscuous *Eco*RI endonucleases bind more favorably to the specific site than the wild-type enzyme. I also showed that tighter mutant enzyme binding derives from a more favorable entropic contribution to the binding free energy change compared to the wild-type. To ask whether differences in the molecular features of the wild-type and mutant complexes can explain the differences in the observed thermodynamic parameters between the wild-type and A138T mutant, I (in collaboration with Dr. John M. Rosenberg) solved the structure of the A138T-specific DNA complex. In this chapter the crystal structure of the A138T-d(TCGCGAATTCGCG) complex is described and compared to the structure of the wild-type enzyme in complex with a DNA of the same sequence. The coordinates of the wild-type model upon which I will base my comparison have not yet been deposited into the Protein Data Bank (PDB). This latest model was obtained by further refinement (Grigorescu, 2003) of PDB entry 1CKQ. The new model is superior to 1CKQ based upon global improvement in the agreement of the model with the diffraction data (lower R_{Free})

(Brünger, 1992) as well as improvement in the local agreement of the model with the electron density (lower real space R factors) (Jones et al., 1991).

3.2. Results and Discussion

3.2.1. DNA numbering scheme

In this chapter, I will use the DNA numbering scheme used by (Grigorescu, 2003) to describe the site used in crystallography experiments. The scheme has the advantage of providing a consistent frame of reference for describing DNAs in which a palindrome is embedded in DNAs of different lengths. The numbering of bases in the oligonucleotide used in the wild-type and A138T structural studies is:

$$5'\text{-}T_{-7}C_{-6}G_{-5}C_{-4}G_{-3}A_{-2}A_{-1}T_1T_2C_3G_4C_5G_6$$

3.2.2. General features of the A138T *Eco*RI-specific DNA complex

In order to make the most meaningful comparisons between the wild-type and mutant structural models, the A138T *Eco*RI-d(TCGCGAATTCGCG) complex was crystallized, cryo-protected and frozen under the same conditions as those used for preparation of crystals of the wild-type complex. We also followed similar procedures for data collection, data processing and refinement to solve the mutant structure that were employed to solve the 1.87 Å resolution structure of the wild-type complex (See Chapter 6: Materials and Methods). The A138T data set was collected at 100 Kelvin at the X6A beam-line of the National Synchrotron Light Source. Processing of the first frame of diffraction data showed that the crystals of the A138T-DNA complex are isomorphous with those of the wild-type complex. Statistics for data processing of the diffraction data and refinement of the atomic model are given in table 3.1.

Quality checks assessing the stereochemistry of the final model (PROCHECK and WHATCHECK (Laskowski, 1993)) and the local agreement of the diffraction data with the

atomic model (SFCHECK (Vaguine et al., 1999)) were conducted; the results of these quality checks are in Appendices E&D respectively.

Global features of the A138T and wild-type *Eco*RI-d(TCGCGAATTCGCG) complexes are very similar, as the two models can be superposed with an root mean square deviation (rmsd) of 0.8 Å. Backbone atoms in the two models have an rmsd of 0.3 Å. Based on these observations, it is not surprising that the secondary structure assignments are the same for all residues between the wild-type and mutant models (Figure 3.1). A plot of protein backbone rmsd between the wild-type and A138T models is shown in figure 3.2A. Among the regions that show greater than average rmsd between the two structures are the inner (residues 116-140) and outer (residues 172-195) segments of the two arms (one in each protomer) that enfold the DNA in the specific complex. However, the average crystallographic B-factor of a residue within the arms is about 50 $Å^2$ in both structures, meaning that the positions of atoms within the arms of *both* structures are not well determined. This makes the interpretation of small changes in the positions of these residues unjustifiable. Figure 3.2B illustrates this point by plotting electron density profiles through a carbon atom with B-factors of 16, 46, or 96 $Å^2$ (Ten Eyck, 1995). The message is that atoms with high B-factors have low and dispersed electron density, necessarily leading to large uncertainty in the position of these atoms.

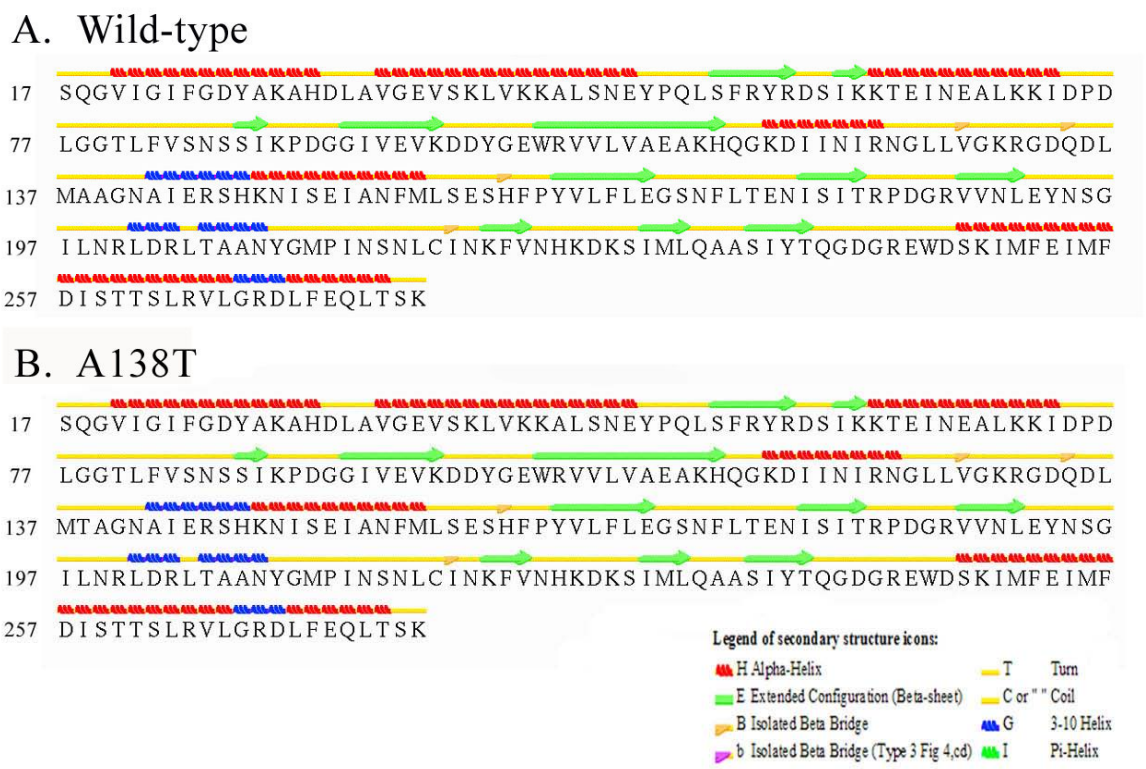
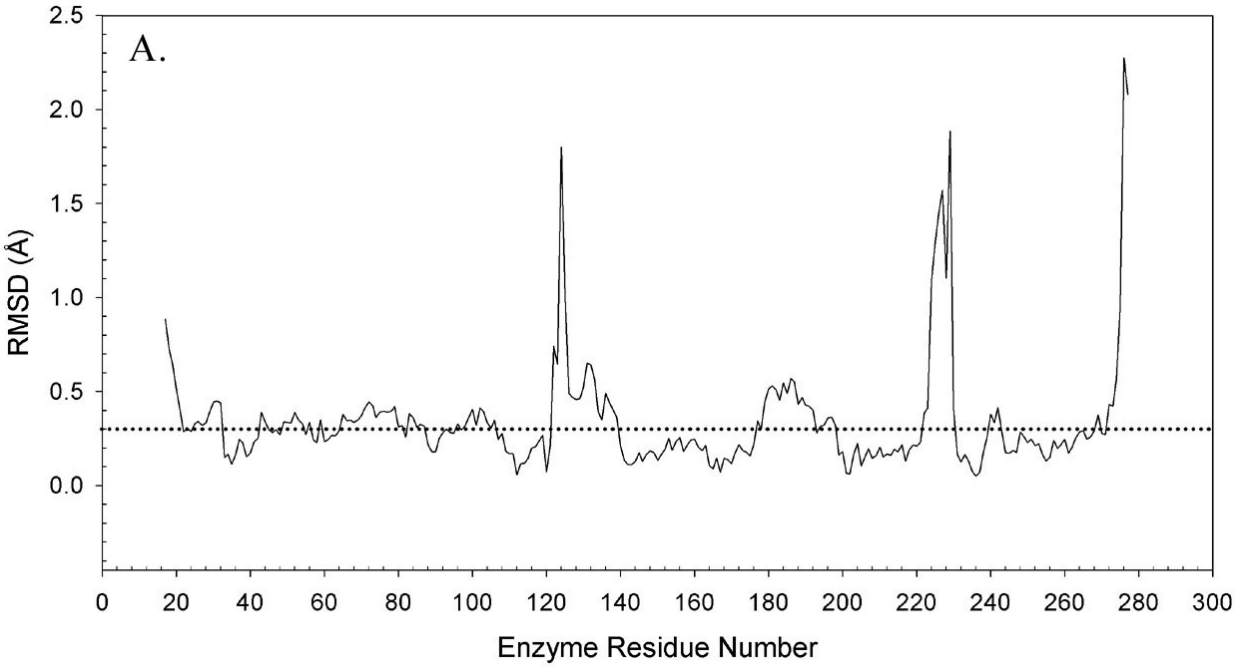


Figure 3.1 Secondary structure assignments in the wild-type and A138T *Eco*RI-d(TCGCGAATTCGCG) complexes.

Secondary structure assignments and this graphic were generated using the program STRIDE (Heinig et al., 2004).

Other regions with larger than average rmsd are the N and C-termini, as well as a loop comprised of residues 224-229 which is located in the back of the molecule, and bathed in solvent. Electron density corresponding to this loop is very weak in both wild-type and mutant maps, indicating that the structure of this regions is not well defined in either complex. It should also be noted that there is no interpretable electron density corresponding to the 16 N-terminal amino-acids of the mutant protein. These 16 residues are also disordered in every wild-type *Eco*RI structure (with and without DNA) solved thus far (pdbids: 1CKQ, 1CL8, 1ER1, 1QPS, 1QRH, 1QR1, 1QC9). Interestingly, experiments using fluorescence resonance energy transfer have suggested that the N-terminus and the 224-229 loop are in close proximity (Liu et al., 1998).



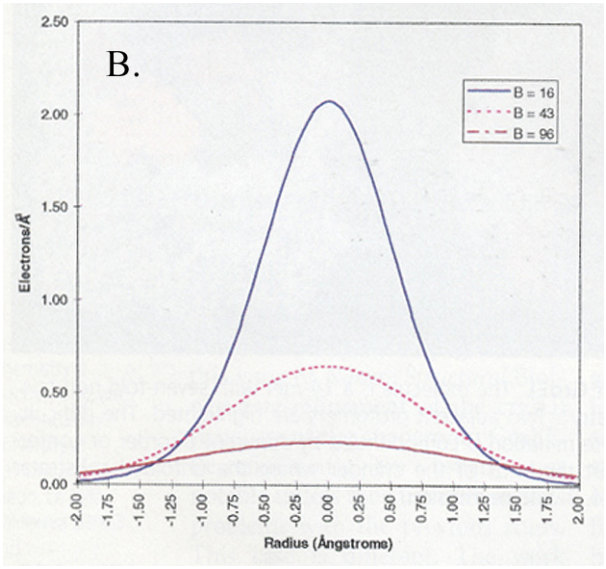


Figure 3.2 Root mean square difference between coordinates of main chain atoms in wild-type and A138T-d(TCGCGAATTCGCG) complexes.

A) RMSD plot, where the dotted line represents the averaged rmsd over all main chain atoms.

B) Electron density profile through a carbon atom with Debye-Waller (B) factors of 16, 43 and 96 $Å^2$ (see legend within figure). The position of an atom with a large B-factor will be ambiguous due to the large spread of the electron density profile and likely overlap with electron density profiles associated with neighboring atoms. Figure taken from (Ten Eyck, 1995).

Table 3.1 Crystal parameters and data collection and refinement statistics for A138T-d(TCGCGAATTCGCG) complex

Crystal Parameters	
Space Group	P321
Unit Cell	a=b= 117.47 Å
	c= 47.48 Å
	$\alpha=\beta= 90°$
	$\gamma= 120°$
Data Collection	
X-ray Source/ λ	Beam line X6A NSLS [a]/ 0.979 Å
Data Collection Temperature	100 K
High Resolution Limit	1.95 Å
Observed Reflections	472,933
Unique Reflections	27,856
Completeness	99.1
R_{merge} [b]	8%
I/σ in highest resolution shell	4.2 (2.01-1.95 Å)
Model Refinement [c]	
Asymmetric unit	1 *Eco*RI monomer, 1 DNA strand
Resolution limits	8 – 1.95 Å
R_{work}	0.20
R_{free}	0.24
rmsd bond length (Å)	0.009
rmsd bond angles (°)	1.3
Ramachandran Plot	
% of residues in most favored regions	90.1%
% of residues in disallowed regions	0%
Average B factors ($Å^2$)	
All atoms (entire structure)	40.3
Protein	
All atoms	40.5
Main chain atoms	39.4
DNA	35.9
Bound solvent (water)	40.4

[a] NSLS is National Synchrotron Light Source, Brookhaven National Labs, Upton NY.
[b] Value from SCALEPACK (Otwinowski, 1997)
[c] Refinement performed with CNS (Brünger et al., 1998)

3.2.3. Local effects of the A138T mutation

Electron density corresponding to the Thr138 side-chain was visible in omit maps during early stages of refinement, and this density was fit with a threonine rotamer from the O database (Kleywegt and Jones, 1998). Residue 138 is located within 'the extended chain motif' (Kim, 1990), which is composed of residues 137 to 142. This element contains multiple amino-acids that make site-specific contacts to the GAATTC site. The β-branched threonine side chain is accommodated within the extended chain motif without any significant structural adaptations (rmsd = 0.16 Å) (Figure 3.3). Figure 3.4 shows electron density corresponding to the extended chain motif and nearby DNA. The quality of this electron density map is indicative of the maps covering the entire recognition interface. All contacts between residues within the extended chain motif and the GAATTC recognition site are intact in the mutant complex (see table 3.2 for list of contacts in the wild-type and A138T complexes). In addition, there are no new contacts made by the A138T enzyme to the GAATTC bases.

Introduction of the threonine 138 side-chain results in changes to the solvent mediated contacts to bases upstream of the recognition site (see table 3.2 for a list of solvent mediated contacts in both structures). [All solvent peaks were assigned (see materials and methods for detailed description of solvent peak assignment criteria) to water molecules in the wild-type and mutant crystallographic models. These solvent molecules will be treated as waters in my discussion, though it should be emphasized that any given solvent peak might also be attributable to other species in the crystallization solution such as NH_4^+, H_3O^+, or OH^-. X-ray crystallographic analysis, at this resolution, cannot distinguish from among these possibilities.]

Table 3.2 Comparison of hydrogen bonds at the protein-DNA interfaces of wild-type and A138T *Eco*RI specific complexes[a]

DNA	Protein	Wild-type Distance (Å)	A138T Distance (Å)
Polar Interactions w/ DNA Bases			
A_{-2} N6	N141 OD1 (B)	2.8	2.8
A_{-2} N7	N141 ND2 (B)	3.2	3.1
A_{-1} N6	N141 OD1 (B)	2.8	2.5
A_{-1} N7	R145 NH2 (A)	3.3	3.2
T_1 O4	A142 N (A)	3.0	2.9
T_1 O4	N141 N (A)	3.3	3.2
C_3 N4	T138 O (A)	2.8	2.8
Polar Interactions w/ DNA Backbone			
G_{-5} O1P	G196 N (B)	2.7	2.7
C_{-4} O1P	R203 NH2 (B)	2.9	2.7
C_{-4} O2P	S87 N (A)	3.0	2.9
G_{-3} O1P	K148 NZ (A)	3.1	2.9
G_{-3} O2P	**N149 ND2 (A)**	**3.4**	**3.1**
G_{-3} O2P	K89 N (A)	2.8	2.6
A_{-2} O1P	R145 NH2 (A)	2.9	2.8
A_{-2} O1P	**K113 NZ (A)**	**3.8**	**2.7**
T_1 O1P	G116 N (A)	2.8	2.9
T_2 O2P	K117 NZ (A)	3.0	3.0
Water Mediated Interactions		**W#/W-Prot. Dist./W-DNA-Dist.**	**W#/W-Prot. Dist./W-DNA-Dist.**
C_{-6} N4	**T138 OG1 (A)**	**No T138 in WT**	**W149/2.6/3.3**
G_{-5} O6	**T138 OG1 (A)**	**No T138 in WT**	**W149/ 2.6/2.5**
G_{-5} N7	**T138 OG1 (A)**	**No T138 in WT**	**E74/ 2.6/2.9**
G_{-5} O2P	**N85 N (A)**	**No W157 in WT Model**	**W157/ 3.0/2.5**
C_{-4} N4	**R200 NH2 (B)**	**W41/3.0/3.2**	**No W41 in A138T Model**
C_{-4} N4	**A197 O (B)**	**E41 /3.3/3.2**	**No W41 in A138T Model**
C-4 O2P	**R2003 NE (B)**	**E43/3.1/2.9**	**No W43 in A138T Model**
G_{-3} O1P	N141 ND2 (B)	W2/3.1/2.8	W2/3.0/2.9
G_{-3} O1P	E144 OE2 (B)	W2/2.5/2.8	W2/2.7/2.9
G_{-3} O2P	K89 O (A)	W59/3.1/3.0	W59/3.3/3.3
G_{-3} O6	A139 O (B)	W1/3.1/3.3	W1/3.0/3.2
G_{-3} O6	R200 NE (B)	W1/3.3/3.3	W1/3.1/3.2
G_{-3} O6	R203 NH1 (B)	W1/3.1/3.3	W1/3.1/3.2
G_{-3} N7	A139 O (B)	W1/3.1/2.8	W1/3.0/2.8
G_{-3} N7	R200 NE (B)	W1/3.3/2.8	W1/3.1/2.8
G_{-3} N7	R203 NH1 (B)	W1/3.1/2.8	W1/3.1/2.8
A_{-2} O1P	K113 NZ (A)	W3/3.7/2.8	W3/2.9/3.1
A_{-2} O1P	R145 NH1 (A)	W3/3.0/2.8	W3/3.1/3.1
A_{-2} O2P	A112 O (A)	W79/3.0/2.5	W79/2.7/2.6
A_{-1} O1P	**H114 N (A)**	**W7/3.6/2.5**	**W7/3.3/2.7**
C_3 O1P	**G129 N (A)**	**E87 /3.2/2.8**	**No W87 A138T model**

[a] Hydrogen bonds were identified with the program HBplus (McDonald et al., 1994). An entry in bold depicts a difference in the contact between the wild-type and A138T models. DNA numbering scheme is:

5'-$T_{-7}C_{-6}G_{-5}C_{-4}\underline{G_{-3}A_{-2}A_{-1}T_1T_2C_3}G_4C_5G_6$.

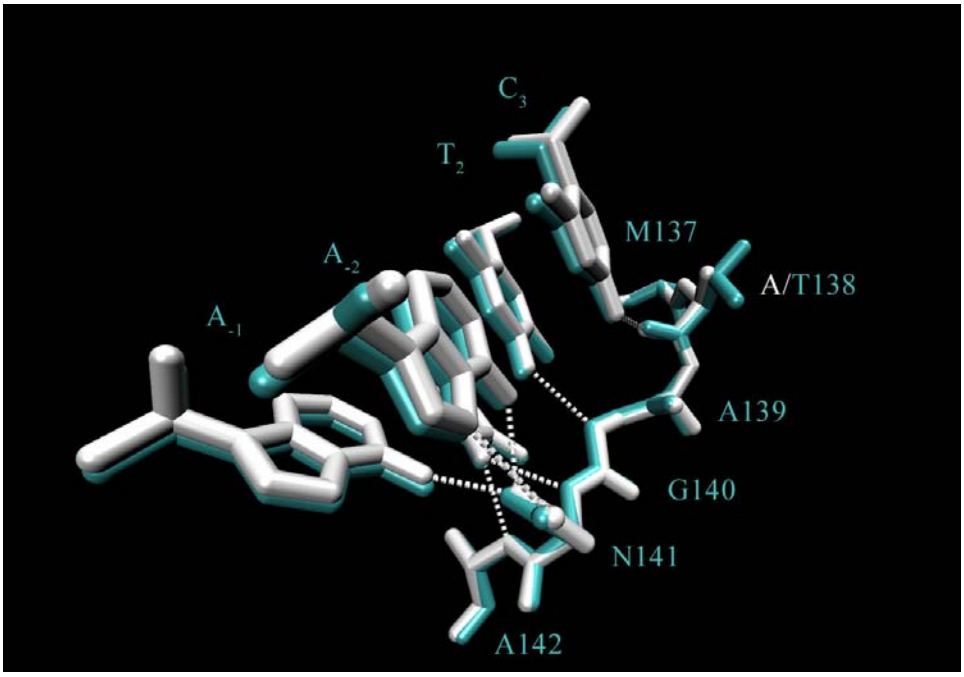


Figure 3.3 Overlay of the extended chain motifs of the wild-type and A138T-d(TCGCGAATTCGCG) complexes.

Wild-type atoms are white and A138T atoms are cyan. The main-chain atoms of residues 137-142 of the two models are superposed here with an rmsd of 0.13 Å. Hydrogen bonds between atoms in the wild-type amino-acids and DNA bases are shown as dotted white lines. The coordinates of the A138T structure also support the presence of these hydrogen bonds in the mutant complex (See table 3.2 for distances). Note that the DNA bases -1&-2 and 2&3 displayed in the figure are in opposite DNA strands.

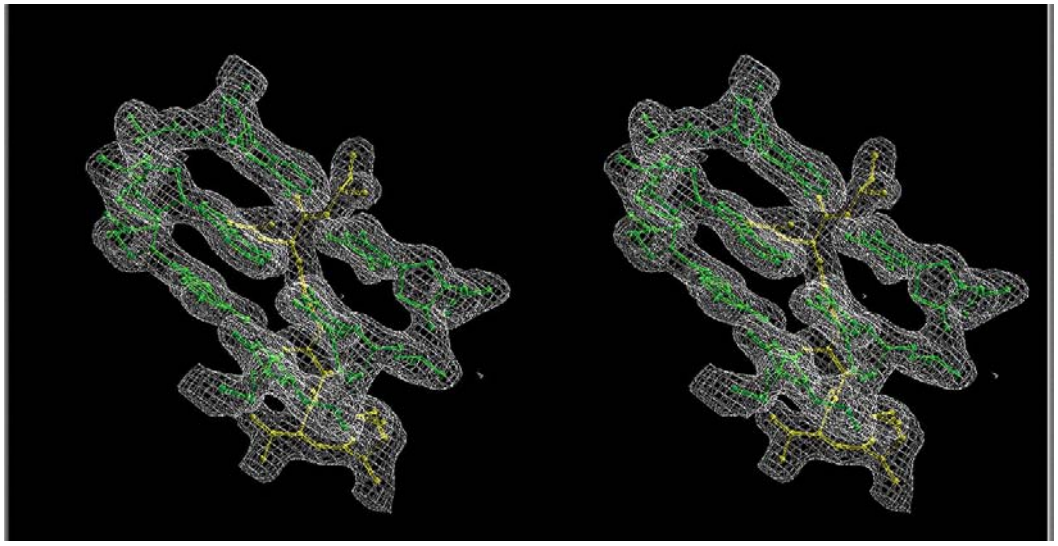


Figure 3.4 Stereo drawing of electron density covering the extended chain motif of the A138T-DNA complex.

The extended chain motif (protein) is colored yellow and the DNA colored green. The map is contoured at 1.5σ. The quality of this electron density map is representative of the quality of the density covering the entire protein-DNA interface.

In the mutant structure, the hydroxyl group of the Thr138 side chain makes contacts, mediated by two different water molecules, to the N7 and O6 functional groups of the guanosine base upstream of the recognition site (G_{-5}). Both guanosine functional groups are hydrated in the wild-type complex (Grigorescu, 2003), but no amino-acid moiety is close enough to bridge the solvent molecule to the DNA (Figure 3.5A&C). Omit maps support the presence of both water molecules in the mutant structural model, however, the molecule solvating the N7 group has stronger associated density than the molecule bridging the interaction with O6 (Figure 3.5B&D). This suggests that the former solvent species may have a higher fractional occupancy in protein-DNA complexes within the crystal lattice than the latter.

To ask if the contact bridged by the higher occupancy (Figure 3.5B&D) solvent molecule is responsible for the improved specific binding of the mutant enzyme, we synthesized oligonucleotides in which G_{-5} is substituted with the base analog, 7-deaza-guanosine (7cG) (Figure 3.6), and measured the binding affinities of the wild-type and mutant enzymes for the modified DNA (Table 3.3). If the isolated contact is responsible for the improved binding of the mutant, we would predict the binding enhancement relative to the wild-type would be reduced with the modified substrate. Surprisingly, both enzymes bind more favorably to the substrate containing the 7cG modification than to the unmodified DNA (Table 3.3). In addition, the substitution results in a larger gain in favorable free energy for mutant complex formation than for wild-type complex formation ($\Delta\Delta G°_{bind,\ A138T}$ = -2.7 kcal/mol vs. $\Delta\Delta G°_{bind,\ Wild\text{-}type}$ = -1.6 kcal/mol). Importantly, a 7cG substitution at a base more distal to the recognition site has no effect on binding in either system (Table 3.3).

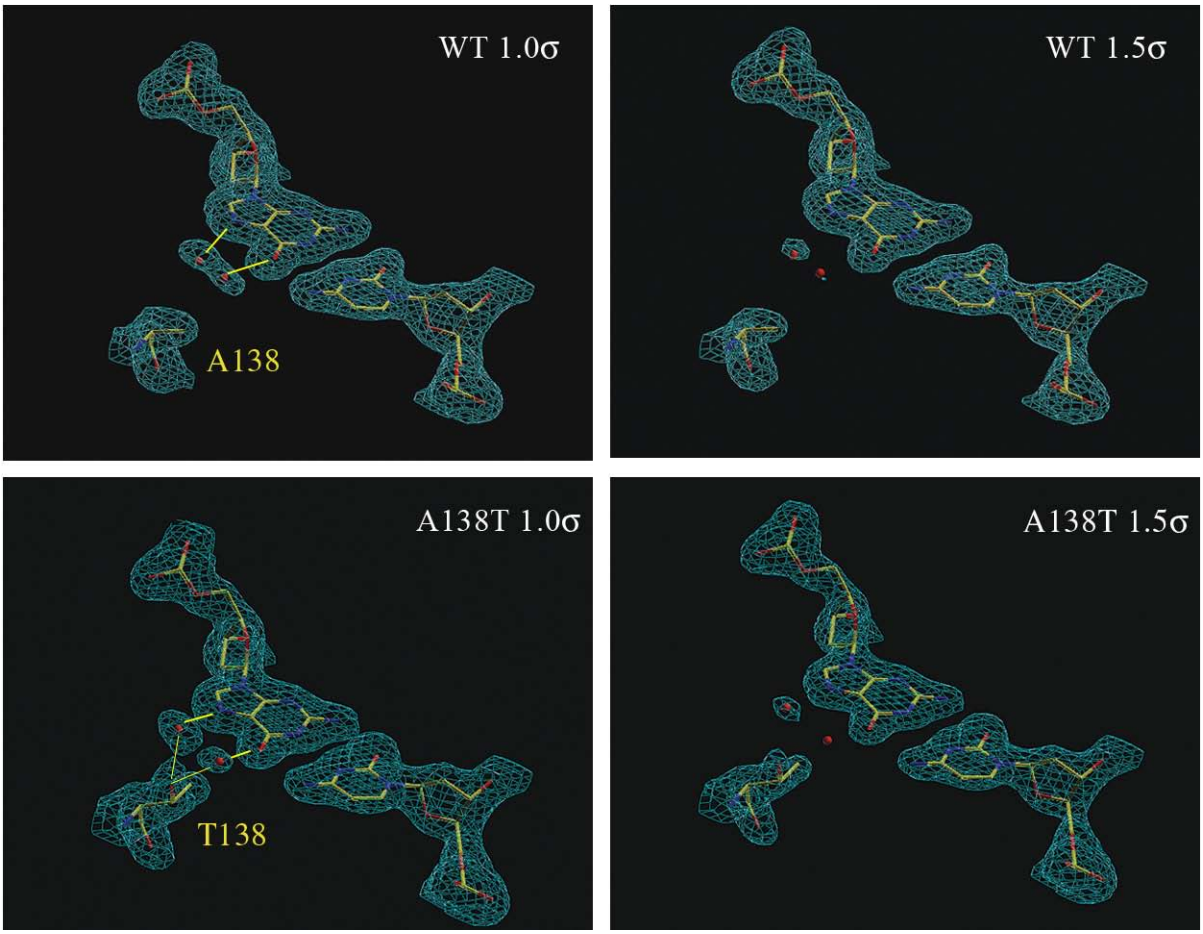


Figure 3.5 Thr138 is involved in novel water mediated interactions with G_{-5}.

Electron density maps are shown covering amino acid 138, the G_{-5}:C_4 base-pair and the water molecules hydrating the O6 and N7 functional groups of G_{-5}. Amino acid 138 and atoms within 3 Å of residue 138 were omitted during map calculation. Maps in the left and right columns are contoured at 1.0 and 1.5σ respectively. Therefore the water molecule hydrating the N7 functional group probably has a higher occupancy or is more spatially restricted than the water molecule hydrating the O6 group. Hydrogen bonds are depicted as yellow lines.

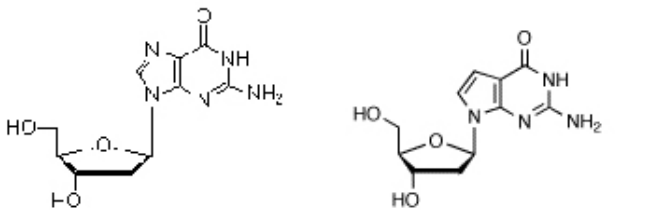


Figure 3.6 Molecular drawings of guanosine and 7-deaza-guanosine nucleosides.

Table 3.3 Effect of 7-deaza-guanosine a substitutions within flanking DNA on EcoRI specific binding

Specific Sequence [b]	K_A [c] (M^{-1})	$\Delta\Delta G°_{bind\ (context)}$ [d] (kcal/mol)	$\Delta\Delta G°_{bind\ (mut.)}$ [e] (kcal/mol)
Wild-type			
GGGCGGGCGCgaattcGCGGGCGC CCCGCCCGCGcttaagCGCCCGCG	7.9 (±0.4)x10^8	0	0
GGGCGGGC**7cG**CgaattcGCGGGCGC CCCGCCCG CGcttaagCGCCCGCG	2.6 (±0.1)x10^9	-0.7 ±0.1	0
GGGCGGGC**7cG**CgaattcG CGGGCGC CCCGCCCG CGcttaagC**7cG**CCCGCG	1.2 (±0.1)x10^{10}	-1.6 ±0.1	0
GGGC**7cG**GGCGCgaattcGCGGGCGC CCCG CCCGCGcttaagCGCCCGCG	5.7 (±1.0)x10^8	0.2 ±0.1	0
GGGC**7cG**GGCGCgaattcGCGGG CGC CCCG CCCGCGcttaagCGCCC**7cG**CG	7.5 (±0.3)x10^8	0 ±0.1	0
A138T			
GGGCGGGCGCgaattcGCGGGCGC CCCGCCCGCGcttaagCGCCCGCG	9.4 (±0.4)x10^9	0	-1.5 ±0.1
GGGCGGGC**7cG**CgaattcGCGGGCGC CCCGCCCG CGcttaagCGCCCGCG	2.3 (±0.3)x10^{11}	-1.9 ±0.1	-2.6 ±0.1
GGGCGGGC**7cG**CgaattcG CGGGCGC CCCGCCCG CGcttaagC**7cG**CCCGCG	8.8 (±1.2)x10^{11}	-2.7 ±0.1	-2.5 ±0.1
GGGC**7cG**GGCGCgaattcGCGGGCGC CCCG CCCGCGcttaagCGCCCGCG	8.9 (±0.1)x10^9	0 ±0.1	-1.6 ±0.1
GGGC**7cG**GGCGCgaattcGCGGG CGC CCCG CCCGCGcttaagCGCCC**7cG**CG	7.3 (±0.2)x10^9	0.2 ±0.1	-1.3 ±0.1

[a] 7cG = 7-deaza-Guanosine

[b] Complete sequence of 24 base-pair oligonucleotide. The specific recognition site, gaattc, is underlined.

[c] Equilibrium association constants were measured in binding buffer plus 0.28 M KCl (pH 7.3, 21°C); means ± std. dev. of ≥ 3 determinations for each sequence.

[d] For each enzyme, the difference in the observed standard binding free energy between the unmodified specific sequence (reference) and each 7cG substituted sequence is calculated as:

$$\Delta\Delta G°_{bind\ (context)} = -RT \ln (K_A^{7cG}/K_A^{Unmod.}) \text{ at } 294 \text{ K.}$$

[e] For each sequence, the difference in the observed standard binding free energy between the wild-type enzyme (reference) and the A138T enzyme is calculated as:

$$\Delta\Delta G°_{bind\ (mut.)} = -RT \ln (K_A^{A138T}/K_A^{Wild\text{-}type}) \text{ at } 294 \text{ K.}$$

Closer examination of the two structural models in the proximity of G_{-5} may partially explain: 1) Why the A138T enzyme binds better to the *unmodified* substrate than wild-type; and 2) Why substitution with 7cG at G_{-5} results in even more favorable DNA binding by the mutant relative to the wild-type. The gamma methylene group of the Thr138 side chain is packed against the N7 group of G_{-5} suggesting a favorable dipole–induced dipole interaction between these two groups in the mutant complex with unmodified DNA (Figure 3.7B). The alanine side-chain in the wild-type model is too far from the DNA to make this contact (Figure 3.7A). This structural difference may contribute to the improved mutant binding affinity for the *unmodified* substrate. It is interesting to note that the A138V mutation also exhibited a relaxed specificity phenotype (Heitman and Model, 1990). Valine, which is isosteric with threonine would be expected to make the Van der Waals interaction with G_{-5}, but not the water mediated interaction. (However, we do not know whether the A138V mutation results in better specific binding than the wild-type). Substitution of the N7 atom of G_{-5} with a carbon atom would result in a hydrophobic interaction between the gamma methylene group of Thr138 and the C7 atom of $7cG_{-5}$. Again, the alanine side-chain would be unable to 'reach' the C7 atom to pack against it without the extended chain motif undergoing some sort of adaptation, which seems unlikely given the importance of the geometry of this motif for direct readout of GAATTC. This may explain why the gain in favorable free energy for binding to the modified DNA is greater for the A138T enzyme than for the wild-type, but we are left unable to explain why the 7cG substitution results in the residual improvement in binding relative to the unmodified DNA for both the wild-type and A138T enzymes. It is possible that the base-analog substitution alters the sequence dependent deformability of the DNA or the guanosine solvation patterns in the free and bound DNAs. It is tempting to suggest that improved binding to the 7cG substituted DNA relative to the canonical substrate derives from a gain in favorable entropy due to the release of water from non-polar surface (C7 atom), however, it is unclear if C7 would be hydrated in the free DNA. Further, if C7 is hydrated in the free DNA, the water molecule may be retained in the complex, as this solvent molecule is part of a hydrogen bond network between the enzyme and the flanking bases (Figure 3.8A).

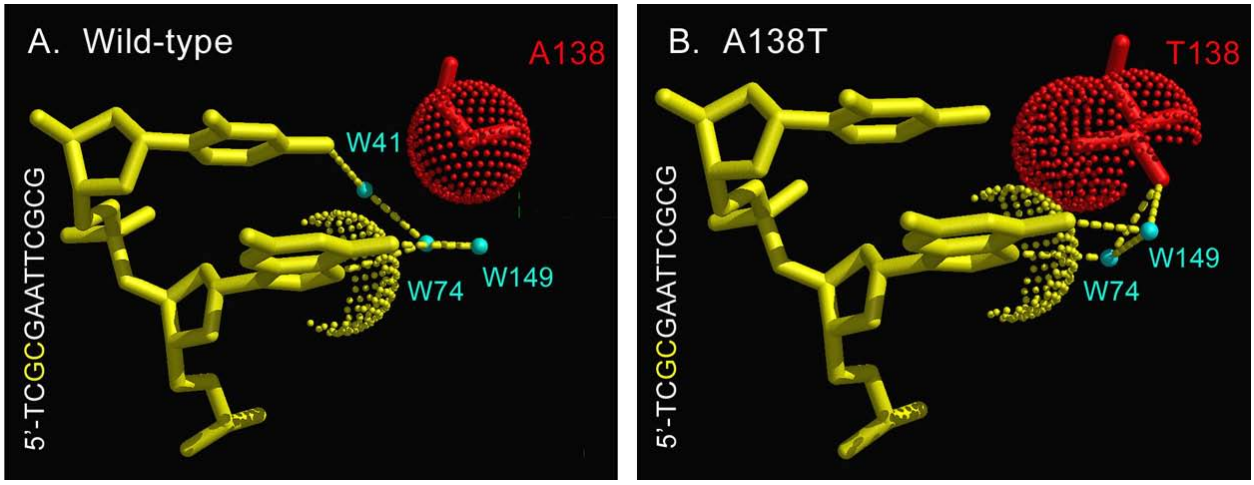


Figure 3.7 Thr138 packs against G_{-5} in the A138T complex.

Panel A, shows the Van der Waals surfaces of N7 of guanosine and Cβ of Ala138 in the wild-type complex. Panel B shows Van der Waals surfaces of the N7 of guanosine and the gamma methylene of T138 indicating a dipole-induced dipole interaction between these two groups. Hydrogen bonds are depicted as dotted lines in both panels.

There is a second significant difference between the wild-type and A138T models involving solvent mediated contacts to GAATTC flanking DNA. In the wild-type model, a water molecule bridges the side-chain of Arg200 and main chain atoms of Ala139 and Ile197 to the 4-amino group of the cytosine immediately upstream of the recognition site (C_{-4}) (Figure 3.8A). The water molecule in the wild-type complex is well ordered as its corresponding electron density is visible in omit maps contoured at 2σ (Figure 3.8C), and the crystallographic B-factor of this atom is similar to the B-factors of waters bridging amino-acid moieties to functional groups within the recognition site. In contrast to the wild-type interface, there is no electron density corresponding to this water molecule in mutant maps (Figure 3.8D). This solvent molecule is expected to be released from the 4-amino group of C_{-4} upon A138T complex formation because the gamma methylene group of the Thr138 side-chain would clash with the solvent molecule if it occupied the same space as in the wild-type complex (Figure 3.8B). Further, the Thr138 side-chain is in Van der Waals contact with the 4-amino group of C_{-4} (Figure 3.8B). Thus, Thr138 directly packs against two groups, N7 of G_{-5} (Figure 3.7B) and the 4-amino group of C_{-4}, which are solvated in the wild-type complex. Therefore, whereas shape complementarity between the wild-type enzyme and the $G_{-5}C_{-4}$ DNA bases relies on ordered solvent molecules to match the surfaces of the two macromolecules, the A138T enzyme contacts

these bases directly. The potential implications of these observations in flanking sequence preferences and the thermodynamic signatures of specific binding will be discussed further in sections 3.2.4 and 3.2.7 respectively.

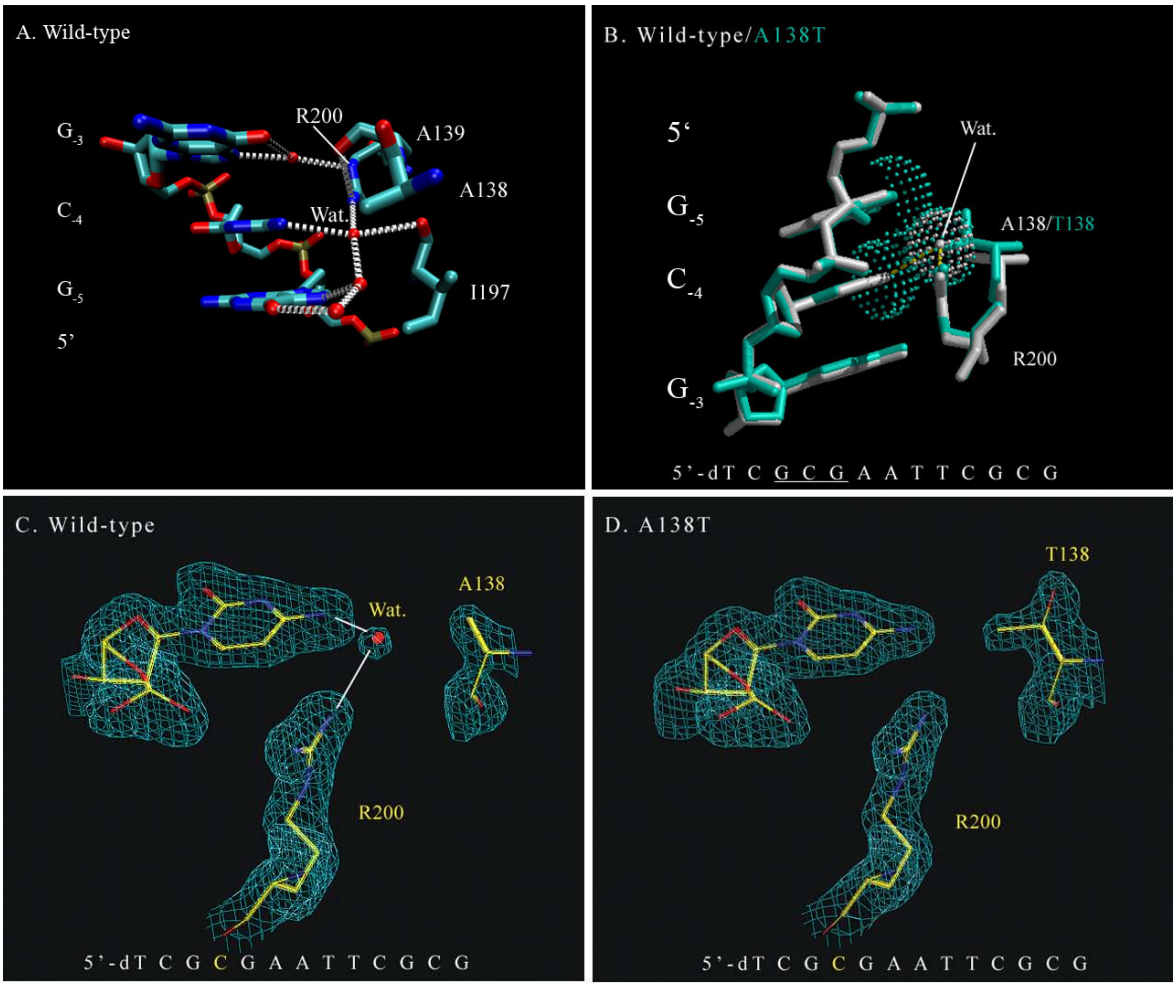


Figure 3.8 A138T complex lacks an ordered solvent molecule that is present at the wild-type-DNA interface.

A) Hydrogen bond network where water molecules bridge interactions between amino-acid main/side-chain groups and the two bases upstream of GAATTC. R200 uses water molecules to contact C_{-4} and G_{-3} (first base of the recognition site).

B) The wild-type (white) and mutant (cyan) models are superposed on one another. Van der Waals representations of the water molecule in the wild-type model and the gamma methylene group of the T138 side-chain in the mutant model show that these two groups would clash if the water were present in the mutant model. In the mutant model, N7 of G_{-3} and 4-NH_2 of C_{-4} are in Van der Waals contact with the gamma methylene group of T138.

C) Omit maps contoured at 2σ show density consistent with a solvent molecule bridging R200 with the 4-amino group of the C_{-4} in the wild-type model.

D) There is no electron density to support a similar (compare with panel C) water molecule at the A138T protein-DNA interface.

3.2.4. Structure of the A138T *Eco*RI-d(TCGCGAATTCGCG) complex provides insight into the changes in GAATTC flanking context preference by the A138T enzyme

The DNA in many site-specific protein-DNA complexes exists in conformations that deviate significantly from the conformation of averaged B-form DNA (Frederick et al., 1984, Schultz et al., 1991, Kim et al., 1993, Winkler et al., 1993). It has been shown in crystallographic, NMR and computational studies of DNA that base-pair steps exhibit sequence dependent differences in the geometry and dynamics (Srinivasan et al., 1987, Lankas et al., 2003). In other-words, certain base-pair steps are more 'deformable' and have different sizes and shapes to their envelopes of accessible conformations than others. It is believed that DNA binding proteins exploit the anisotropic flexibility of DNA to add an additional level of sequence specificity beyond the direct read-out of DNA functional groups in the grooves of DNA (Otwinowski et al., 1988, Dickerson and Chiu, 1997, Olson et al., 1998). That is, site-specific DNA binding proteins have evolved to recognize DNA sites whose anisotropic flexibility favors the formation of a complementary interface with the binding surface of the particular protein. A corollary of this hypothesis is that the DNA deformation energy in going from the averaged B-DNA conformation of the free DNA to the distorted conformation observed in a complex is expected to be minimized for the specific recognition site (Lesser et al., 1990, Lesser et al., 1992, Paillard et al., 2004).

Jen-Jacobson proposed (Jen-Jacobson, 1997) that the anisotropic flexibility of the base-pairs *outside* the GAATTC site also plays a role in binding affinity, and might dictate the range in DNA binding affected by changing the flanking context of the *Eco*RI recognition site. This hypothesis is based on the following observations: 1) Base functional groups within the most favored triplet flanking GAATTC are not directly contacted by the protein with the exception of a Van der Waals contact between Ile197 and C_{-6}; therefore differences in binding affinity are not due to differences in direct protein-base contacts; 2) Although the phosphates upstream of the recognition site are contacted by the protein, ethylation interference patterns for *Eco*RI binding to the GAATTC site embedded in multiple flanking contexts are the same (Michael R. Kurpiewski, *unpublished*). This observation coupled with the fact that the salt dependence of endonuclease binding to specific DNAs in all flanking contexts is also the same (Michael R. Kurpiewski and Paul J. Sapienza, *unpublished*), supports the inference that phosphate contacts in

the different complexes are also unchanged.; 3) DNA cleavage rates, which are exquisitely sensitive to changes in the geometry of the protein-DNA interface, are identical for the GAATTC site embedded in all flanking contexts tested (with the exception of four base-pair A-tract which produces a small inhibition of the cleavage rate constant), indicating that global features of the bound states of all specific complexes irrespective of the flanking context are the same (Jen-Jacobson, 1997). These observations are consistent with the idea that the GAATTC flanking sequence influences the energy required to achieve the distorted DNA conformation observed in the specific complex, and does not alter the direct protein-DNA contacts.

Jen-Jacobson and co-workers observed that the enthalpy ($\Delta H°$), entropy ($\Delta S°$) and heat capacity ($\Delta C°_p$) changes associated with specific binding are correlated with the free energy changes for binding to GAATTC in a number of flanking contexts (Jen-Jacobson, 2000). That is, formation of complexes with the best specific flanking triplet is accompanied by a more favorable enthalpy change, less favorable entropy change and larger (more negative) heat capacity change as compared to binding to sites in sub-optimal contexts. It was proposed that favorable $\Delta H°$ for binding to the best flank compared to the worst flank originates from the lower cost of DNA distortion as well as the optimization of protein-base and protein-phosphate contacts across the entire interface. This optimization of contacts results in a deeper, less "rugged" binding energy well, where molecular vibrations are more restricted and fewer microstates are significantly populated. These factors translate into a less favorable entropy change for binding to the optimal flanking sequence, which is consistent with the thermodynamic data (Jen-Jacobson, 2000). Finally, since the stoichiometry of water release is the same for binding to all flanking sequences tested (Cao, 2001), the observed differences in $\Delta C°_p$ cannot be explained by differences in the number of waters released and/or retained at the interface. Therefore the more negative $\Delta C°_p$ associated with formation of complexes with more favored flanking sequences is proposed to originate from a reduction in vibrational motions across the interface which might have contributions from changes in the coupled folding reaction involving the enfolding "arms".

Recently, the crystal structure of the *Eco*RI restriction endonuclease in complex with GAATTC flanked by a CAC triplet was solved (Grigorescu, 2003) and compared to the structure of the canonical *Eco*RI endonuclease–d(TCGCGAATTCGCG). The CAC flanking context produces a 13-fold inhibition of binding ($\Delta\Delta G°_{bind}$ = 1.5 kcal/mol, Michael R. Kurpiewski,

unpublished) relative to the CGC flank. The protein-base contacts, protein phosphate contacts, and DNA conformations are the same in the two structural models. The only difference involves a reduction in the number of observable solvent mediated interactions between the protein and GAATTC flanking bases in the lower affinity complex. Grigorescu suggested that the water molecules are present in both complexes, but the associated electron density was less visible, due to higher mobility or lower occupancy in the complex with the less optimal flanking sequence (Grigorescu, 2003). It was also inferred from differences in crystallographic B-factors between the two structures that the flanking bases and the residues proximal to these bases are more mobile in the lower affinity complex. Interestingly, an analysis of the *Eco*RV enzyme in complex with its recognition site in two different flanking contexts also showed changes in water mediated contacts to be the only significant difference between the two structures (Horton et al., 1998). Although these studies represent comparisons of structures with only two different flanks, we must consider differences in water mediated complementarity as a factor that contributes to the observed preferences by restriction enzymes for their recognition sites in different flanking contexts. It should be emphasized that the structural findings described above are consistent with the hypothesis of Jen-Jacobson at al. (Jen-Jacobson, 2000) as to the molecular origins of the flanking sequence effect, since the analysis of crystallographic B-factors of the complexes in the two flanking contexts show that the dynamics of the flanking DNA, the hydrating water, and the protein are coupled, suggesting that the protein-DNA interface is more dynamic in the lower affinity complex.

In section 2.2.2, I presented data showing differences in the GAATTC flanking context binding preferences between the wild-type and A138T endonucleases. The energy diagram illustrating these differences is shown again in Figure 3.9. Having determined the structure of the A138T-specific DNA complex, we wanted to analyze the mutant structure and ask: 1) Are there are any differences in DNA conformations between the two models that might lead to the observed differences in flanking sequence preference? 2) Are there additional contacts (water mediated or otherwise) to the DNA bases/backbone flanking the recognition site that might be consistent with the changes in the GAATTC flanking triplet hierarchy?

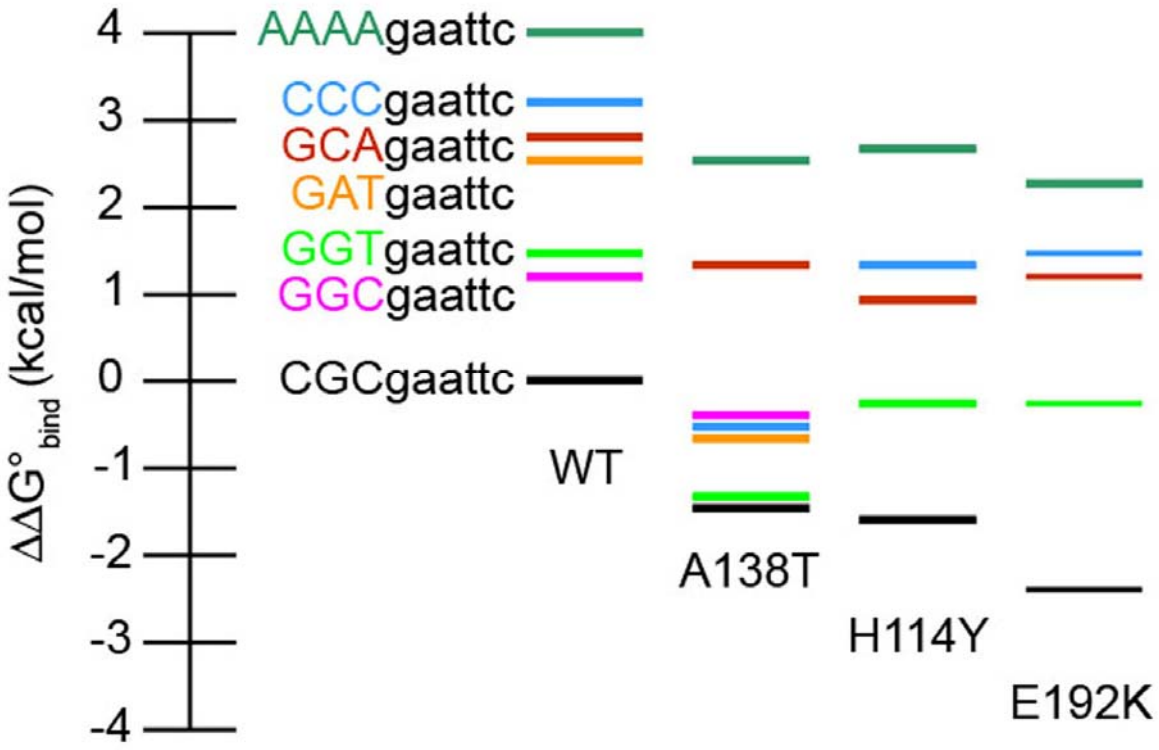


Figure 3.9 Binding of *Eco*RI endonuclease to the specific DNA site in different flanking contexts.

Scale at left has its origin ($\Delta\Delta G°_{bind}$ = 0) at the wild-type interaction with its most favored specific sequence. Similar flanking triplets have bars of the same color in each enzyme's column. Note that only the three bases 5' of the recognition site are shown, however, the three flanking sequences are symmetrical in both strands (see Table 2.1 notes or Appendix A for exact sequences). The wild-type, H114Y and E192K enzymes have the same hierarchy of flanking preferences, but the A138T mutation results in the modification of this hierarchy.

The DNA conformations in the two complexes are very similar; and can be superimposed an rmsd of 0.42 Å. The average rmsd can be subdivided into an rmsd of 0.22 Å for the GAATTC site base-pairs, and 0.54 Å for the six flanking base-pairs (three on each side of the recognition site). Consistent with this global similarity is the fact that the local parameters describing the geometry of base-pairs and base-steps are virtually indistinguishable for the DNAs in the two complexes. (Plots comparing these parameters for DNAs in the wild-type and mutant complexes are shown in Appendix F). The only significant change in a local DNA structural parameter that can be correlated with the A138T mutation is base-pair 'opening' (Figure 3.10). In the wild-type complex, the G_{-5}:C_5 base-pair is opened slightly towards the minor groove,

whereas the same base-pair in the A138T-bound DNA has a base-pair opening value similar to that associated with a base-pair in averaged B-DNA. In consequence of the small change in the orientation of this base-pair, the N7 group of G_{-5} is shifted towards the protein in the mutant complex such that the gamma methylene group of Thr138 can come into Van der Waals contact with the N7 atom (contact shown in Figure 3.7B).

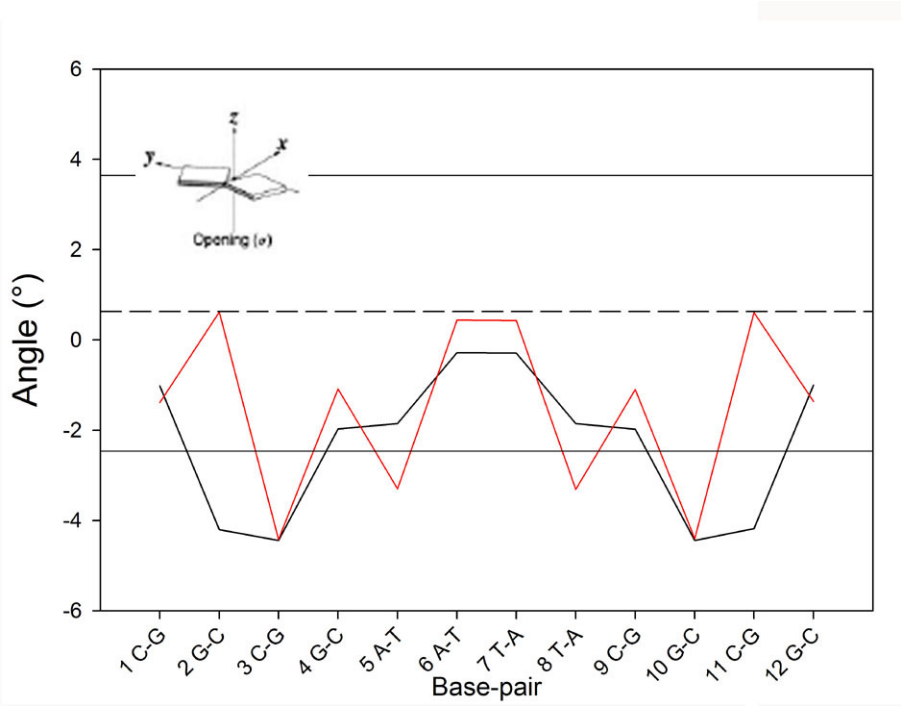


Figure 3.10 "Opening" base-pair parameter for DNAs in wild-type and A138T-specific complexes.

Values for opening were calculated with the program 3DNA (Lu et al., 2003). Base-pair opening values for DNAs in the wild-type and A138T complexes are plotted in black and red respectively. The dotted line represents the average value of opening observed in an analysis of B-DNA structures from the nucleic acid database (Berman et al., 2002). The solid straight lines above and below the dotted line are ± 1 standard deviation from the mean value of opening. The schematic (top left of the figure) provides orientation by shading the minor groove edge of the base-pair.

Interestingly, the flanking triplets for which I observe two of the three largest differences in binding free energy between the wild-type and A138T enzymes ($\Delta\Delta G°_{bind} = \Delta G°_{bind,Mutant} - \Delta G°_{bind,Wild\text{-}type}$), $G_{-6}G_{-5}T_{-4}$ and $G_{-6}A_{-5}T_{-4}$, contain a purine in the -5 position and a thymine in the

-4 position. The mutant's elevated preference for a purine in the -5 position might be explained based on the packing of the Thr138 side-chain against N7 (present in both adenosine and guanosine). Also of note are data presented in chapter 4 which show the effect of flanking context on *miscognate* DNA *cleavage*. Flanking triplets that showed the largest A138T enzyme cleavage rate enhancements relative to the wild-type enzyme all have a purine in the -5 position (data in table 4.4). This observation is consistent with a potential role of the Thr138 contact with N7 of the -5 base in stabilization of the mutant miscognate DNA cleavage transition state.

Recall that the wild-type enzyme prefers binding to the GAATTC site flanked by the optimal CGC triplet over GGT by 1.5 kcal/mol, but the A138T enzyme does not discriminate between CGC and GGT (Figure 3.9, raw data in Table 2.1). The reason for the preference of a thymine at position -4 in the A138T complex is not clear from analysis of the structural coordinates. In the crystal structure, the 4-amino group of C_{-4} is packed against the Thr138 side-chain (Figure 3.8B). A thymine at position -4 would result in replacement of the 4-amino and 5-hydrogen substituents of cytosine with 4-keto and 5-methyl groups respectively. Nitrogen and oxygen atoms have similar Van der Waals radii (Lide, 2004), thus it is reasonable to expect that the Thr138 side-chain could also pack against the 4-keto group of thymine without significant adjustments in the positions of the surrounding atoms.

Table 3.4 Effect of uracil substitutions within GAATTC flanking DNA on *Eco*RI specific binding

Specific Sequence [a]	K_A [b] (M^{-1})	$\Delta\Delta G°_{bind}$ [c] (kcal/mol)	$\Delta\Delta G°_{bind}$ [d] (kcal/mol)
Wild-type			
GGGCGGGCGCgaattcGCGGGCGC CCCGCCCGCGcttaagCGCCCGCG	7.0 (±0.6)x10^8	0	-1.7 ±0.1
GGGCTGGGGTgaattcACCGGAGC CCCGACCCCActtaagTGGCCACG	3.6 (±0.3)x10^7	1.7 ±0.1	0
GGGCTGGGG**U**gaattcACCGGAGC CCCGACCCCActtaagTGGCCACG	1.9 (±0.2)x10^8	0.8 ±0.1	-1.0 ±0.1
GGGCTGGGG**U**gaattcACCGGAGC CCCGACCCCActtaag**U**GGCCACG	1.1 (±0.2)x10^9	-0.3 ±0.1	-2.0 ±0.1
GGGC**U**GGGGTgaattcACCGGAGC CCCGACCCCActtaagTGGCCACG	4.5 (±0.5)x10^7	1.6 ±0.1	0 ±0.1
GGGC**U**GGGGTgaattcACCGGAGC CCCGACCCCActtaagTGGCC**U**CG	3.8 (±0.2)x10^7	1.7 ±0.1	-0.1 ±0.1
A138T			
GGGCGGGCGCgaattcGCGGGCGC CCCGCCCGCGcttaagCGCCCGCG	9.4 (±0.4)x10^9	0	0 ±0.1
GGGCTGGGGTgaattcACCGGAGC CCCGACCCCActtaagTGGCCACG	1.0 (±0.1)x10^{10}	0 ±0.1	0
GGGCTGGGG**U**gaattcACCGGAGC CCCGACCCCActtaagTGGCCACG	7.6 (±0.7)x10^{10}	-1.2 ±0.1	-1.2 ±0.1
GGGCTGGGG**U**gaattcACCGGAGC CCCGACCCCActtaag**U**GGCCACG	3.2 (±0.2)x10^{11}	-2.1 ±0.1	-2.0 ±0.1
GGGC**U**GGGGTgaattcACCGGAGC CCCGACCCCActtaagTGGCCACG	1.1 (±0.1)x10^{10}	-0.1 ±0.1	0 ±0.1
GGGC**U**GGGGTgaattcACCGGAGC CCCGACCCCActtaagTGGCC**U**CG	1.2 (±0.1)x10^9	-0.1 ±0.1	-0.1 ±0.1

[a] Complete sequence of 24 base-pair oligonucleotide is given in the table. The specific recognition site, GAATTC is underlined.

[b] Equilibrium association constants were measured in binding buffer plus 0.28 M KCl (pH 7.3, 21°C); means ± std. dev. of ≥ 3 determinations for each sequence.

[c] For each enzyme, the difference in the observed standard binding free energy between the CGC specific flanking sequence (reference) and each additional sequence is calculated as:

$$\Delta\Delta G°_{bind} = -RT \ln (K_A^{Flank}/K_A^{CGC\ Flank}) \text{ at } 294 \text{ K.}$$

[d] For each enzyme, the difference in the observed standard binding free energy between the GGT flanking sequence (reference) and each additional sequence calculated as:

$$\Delta\Delta G°_{bind} = -RT \ln (K_A^{Flank}/K_A^{GGT\ Flank}) \text{ at } 294 \text{ K.}$$

We have tested whether the presence of the 5-methyl group of thymine within the GGT flanking triplet is responsible for the preferential binding of the mutant enzyme by measuring the effects of removal of the 5-methyl group from base$_{-4}$. If a protein-DNA contact with or solvent release from the thymine methyl group were responsible for the large improvement in binding to GAATTC in the GGT flanking context observed for the A138T enzyme, substitution of T$_{-4}$ with uracil would be expected to selectively reduce the affinity of the mutant complex. However both enzymes bind *better* to the specific DNA with a GGU as compared to a GGT flanking triplet (Table 3.4).

Why might removal of the 5-methyl group from T$_{-4}$ result in improved binding? Evidence suggest that DNA flexibility immediately upstream of the recognition site promotes tight binding by the *Eco*RI endonuclease, and that Uracil$_{-4}$ may improve binding affinity by increasing the DNA flexibility at the U$_{-4}$G$_{-3}$ base step: The best GAATTC flanking sequences have a Y$_{-4}$G$_{-3}$ step (Y = pyrimidine) (Michael R. Kurpiewski and Linda Jen-Jacobson, *unpublished*), and YR (R = purine) steps have been shown to be the most flexible/deformable (Dickerson and Chiu, 1997, Olson et al., 1998). Further, inflexible elements such as A-tracts (Zhurkin et al., 1991) abutting the GAATTC site strongly oppose binding (Jen-Jacobson, 1997). Taken together, these observations suggest that flexibility at the N$_{-4}$G$_{-3}$ di-nucleotide step facilitates tight binding to the GAATTC site. Although T$_{-4}$G$_{-3}$ and U$_{-4}$G$_{-3}$ are both YR steps, several experiments have shown that removal of the thymine 5-methyl group produces increased flexibility at UN base-steps relative to TN. Notably, changing the GA$_4$T$_4$C A-tract to GA$_4$U$_4$C creates changes in electrophoretic mobility consistent with an increase in the magnitude of DNA curvature (Hagerman, 1990); and substitution of the 3'most thymine produces the same effect as substituting all T residues with U. Another study concluded that the thymine 5-methyl group contributes to the rigidity of A-tracts by using nuclear magnetic resonance to demonstrate that removal of this methyl increases base-pair "breathing" (Warmlander et al., 2002). Finally, experiments with the *Eco*RI endonuclease where an A$_{-4}$-U$_4$ base pair was placed immediately adjacent to the GAATTC site (i.e. interruption of A$_{-7}$A$_{-6}$A$_{-5}$A$_{-4}$/T$_4$T$_5$T$_6$T$_7$ with a single T-U base-pair (AAAA/UTTT)) rescues DNA binding affinity significantly relative to the A$_{-4}$-T$_4$ base-pair (David Chi and Linda Jen-Jacobson, *unpublished*). Therefore, we hypothesize that the binding preference for the specific site in the GGU context relative to GGT is derived from the increased flexibility at a functionally important base-pair step in the uracil containing construct.

Early in this discussion I suggested that differences in the sequence dependent deformability of DNA was at least in part responsible for the observed range in binding affinities exhibited by the *Eco*RI endonuclease for specific DNAs in different flanking contexts. Structural and energetic studies point to additional factors that likely lead to the range in binding affinities; among them are differential complementarity between the enzyme and DNA hydration patterns, and differences in the dynamics of the protein and flanking DNA. Although we have only examined a subset of 7/64 possible flanking triplets, the wild-type and A138T enzymes share the same optimum flanking triplet, CGC and the least preferred triplet, AAA among the sequences tested. From this we conclude that many if not all of the same factors affecting wild-type flanking sequence preference also appear to be operative in the A138T mutant system.

We expect the effect of differential deformability on flanking sequence preference to be similar or the same for the wild-type and A138T enzymes since the DNA conformations in the two complexes and global features of the two interfaces are extremely similar. I have therefore focused my discussion of altered flanking sequence preference by the A138T enzyme on contacts to the flanking bases. Direct readout of functional groups is not relevant to wild-type flanking sequence preference because the enzyme does not directly contact the flanking bases. However, the increased length of the Thr138 side-chain (relative to the canonical Ala) is able to reach into the major groove and contact two bases in the GAATTC flanking triplet. I have also focused my discussion on differences in protein-DNA contacts, because these can be "seen" in crystal structures. The technique of X-ray crystallography is less well suited for detecting small differences in coordinates manifested by different degrees of protein-DNA contact optimization, or differences in protein/DNA dynamics; both of these may have significant energetic consequences.

Paillard et al. (Paillard and Lavery, 2004) reported a computational approach for identifying the basis for DNA site selectivity of specific DNA binding proteins. The program uses two energetic terms to identify the optimum recognition sequence for a given DNA binding protein. The first, or direct read-out term, measures the energy associated with formation of protein-DNA contacts to all possible recognition sites. The second, or indirect readout term, is obtained from using a force-field (Cheatham et al., 2000, Giudice et al., 2003) to determine the DNA deformation energies associated with converting all possible sequences from B-DNA to the conformation observed in the crystal/NMR structure of the specific complex. This analysis

showed that for some proteins which significantly distort their DNA site upon binding (*Eco*RV restriction endonuclease and TATA binding protein), the deformation term is at least as important in specifying the recognition site as the protein-DNA contact term. The *Eco*RI endonuclease was not included in this study, but it would be interesting to use the Paillard et al. program to ask:

1) Is the deformation energy associated with achieving the DNA conformation observed in the wild-type *Eco*RI-DNA complex lowest for GAATTC relative to other DNA sites?

2) Are the base-pairs *flanking* the GAATTC recognition site predicted to affect the direct contact or DNA deformability free energy terms or both?

3) Are the calculations predictive of the differences in flanking sequences preference between the wild-type and A138T enzymes?

4) Do the differences in flanking sequence preference between the wild-type and A138T enzymes originate in the DNA deformability, direct contact terms, or both? (My analysis of the A138T structure would indicate that the direct contact term will dominate the differences.)

It would also be interesting to compare structures of the wild-type and A138T enzyme complexes with GAATTC in a flanking context where the difference in binding affinities between the two enzymes is larger than for the CGC triplet ($\Delta\Delta G°_{bind}$ = -1.5 kcal/mol between wild-type and A138T). Structures with the GGT ($\Delta\Delta G°_{bind}$ = -2.9 kcal/mol) or GAT ($\Delta\Delta G°_{bind}$ = -3.1 kcal/mol) flanking triplets are candidates since the larger differences in binding free energy may manifest themselves in more exaggerated differences in structure. I attempted to grow crystals of wild-type and A138T complexes with d(TGGTGAATTCACC) and d(TGATGAATTCATC), with no success. Absence of crystal growth was correlated with poor quality of duplex DNA, and I have not pursued the problem further to ask if poor duplex quality is intrinsic to the DNA sequences or the result of problems with synthesis and/or purification.

Molecular dynamics (MD) simulation of the wild-type and A138T specific complexes represents an alternative method of asking how the enzymes interact with flanking triplets of different composition. It is obviously easier to make the changes to the DNA flanking triplet *in silico* and use the one experimentally determined structure as the starting point in MD simulations than to solve multiple structures by x-ray crystallography. As a control, we would need to verify that we can reproduce the experimentally observed differences between the wild-

type A138T complexes by making the A138T mutation *in silico* and using the wild-type crystal structure as the starting point in MD simulations. Finally, we would like to characterize more A138T 'flanking sequence space' by testing the binding affinity for a greater number of flanking triplets to see if any additional patterns in the differences between wild-type and A138T preferences emerge.

3.2.5. Distal effects of the A138T mutation: Phosphate contacts, the active site and the 'arms'

Comparison of the structures of the wild-type and A138T-DNA complexes shows that the two enzyme form identical hydrogen bonds to the DNA phosphate backbone (Table 3.2). However, two of these interactions appear to be more constrained in the mutant complex. In the A138T model, the amide nitrogen of Asn149 forms a hydrogen bond (3.1 Å) to a non-bridging phosphoryl oxygen at C_{-4}**p**G_{-5}, whereas the same groups are farther apart (3.4 Å) in the wild-type complex, indicating a less constrained bond at the wild-type interface. The difference between these two distances, (0.3Å) is small, but the difference in hydrogen bond donor-acceptor distances for 15 out of 17 protein-base and protein-phosphate hydrogen bonds is 0.2 Å or less. This indicates that the difference in this contact distance cannot be explained by random errors in the two structures. In addition, analysis of the difference in Debye-Waller (B) factors between atoms in the two complexes is consistent with the notion that both the DNA backbone *and* the Asn149 side-chain are less mobile (more positionally constrained) in the mutant complex.

Briefly, it is possible to calculate the expected diffraction from an atom, modified by a Gaussian function that is related to the estimated mean square displacement of that atom in real space. The form of the Gaussian is:

$$e^{(-B\sin^2\theta/\lambda^2)} \qquad (3.1)$$

B is referred to as the Debye-Waller factor, and is related to the mean square displacement of the atom by:

$$B = 8\pi\left\langle x^2 \right\rangle \qquad (3.2)$$

Therefore, the position of an atom with a low B-factor is well fixed in space, and an atom with a high B-factor is more mobile.

There are two main contributors to atomic B-factors; contributions from dynamics of the protein and contributions from static disorder of the crystal lattice. The dynamic contribution can be subdivided into three components: Atomic fluctuations (vibrations), collective motions, and triggered conformational changes (Ringe et al., 1986). We should bear in mind that diffraction data from both wild-type and A138T-DNA co-crystals were collected at 100 K, which is below the glass transition temperature where all internal motions are suppressed except for atomic vibrations (Ringe et al., 2003). Thus, the influence of segmental motions on atomic B-factors does not arise from collective motions over the time course of diffraction data collection, but rather from the spatial distribution of many frozen microstates within the crystal lattice. It has been proposed that B-factors are dominated by concerted rigid body motions, which include the collective motion and lattice disorder components mentioned above (Kuriyan et al., 1991, Kuriyan et al., 1991). There is currently no way to accurately dissect those contributions to atomic B-factors that may be relevant to function (vibrational and collective motions) from confounding lattice disorder contributions.

We took the following steps to make the most accurate comparisons between B-factors in the wild-type and A138T structures. For restrained individual B-factor refinement, the 1.87 Å wild-type data were truncated at 1.95 Å to avoid contributions from differences in resolution to differences in B-factors. Further, the same procedures for B-factor refinement, including use the same weights for B-factor restraints were applied for both structures. One suggested method of placing B-factors from two independently refined structures on the same scale and thereby minimizing the contribution of lattice disorder involves adjusting the B-factors from the structure with the higher values such that the lowest values from both structures are equal (Ringe and Petsko, 1986). In the *Eco*RI complexes, the atoms with the lowest B-factors are at the protein-DNA interface. Since the A138T mutation makes a change at the protein-DNA interface, differences in B-factors between interfacial atoms may result from the mutation itself, and not differences in lattice disorder. As an alternative, we chose to normalize the B-factors to values of the atoms within the hydrophobic core of the proteins. In normalizing the B-factors in this

way, we are making the untested assumption that the hydrophobic cores of the two enzymes have the same thermal stability. (We find it unlikely that an amino change within a protuberance (inner segment of the arm) from the core domain would affect the stability of the core domain).

The average B-factor of a residue within the hydrophobic core of the A138T enzyme is 32.2 $Å^2$, which is 2.7 $Å^2$ higher than the per residue average within the wild-type core. In fact, B-factors for both protein and DNA atoms are on average higher in the A138T structure than in the wild-type structure (Figure 3.11). However there are localized regions of the protein DNA interface where atoms in the A138T structure have *lower* B-factors than the wild-type counterparts. Notably the DNA backbone at the d(TCG**p**C**p**G**p**A**p**ATTCGCG) phosphates exhibits lower mobility in the A138T complex (Figure 3.12). The fact that the B-factors of atoms at both sides (Table 3.5) of the interface are lower in the A138T complex supports the idea that a more optimized contact between Asn149 and d(TCGC**p**GAATTCGCG) (hereafter referred to as pG) is formed in the mutant complex, as a hydrogen bond of increased strength would decrease the vibrational amplitudes of bonds of both the donor and acceptor atoms. Interestingly, the B-factors of the other amino-acid atoms (N of Lys89 and NZ of Lys148) that contact the phosphoryl oxygens at pG are also reduced in the A138T structure (Table 3.5), indicating that multiple contacts to this phosphate are optimized relative to those in the wild-type complex. The B-factors of DNA backbone atoms at the d(TCG**p**CG**p**A**p**ATTCGCG) phosphates are also reduced in the A138T complex. However, I do not observe commensurate reduction in the B-factors of amino acids (Ser87 and Arg203) contacting the d(TCG**p**CGAATTCGCG) phosphate. The ND1 atom of His114 is marginally (by 0.2 Å) closer to O1P of the ApA phosphate in the mutant complex than in the wild-type. One could naively suggest that this decreased separation of the opposite charges would lead to a stronger Coulombic interaction between the two atoms in the A138T complex. We do observe a parallel decreases in the B-factors of ND1 of His114 and O1P of the ApA phosphate, but the B-factor of the protein atom is 50 $Å^2$ (Table 3.5), indicating the slight change in the His114 positions in the two models is not significant (see figure 3.2B).

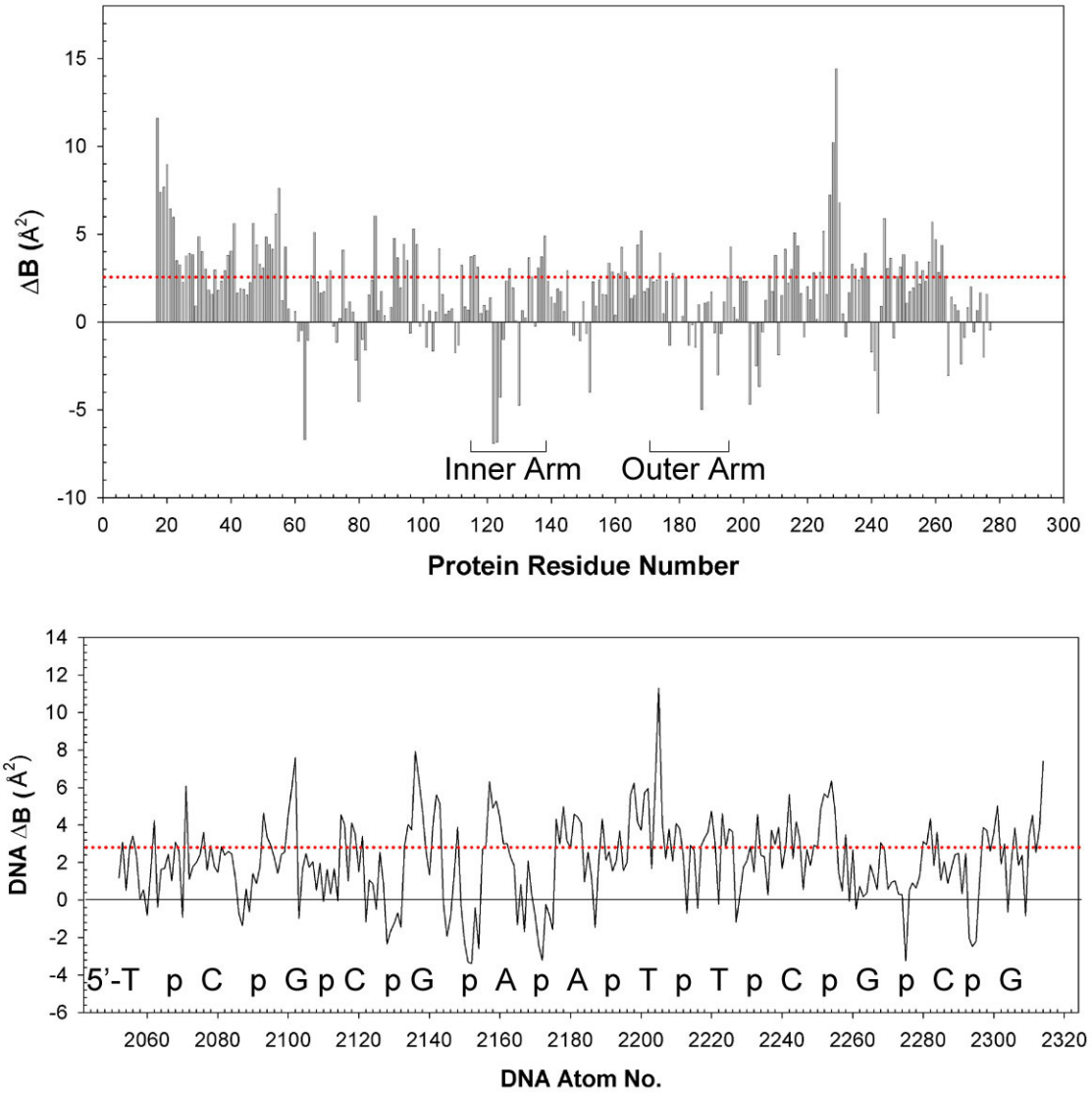


Figure 3.11 Differences in crystallographic B-factors in the structures of the wild-type and A138T specific DNA complexes.

Top Panel) The wild-type per amino-acid residue B-factors were subtracted from those of the corresponding A138T residues (B_{A138T}-B_{WT}).

Bottom panel) B-factors for wild-type DNA atoms were subtracted from the corresponding DNA atomic B-factors in the A138T structure. The red dotted lines in both plots represent the average increase in the B-factor of a residue within the hydrophobic core of the A138T enzyme (+2.6 Å^2).

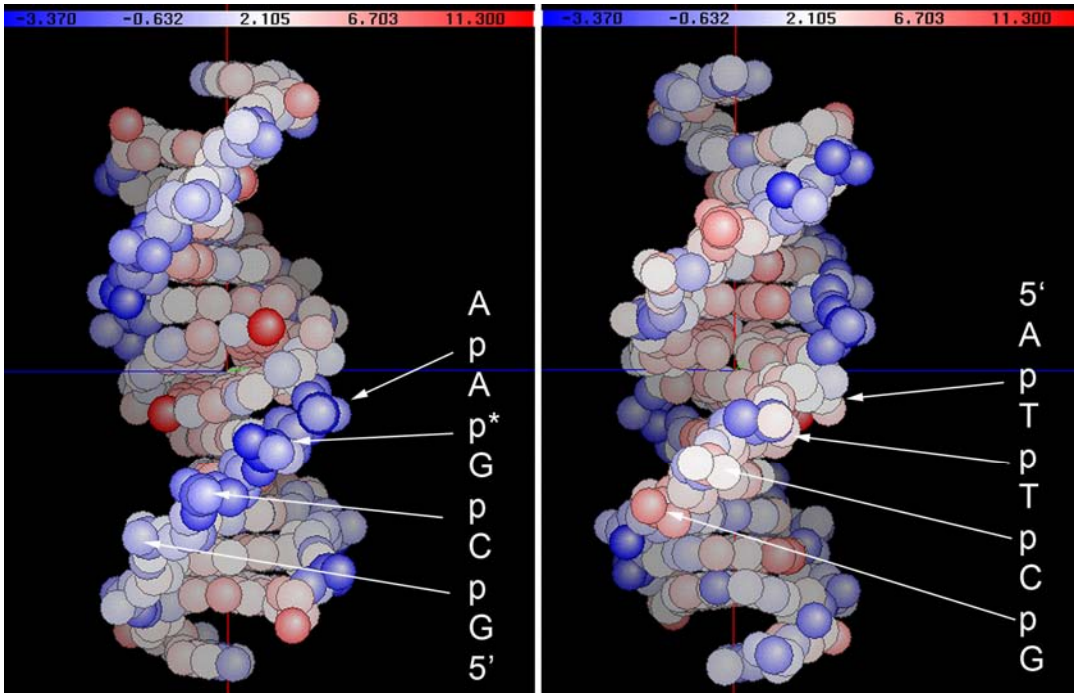


Figure 3.12 Functionally important regions of the DNA backbone exhibit reduced mobility in the A138T–specific DNA complex.

ΔB values ($B_{A138T} - B_{WT}$) for DNA atoms are painted onto the structure of the DNA in the wild-type specific complex using the program GRASP (Nicholls et al., 1991). Atoms colored blue have lower temperature factors in the A138T complex relative to the wild-type complex, and atoms colored red have higher temperature factors in the A138T complex. Dr. Eric Polinko wrote the PERL script used apply the differences in B-factors to the wild-type co-complex pdb file.

Table 3.5 Correlation between changes in Debye-Waller (B) factors at both sides of the pp protein-DNA interface

DNA Atom	B-factor (Å^2)[a]	Normalized ΔB[b]	Contacting Enzyme Atom	B-factor (Å^2)	Normalized ΔB
G_{-3} O1P	31.2	-3.9	K148 NZ	25.0	-5.8
G_{-3} O2P	30.4	-3.4	N149 ND2	27.9	-4.4
G_{-3} O2P	30.4	-3.4	K89 N	29.8	-1.2
A_{-2} O1P	36.0	-6.1	K113 NZ	34.2	-7.6
A_{-1} O1P	41.0	-2.9	H114 ND1	50.52	-3.5
A_{-1} O2P	40.4	-3.5	Wat 176	51.12	No Wat 176 in WT

[b] Atomic B-factor from the model of the A138T-d(TCGCGAATTCGCG) complex

[a] ΔB is the difference in atomic B-factors ($B_{A138T} - B_{wild\text{-}type}$). To normalize the B-factors, the average difference in B-factors of atoms within the enzyme hydrophobic core has been subtracted from the absolute difference in B-factors between the two structures. These normalized values are listed in the table. The DNA numbering scheme is: 5'-$T_{-7}C_{-6}G_{-5}C_{-4}\underline{G_{-3}A_{-2}A_{-1}T_1T_2C_3}G_4C_5G_6$.

The final DNA backbone element for which we see a reduction in mutant B-factors is the scissile phosphate. Here we observe reduced B-factors for both of the phosphoryl oxygens, and a contacting atom, NZ of the active site residue, Lys113 (Table 3.5). The reduced mobility of these two groups is consistent with the formation of a novel hydrogen bond that is formed between Lys113 and the scissile phosphate in the A138T complex (Table 3.2). The difference between the wild-type and mutant structures derives from a $\chi 4$ rotation in Lys113 that places the NZ atom of the side-chain in position to make a hydrogen bond (2.8 Å) with O1P of the scissile phosphate in the A138T complex. The distance (3.8 Å) and angle between these two groups in the wild-type structure (Figure 3.13A) support a Coulombic interaction, not a constrained hydrogen bond. Importantly, electron density maps support the difference in Lys113 side-chain conformation between the two models (Figure 3.13B&C). It is interesting to note that conformation of Lys113 in the A138T metal free complex resembles the conformation of this amino-acid observed during molecular dynamics simulations of the wild-type complex in the presence of the Mg^{++} co-factor (Kurpiewski et al., 2004), during which the Lys113 side-chain rotates to form a hydrogen bond with O2P of the scissile phosphate within 100 picoseconds of adding the Mg^{++}. The interaction of this positively charged residue (along with Arg145) with the scissile phosphate was proposed to stabilize the phosphorane intermediate during catalysis (Kurpiewski et al., 2004). Since the first order rate constant for cleavage of specific DNA is on the order of seconds, and rotation of Lys113 occurs on the order of 0.1 nanoseconds, the necessity for this side-chain rotation in the wild-type complex probably does not contribute to the slower rate of specific DNA cleavage by the wild-type compared to the A138T enzyme (Table 2.2). The B-factors of the active-site residue, Glu111 are also lower (average B= 38 $Å^2$, normalized ΔB= -3.8 $Å^2$) in the A138T complex compared to the wild-type. The WHATIF (Hooft et al., 1996) x-ray model quality check program finds that Glu111 adopts an unusual rotamer in the wild-type complex, whereas this flag is not raised upon analysis of the mutant complex. These two observations are consistent with the idea that further neutralization of the scissile phosphate by Lys113 in the mutant complex reduces the mobility of Glu111 and permits this residue to adopt a less strained rotamer.

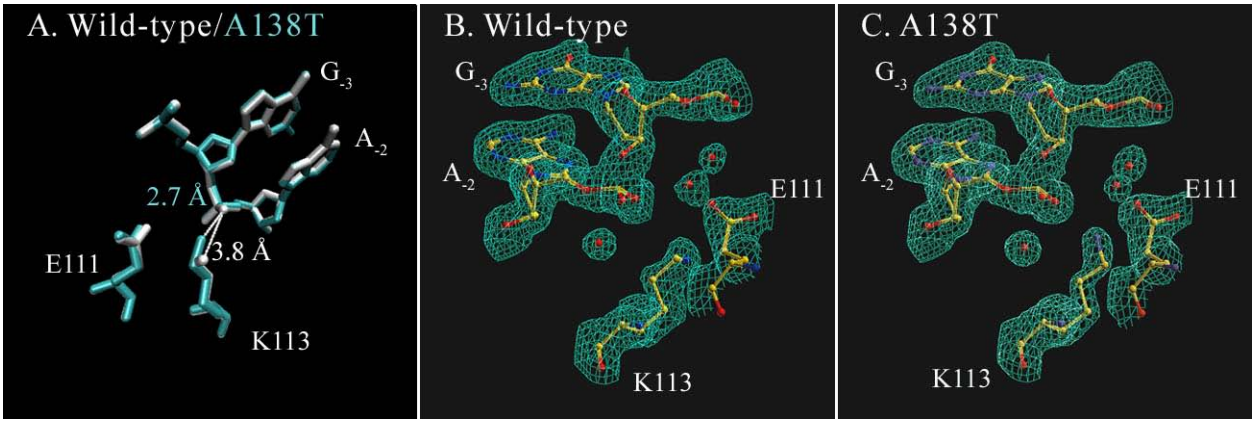


Figure 3.13 Hydrogen bond between Lys113 and the scissile phosphate in the A138T-specific DNA complex.

A) Comparison of the wild-type (white) and A138T (cyan) models shows that the active site residue, K113 makes a hydrogen bond with a non-bridging oxygen of the scissile phosphate in the mutant model. The two atoms are farther apart in the wild-type complex, consistent with a less constrained interaction.

In panels B and C, electron density omit maps (1.5σ contour, with K113 and a surrounding 5 Å sphere of the model omitted from map calculation) covering atoms that make up the active site (remaining active site residue, D91 is not shown in the figure for clarity) show that the K113 conformations in the models are supported by the density.

There are additional enzyme elements that show localized differences in B-factors between the wild-type and mutant structures (Figure 3.11), including the N-terminus and the loop comprised of residues 224-229; both of which have larger than average *increases* in atomic B-factors in the A138T complex. B-factors for atoms within these elements of both structures are very high, most likely indicating that the differences in B-factors are insignificant. Interestingly, B-factors of residues within the inner and outer segments of the enzyme's arms are reduced in the A138T complex. The average normalized B-factor within the α-helix composed of amino acids 118-123 of the inner arm is 6 $Å^2$ lower than the corresponding B-factors in the wild-type complex. The outer segment of the arm also appears to be more ordered in the mutant complex than in the wild-type complex. It should be pointed out that the average B-factors of the arms in the A138T complex are still very high (~50 $Å^2$), which suggests that the arms have significant mobility in both the wild-type and mutant complexes. Although the B-factors of the residues within the arms are high, and the B-factor differences between the two models relatively small, we feel the differences in this case may be real. First the reductions in B-factors are

localized, and occur in both arms. Grigorescu provided evidence which suggests that the inner and outer segments of the arms both fold cooperatively upon binding to specific DNA as a lattice contact involving only a four amino-acid segment of the outer arm resulted in stabilization of both segments in the absence of DNA (Grigorescu, 2003). Thus the fact that we observe reductions in B-factors in both segments supports the idea that the collective motions of the arms may be altered by the A138T mutant. Also, it is reasonable for changes at amino-acid 138 to affect the dynamics of the arms since this residue lies at the base of the inner arm. In chapter 5, I propose experiments designed to probe the conformation and dynamics of the arms in the different (specific, miscognate, non-specific) complexes.

Most of the differences in atomic coordinates and B-factors between the two models are relatively modest and occur in regions that are distal to the site of the A138T mutation. It is not clear how an amino acid change within the inner segment of the arm affects the conformation of the β-sheet at the base of the arm (contains residues contacting the active site and the $A_{-2}pA_{-1}$ phosphate), or the α-helix at the base of the inner arm (contains the amino acids recognizing pG_{-3}). However, it is not reasonable to reject the notion that these changes are somehow attributable to the mutation simply because we cannot explain how the mutation's effect is propagated to these distal sites. Consistent with this idea is a study showing that a host of amino amino-acid changes affected similar regions of a given protein's structure, regardless of the spatial or sequential distance from the mutation (Sinha et al., 2001).

We should point out that there are more trivial alternative hypotheses that could also explain these small differences. It must be borne in mind that the comparisons made between the two structures are based on data collected from a single co-crystal of each enzyme-DNA complex. We do not know how the rotamers of the side-chains in question vary among models derived from data collected from different crystals of the wild-type or mutant complexes. Slight changes in crystallization or cryoprotection buffers, size of the crystal (leading to different kinetics of crystal temperature change during freezing), or other factors may very well contribute to some conformational variability from crystal to crystal. In addition, x-rays are diffracted from crystals where the dynamic nature of macromolecules (Frauenfelder et al., 1991) dictates that the conformations of the macromolecules are heterogeneous from unit cell to unit cell. The resolution of the diffraction pattern is usually insufficient to model this heterogeneity; therefore crystal structures are usually presented as a single conformation that best fits the diffraction

pattern. (This is true in the case of the models of the wild-type and A138T-DNA complexes). Unfortunately, the inability to properly account for conformational heterogeneity introduces degeneracy into to the structure determination processes. This has been confirmed by the analysis of several test cases where many models differing in both side-chain and main-chain conformations fit a diffraction data-set equally well according to the existing criteria for model accuracy (DePristo et al., 2004). Thus small conformational differences among structures must be viewed with some skepticism.

Along the same lines, it is possible that the observed changes in B-factors represent an artifact of refinement. The refinement process involves minimization of the differences in the coordinate parameters (atomic x, y, z Cartesian coordinates and B-factor) to satisfy a maximum likelihood residual. Therefore, local coordinate parameters may be incorrect at the expense of the 'global' minimization process. Two lines of evidence argue against this as an explanation for some of the differences in B-factors between the two models. First, the changes in B-factors discussed in this section involve examples in which the B-factors for localized regions of the A138T model are *lower* than those of the wild-type, while B-factors are on average, *higher* in the mutant complex. Second, the correlation between B-factor changes for atoms at both sides of the protein-DNA interface suggests that the changes originate from reduction in the mobility of the atoms rather than an artifact of refinement.

We do have the luxury of having collected another high resolution data set (2.1 Å) from a different crystal of the A138T specific complex. An independent refinement of this data-set would help address the question as to whether various changes in conformation and atomic B-factors are reproducible from crystal to crystal. Unfortunately, we would need to refine models derived from multiple data sets collected from many A138T *and* wild-type co-crystals in order to get a feel for whether B-factor and coordinate differences can be attributable to the mutation or one of the more trivial factors discussed above.

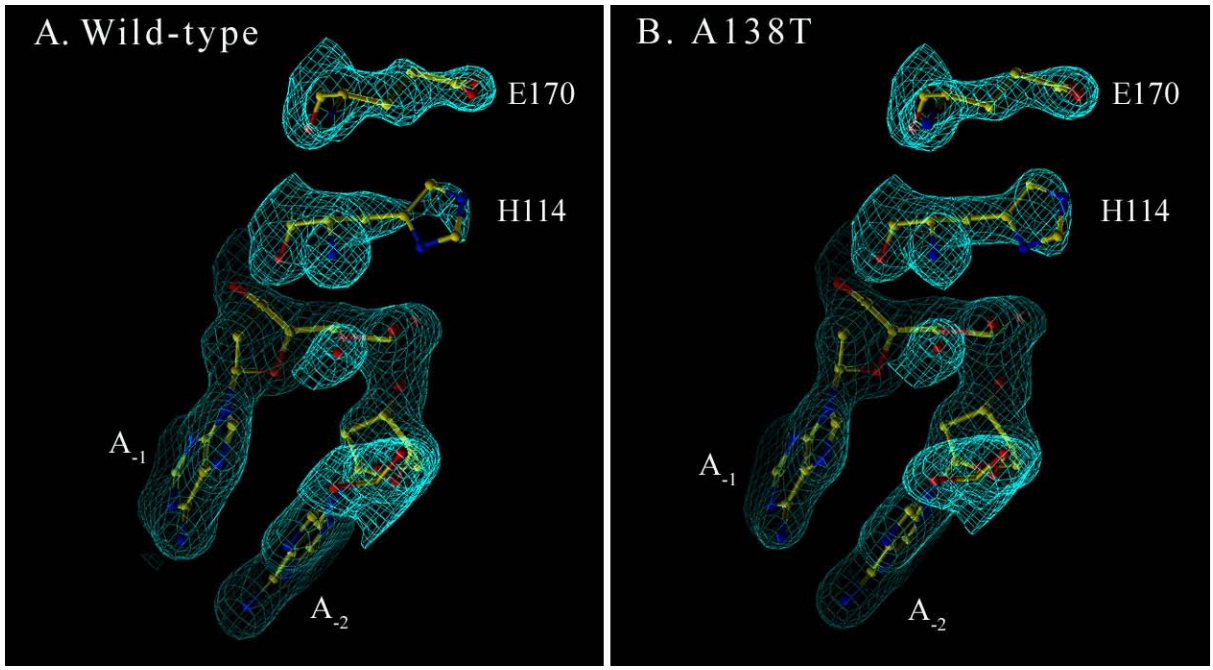


Figure 3.14 The side-chain of histidine 114 is more ordered in the A138T complex than in the wild-type complex.

Electron density maps were calculated while omitting atoms within 4 Å of and including H114. The maps are contoured at 1.0σ in both panels. There is very little electron density for the H114 side-chain in the wild-type map, whereas density corresponding to the imidazole ring is present in the A138T map.

3.2.6. Distal effects of the A138T mutation: pH dependence of specific binding and potential crosstalk with histidine 114

In examining the structure of the A138T specific complex, I wanted to ask if the A138T mutation affects the structure vicinal to other residues implicated in relaxed specificity by the Heitman and Model screen. The structure of the outer arm around residues Gle192 and Tyr193 (E192K and Y193H mutations confer relaxed specificity) is unchanged within the expected error of the coordinates of the two structures. The B-factors of these residues are lower in the A138T complex, but the side-chains remain solvent exposed and flexible. Histidine 114, like Ala138, lies at the base of the inner arm segment, and is therefore a more likely candidate to be affected by the A138T mutation. A change around residue 114 is also reasonable since I already demonstrated in the previous section how the A138T mutation is propagated to a residue next to His114, Lys113. A difference in the conformation or dynamics of His114 in the two complexes

is supported by omit maps (Figure 3.14) showing very little electron density corresponding to the imidazole side-chain visible in wild-type maps, indicating that this side-chain is disordered. In contrast, there is electron density corresponding to the His114 side-chain in mutant maps.

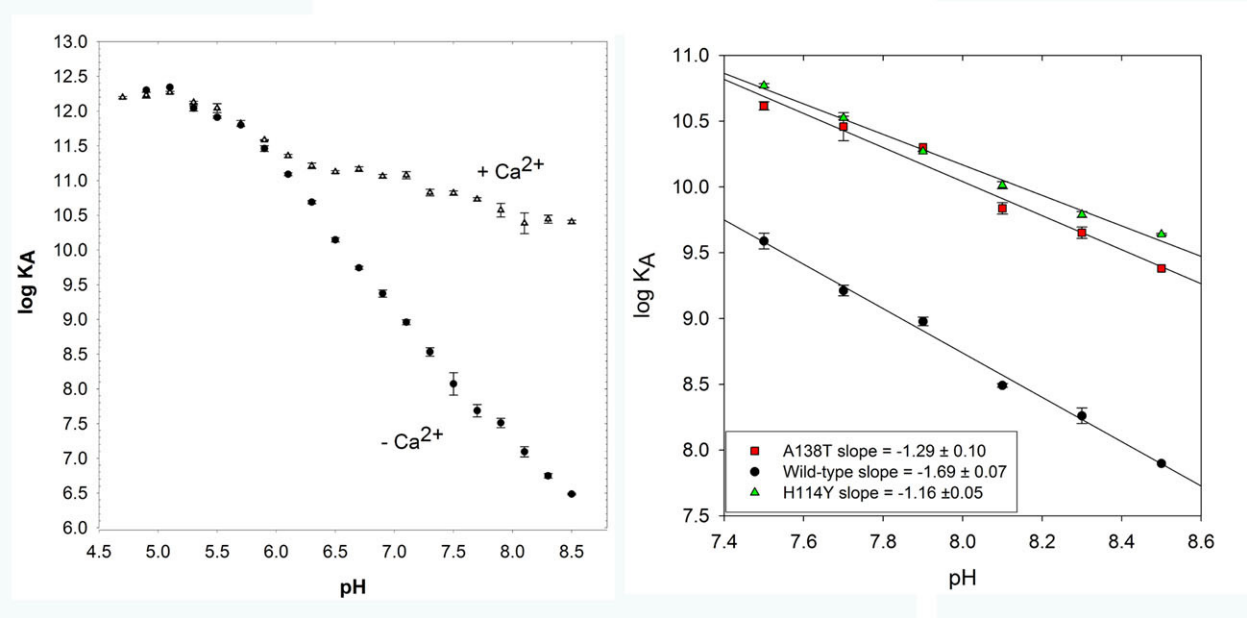


Figure 3.15 pH dependence of the *Eco*RI-specific DNA interaction in the presence and absence of calcium.

Left Panel) pH dependence of the wild-type binding to GAATTC (24 base-pair oligomer in the GCA flanking context; see Appendix A for complete sequence). Reactions were buffered by citrate for pH ≤7.5 and HEPES for pH ≥7.5. Binding constants in both buffers were found to be the same at pH 7.5. Conditions were: (-$CaCl_2$): Binding buffer (0.24M KCl, 21°C), (+$CaCl_2$): Binding buffer (0.23M KCl, 5mM $CaCl_2$, no EDTA, 21°C).

Right Panel) pH dependence of the wild-type, A138T, and H114Y are shown in the absence of $CaCl_2$ over the restricted pH range of 7.5 to 8.5. The specific sequence in the CGC flanking context (Appendix A for complete sequence) was used for this set of experiments.

The pH dependence of *Eco*RI endonuclease specific DNA binding may also indicate a change in the environment of His114 in the A138T mutant complex. The *Eco*RI endonuclease exhibits a steep pH dependence for specific binding; where binding affinity increases as solution pH is decreased (Figure 3.15A), indicating that a net uptake of protons in the complex is coupled to DNA binding (Kozlov and Lohman, 2000, Lundback et al., 2000). It has been shown in the *Bam*HI system that the pH dependence of binding is virtually eliminated if either calcium is

included in the binding reaction, or an acidic residue in the active site is mutated to alanine (Lisa E. Engler, Jennifer Ames, Martina Lefterova, Oanh Pham, and Linda Jen-Jacobson, *unpublished*). X-ray crystal structures reveal that the scissile phosphate and acidic amino acids that constitute the active site provide the ligands that coordinate Ca^{2+} in the specific DNA complexes of *Bam*HI (Viadiu and Aggarwal, 1998) and *Eco*RV (Horton et al., 1998), among others, although the metal does not permit cleavage (Engler et al., 2001). We have deduced that Ca^{2+} also binds to the active site in *Eco*RI complexes based on the observation that the metal is a potent inhibitor of DNA cleavage (K_I = 29 μM at 0.1 M NaCl; plot in Figure 5.3). Therefore, the fact that pH dependence of *Eco*RI specific binding is severely reduced in the presence of Ca^{2+} suggests that protonation of the acidic active site residues is largely responsible for the observed pH dependence (Figure 3.15). The pH dependence also indicates that the titratable groups have pKa values in the 6-9 range; normal Glu and Asp residues have pKa values near 4. We have hypothesized that the highly negatively charged microenvironment of the active sites of restriction endonucleases is responsible for the anomalously high pKa values of these side-chains. I find that slopes of log K_A vs. pH in the pH 7.3 – 8.5 range are different for wild-type and A138T EcoRI binding to DNA (Figure 3.15B), where wild-type binding is more strongly dependent on pH in this range, indicating a difference in the pKa(s) of a residue(s) between the two complexes. The relaxed specificity H114Y mutant shows a shallower slope for log K_A vs. pH than the slopes for wild-type and A138T binding (Figure 3.15B). This could mean that the pKas of active site amino acids are shifted even higher in the A138T and H114Y complexes, or that titration of the imidazole side-chain of His114 is a contributor to the pH dependence of binding in the pH 7.3-8.5 range in the wild-type and A138T *Eco*RI systems. The side-chains of free histidines typically have pKa values around 6, but it is not unusual for the microenvironment to affect the pKa of histidines in proteins (Lundback et al., 2000, Spitzner et al., 2001, Klingen et al., 2004). Histidine 114 lies at the protein-DNA interface, about 4 Å from the negatively charged DNA backbone; thus an anomalously high pKa value for this side-chain is reasonable. We have data which show that His114 is more ordered in the A138T complex (Figure 3.14), which is consistent with this amino acid having a higher pKa in the mutant complex due to a more fixed position of the side-chain, closer to the negatively charged DNA backbone. However, since the H114Y mutation does not completely abolish the pH dependence in the pH

7.3-8.5 range, a change in the pKa other residues might be responsible for the differences in pH dependence slopes between the wild-type and mutant proteins.

3.2.7. Hypotheses on the origins of the differences in ΔH° and ΔS° accompanying specific DNA binding by the A138T and wild-type enzymes

The goal of free energy dissection is to relate the components, $\Delta H°$, $\Delta S°$ and $\Delta C°_p$ to the underlying structural and dynamic observables, with the ultimate aim of using the information to aid in rational protein engineering and drug design. Therefore, systems for which structural and thermodynamic data are available represent excellent opportunities to make progress towards the above aim. In this section, I will discuss the changes in the thermodynamic parameters associated with formation of the wild-type and A138T specific complexes (Table 3.6) in the context of the observed differences between the two structural models. Importantly, the DNA sequences used in Van't Hoff and crystallographic studies are essentially the same. The flanking sequences are the same (CGC triplet), with the only difference being that the DNA used in the solution experiments is a 24mer and the DNA in the crystal structures are dodecamers with a 5' dangling thymine. The difference in DNA length is insignificant as the *Eco*RI endonuclease binds equally well to the 13mer or 24mer containing the specific site (Michael Kurpiewski and Arabela Grigorescu, *unpublished*).

Table 3.6 Thermodynamic parameters accompanying *Eco*RI-specific DNA binding at 25°C [a]

	ΔC°p (kcal/mol K)	T_H	T_S	ΔG° (kcal/mol)	ΔH° (kcal/mol)	T ΔS° (kcal/mol)
WT	-1.84 ± 0.04	289.8 ± 0.1	296.9 ± 0.1	-13.1	-15.4 ± 0.4	-2.3 ± 0.2
A138T	-1.90 ± 0.04	291.4 ± 0.1	299.1 ± 0.2	-14.6	-12.8 ± 0.3	1.8 ± 0.3
E192K	-1.79 ± 0.08	292.4 ± 0.2	301.0 ± 0.3	-15.4	-10.3 ± 0.5	5.1 ± 0.6
H114Y	-2.13 ± 0.05	291.8 ± 0.1	298.7 ± 0.1	-14.7	-13.5 ± 0.4	1.2 ± 0.3
	ΔΔC°p [b] (kcal/mol K)			ΔΔG (kcal/mol)	ΔΔH (kcal/mol)	TΔΔS) (kcal/mol)
WT	0			0	0	0
A138T	-0.06 ± 0.06			-1.5	2.6 ± 0.5	4.1 ± 0.4
E192K	0.05 ± 0.09			-2.3	5.1 ± 0.6	7.4 ± 0.6
H114Y	-0.29 ± 0.06			-1.6	1.9 ± 0.6	3.5 ± 0.4

[a] Equilibrium association constants were measured as a function of temperature and fit to the van't Hoff (equation 2.12, see text) in order to obtain $\Delta C°_p$, T_H and T_S. These values were then plugged into equations 2.13 and 2.14 in order to calculate the ΔH° and ΔS° values for the binding reactions at 25°C. Binding constants were measured using the oligonucleotide 5' GGGCGGGCGCGAATTCGCGGGCGC in 10mM cacodylate, 0.27M KCl, pH 7.3.
[b] All ΔΔ parameters are referenced to the wild-type.

The major differences between the wild-type and mutant structures that can be directly attributed to the A138T amino acid change and are also likely to affect the thermodynamics of the protein DNA interaction are changes in the water mediated contacts between the enzyme and the DNA bases flanking the recognition site. Well ordered water molecules are often observed bound to the surfaces of macromolecules and trapped at the interfaces of intermolecular complexes. It was believed that water, due to it's small size and dual hydrogen bonding character allows for a level of adaptability at a recognition interface, but water can also serve to increase the affinity and specificity of macromolecular interactions (Janin, 1999). Calculations by Dunitz (Dunitz, 1995) showed that the net contribution of an isolated hydrogen bond to binding free energy is close to zero because the favorable enthalpic contribution is opposed by the unfavorable entropy associated with the restriction of rotational, translational and vibrational degrees of freedom of the groups involved in the bond. The same enthalpy-entropy tug of war is expected when water mediated interactions exist at a binding interface between two molecules.

A major question is what effect does the release and/or burial of tightly bound water molecules have on the thermodynamic parameters associated with macromolecular interactions? An upper limit on the entropic penalty associated with 'freezing' a water molecule at a molecular interface was proposed based on a comparison of the standard entropies of a number of

crystalline hydrates with the standard entropies of the corresponding anhydrous salts (Dunitz, 1994). From this analysis, a maximum value of 2.1 kcal/mol (295.15 K) was derived for the transfer of a tightly bound water to bulk solvent. A similar approach was followed in order to get a handle on the enthalpic signature associated with the release of a bound water molecule (Connelly, 1997). The result is an estimate of 3.8 kcal/mol for the enthalpy of transfer of a water molecule from a crystalline salt hydrate to bulk water. The authors pointed out the entropic limit obtained from analysis of the crystalline salt hydrate thermodynamic data is probably applicable to macromolecular systems because the number of degrees of freedom of a bound water molecule is expected to be independent of environment. Conversely, accurate bounds for the enthalpy of water release from macromolecules are more difficult to assign because the enthalpy of a bound solvent molecule is likely dependent on the local dielectric of the environment for example. Further, the enthalpy associated with release of the molecule depends on the number and nature of the intermolecular hydrogen bonds involving the water molecule in the bound state. From these analyses, it is clear that opposing entropic and enthalpic factors are also at work in determining the net free energy associated with trapping or releasing of a tightly bound water molecule.

In section 3.2.3 I showed that water molecules bridge interactions between the Thr138 side chain and G_{-5}; these bridging interactions are not present within the wild-type interface (Figure 3.5). In addition, the N7 atom of G_{-5} is in Van der Waals contact with the Thr138 side-chain (Figure 3.7A), whereas the Ala138 side-chain does not contact the DNA (Figure 3.7A). In proposing the thermodynamic consequences of these observations, we must bear in mind that the structure of the *free* DNA dodecamer shows two solvent molecules bound to G_{-5}, one interacting with N7 and the other to O6 (Shui et al., 1998), and these water molecules are also observed in the specific complexes formed by *both* enzymes (Figure 3.5). Thus, the entropic penalties associated with trapping these two solvent molecules at both interfaces should be about the same. The fact that these waters bridge a contact between the Thr138 side-chain and G_{-5}, coupled with the Van der Waals contact to N7 in the mutant complex would lead to a more favorable $\Delta H°$ for the binding reaction involving the mutant enzyme. We should bear in mind that experiments with the 7cG base-analog (Table 3.3) do not support a role of the water-mediated hydrogen bond to N7 in enhancing binding affinity of the mutant; but are consistent with the existence of the Van der Waals contact in the mutant complex.

As a guide to aid in the estimation of the enthalpic contribution of a Van der Waals interaction, we can use the extensive studies of interactions between DNA binding proteins and C5 methyl groups of thymines. These interactions have been shown to be worth between 0.6 and 1.6 kcal/mol in binding *free energy* (Goeddel et al., 1977, Dubendorff et al., 1987, Takeda et al., 1989, Jen-Jacobson, 1997). The question then becomes whether this interaction free energy should be ascribed to favorable enthalpy of the Van der Waals interaction or favorable entropy of solvent release from non-polar surface? The free energy of methane-methane interaction in the gas phase is around 0.2-0.4 kcal/mol (Novoa et al., 1991), which is insufficient to account for the energy of interaction observed in the protein-DNA complexes. Molecular dynamics simulations have suggested that the favorable free energy from protein interaction with the thymine C5 methyl groups results from favorable changes in solvation free energy originating from the removal of hydrophobic surface from solvent, rather than interaction free energy (Plaxco et al., 1994, Saito et al., 2003). Hence, the contribution of the Van der Waals interaction observed in the mutant complex is expected to contribute little to the difference in $\Delta H°$ between the wild-type and mutant enzymes. However, the Van der Waals interaction might contribute to the more favorable $\Delta G°$ for mutant complex formation deriving from favorable $\Delta S°$ associated with the removal of the hydrophobic Thr138 γ-methylene group from solvent.

A second difference between the two models that is likely to have an impact on the thermodynamics of binding involves a water molecule trapped at the protein-DNA interface, which bridges interactions between the wild-type enzyme and the DNA. This water is absent from the mutant enzyme-DNA interface (Figure 3.8). The release of this water molecule from the 4-amino group of C_{-4} upon A138T complex formation will result in a more favorable entropy change relative to the wild-type binding reaction, consistent with my measurements (Table 3.6). An upper limit for the effect on $T\Delta\Delta S°$ ($T\Delta S°_{Mut}$ - $T\Delta S°_{WT}$) originating from releasing a single trapped water molecule is +2.1 kcal/mol (295K) based on the estimates from crystalline hydrate data (Dunitz, 1994). It is quite possible that immobilization of this water molecule in the wild-type complex exacts the full 2.1 kcal/mol entropic penalty, as the water is highly immobilized due to its participation in a hydrogen bond network involving functional groups from three amino acids (including two backbone atoms) and another solvent molecule that contacts the DNA (Figure 3.8A). On the other hand, release of this solvent molecule upon A138T complex

formation is likely to reduce the favorable enthalpy change accompanying the mutant binding reaction originating from loss of the water mediated contact and the enthalpic penalty for polar desolvation. Although the 4-amino group (C_{-4}) is packed against the Thr138, I do not expect this interaction to make up for the reduction in favorable enthalpy associated with the factors above (see above reference to methane-methane interaction energy). This is consistent with the unfavorable enthalpy change associated with mutant binding relative to wild-type (Table 3.6). Finally, immobilized waters, and hydrogen bond networks have been proposed to contribute to the negative heat capacity change associated with binding reactions in a number of systems (Ladbury et al., 1994, Cooper, 2000, Bergqvist et al., 2004). Large reductions in the magnitude of the heat capacity change have been observed upon disruption of hydrogen bonding networks involving trapped water molecules (Bergqvist et al., 2004). That there is no difference in $\Delta C°_p$ associated with the formation of the wild-type and A138T specific complexes (Table 3.6) is consistent with this particular local change in the structure, since the absence of the water molecule from the mutant interface leaves the remainder of the hydrogen bond network intact (Figure 3.16B). Having one more trapped water at the interface would be expected to make the $\Delta C°_p$ for wild-type binding more negative (Connelly, 1997) than for the mutant, but the difference (~10cal/mol K per water) is very small and certainly well within the error of the van't Hoff measurements.

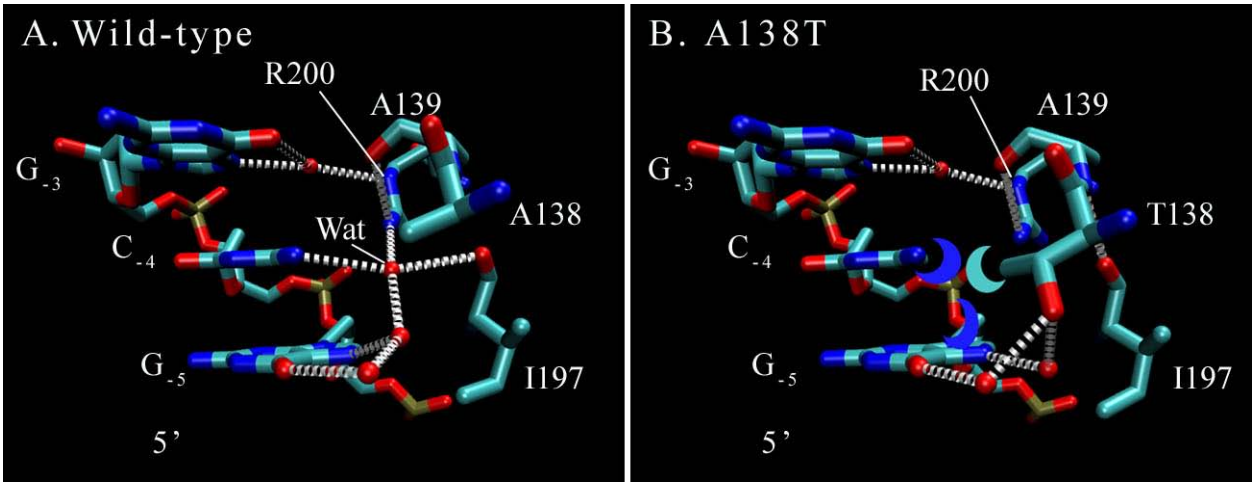


Figure 3.16 Water mediated hydrogen bond network couples contacts to both the GAATTC site and the flanking bases.

A) The labeled water molecule bridges contacts between the side-chain of R200 and the main-chain of I197 to the 4-amino group of C_{-4} in the wild-type complex.

B) In the mutant complex, the water is absent owing to the presence of the T138 side-chain. However, the network remains intact as the water mediated interactions are replaced by packing interactions (depicted as crescent shapes in this figure).

In her analysis of the structure of the *Eco*RI-d(TCGCGAATTCGCG) complex, Grigorescu proposed (Grigorescu, 2003) that the metastability of the arms is due to imperfections in the tertiary interactions within the arms. She found examples of poorly packed hydrophobic side-chains, charged residues buried in non-polar environments, and electrostatic repulsions between residues of similar charge (Grigorescu, 2003). (I do not wish to imply that improvement upon these 'imperfections' would lead to better enzyme, as the metastability of the arms is likely critical to the function of *Eco*RI).

I analyzed the arms in the A138T complex to ask if these 'imperfections' are still observed and whether there are any additional examples of sub-optimal design that would contribute to the reduction in magnitude of favorable enthalpy change associated with mutant complex formation relative to wild-type. There are no changes in the environments of the hydrophobic residues, Leu126, Met137, Val188, Val189 and Ile197, which are poorly packed in the wild-type complex. Further, Asp118 finds itself buried in a hydrophobic pocket in the inner

arms of both the wild-type and mutant complexes. There are examples of electrostatic repulsion between like-charged residues inferred from the coordinates of the mutant structure that are not predicted based on the wild-type model. However, these putative interactions involve residues (Arg123, Lys130, Arg131, and Arg244) whose side-chain B-factors are greater than 50 $Å^2$. The large B-factors indicate that the precise conformations of these residues are not specified by the data. Thus it is likely that the side-chains in both complexes are quite flexible, and that these interactions do not contribute to differences in the thermodynamic parameters. Next, I used WHAT IF (Hooft et al., 1996), the macromolecular model quality check program, to identify unusual side-chain rotamers, unsatisfied hydrogen bonds and steric clashes in the wild-type and mutant models that may differentially affect the thermodynamic parameters associated with formation of the two complexes. To summarize, the number of unsatisfied hydrogen bonds and steric clashes are greater in the wild-type complex than in the mutant complex. Also, the only difference in the observed unfavorable side-chain rotamers between the two complexes involves the active site residue, Glu111, which adopts an unusual rotamer in the wild-type complex and a more common rotamer in the A138T complex. Each of these differences would argue for a more favorable enthalpy change associated with mutant complex formation compared to wild-type complex formation, which is opposite to what the van't Hoff data suggest (Table 3.6). However, I do not lend much credence to these differences, largely because they lie in regions where the B-factors are too high to know the position of the atoms with certainty. Other structural observations might also lead to a more favorable enthalpy change for mutant complex formation. Among them are the hydrogen bonds to pG_{-3} and the scissile phosphate that are less optimal in the wild-type specific complex. One might also propose that the novel interaction with the scissile phosphate by the positively charged Lys113 side-chain alleviates some of the electrostatic repulsion that results from the juxtaposition of the scissile phosphate with the cluster of acidic residues comprising the active site. Reduction of electrostatic repulsion manifests itself as an increase in favorable enthalpy (Lisa E. Engler, *unpublished*). However, Ca^{2+}, which is expected to be coordinated by a scissile phosphoryl oxygen and active site carboxylate groups results in similar improvements in binding affinity for both enzymes (data shown in Chapter 5, Table 5.1). This would indicate that the unfavorable binding enthalpy component attributable to active site repulsion is the same for both complexes.

In this section I have pointed out two characteristics of the mutant structure that would lead to a more favorable ΔS° of binding for the A138T mutant: 1) Removal of the non polar portion of the Thr138 side-chain from solvent, and 2) The release of an ordered water molecule from the mutant interface that is present in the wild-type. A source of unfavorable ΔH° of mutant binding relative to wild-type binding is the loss of the water mediated hydrogen bond to C_{-4}, and the penalty associated with polar desolvation of the 4-amino group of the same base. These results are consistent with the differences in the thermodynamic parameters obtained from van't Hoff analysis. However, the loss in favorable ΔH° might be offset by the two hydrogen bonds that are optimized (shorter donor – acceptor distance) to phosphoryl oxygens, or the packing interactions between Thr138 and the GAATTC flanking bases observed in the mutant complex.

It is known that the observed favorable binding free energy for DNA-binding proteins is the net of large and opposing energetic forces (Figure 2.4). An understanding of these forces that determine the thermodynamics of binding requires the knowledge of the structure *and* dynamic properties of both the free *and* the bound macromolecular species. Our interpretations of the thermodynamic data are based solely on comparison of the enzyme DNA complexes. We have no knowledge of how the A138T mutation affects the free enzyme. In addition comparisons based on structural coordinates lead to an "enthalpy-centric" interpretation of the data as macromolecular structures solved by x-ray crystallography, while unparalleled in their ability to convey spatial features, are not very rich in information pertaining to the dynamics of the system. Thus, it remains possible that the A138T mutants exerts its effect on the differences in thermodynamic parameters by perturbing the structure and/or dynamics of the free enzyme, or the dynamics of the complex, aspects of which are largely "invisible" to the technique of x-ray crystallography. Changes in accessible polar/non-polar surface area, changes in intra-molecular contacts, and changes in configurational entropy are all expected to accompany the folding of the arms which is coupled to binding, and any number of these features may be affected by the A138T mutation.

4. PROMISCUOUS MUTANTS AND THE NON-COGNATE DNA INTERACTION: INSIGHTS INTO THE BASIS FOR RELAXED *IN VIVO* SPECIFICITY

4.1. Introduction

The phenomena of restriction and modification were first observed in the early 1950's when Luria and Human (Luria et al., 1952), and Bertani and Weigle (Bertani et al., 1953) found that λ phage grown on *E. coli* strain C, grew poorly when propagated on a different *E. coli* strain, K12. It was later shown that site specific DNA endonucleases and methylases formed the molecular basis for the above genetic observations (Roberts, 1976, Arber, 1979). Indeed, bacterial expression of many restriction endonucleases has been shown to severely restrict phage growth in bacteria, leading to the hypothesis that the selective advantage of carrying a restriction-modification system lies in their function as defense tools (Arber, 1965, Rambach et al., 1974). However, several pieces of evidence suggest that a function as bacterial 'immune systems' against the invasion of foreign DNA was not the only selective pressure for the evolution and maintenance of restriction modification systems (RMSs). First, many cloned restriction modification systems do not restrict phage growth when expressed in *E. coli* (Heitman, 1993). This could be due to the fact that many bacteriophage have evolved multiple strategies to avoid destruction of their genomes at the hands of restriction endonucleases such as DNA modification (Kruger et al., 1983, Birge, 1994, Campbell, 1996), and the avoidance of palindromic DNA sites within their genomes (Rocha et al., 2001). Second, the restriction endonucleases discovered to date are only capable of cleaving double stranded DNA, thus single stranded DNA and RNA phage can infect bacteria despite the presence of RMSs (Levin, 1993). The last argument against the universal applicability of the defense hypothesis is the existence of rare cutting restriction endonucleases which recognize 8 base-pair sequences that are typically

absent from phage genomes due to their small size. This observation suggests that at least in some cases, selection for the maintenance of RMSs occurs at the level of the host genome, not the phage genome.

In accord with this idea, Kobayashi and colleagues (Naito et al., 1995) proposed that RMSs be considered selfish genetic elements that invade genomes, and are maintained without necessarily providing selective advantages to the "host" cell. It has been shown that once many RMSs are acquired through transformation via plasmids or horizontal gene transfer (Jeltsch et al., 1996), RMSs become essential for the survival of the bacteria (Kulakauskas et al., 1995, Kusano et al., 1995, Handa et al., 2000). The hypothesis explaining this observation posits that if a bacterium loses the RMS genes, dilution of the RMS gene products through subsequent rounds of cell division will result in an inability of the methylase to protect all recognition sites, such that the remaining endonuclease molecules can cleave unprotected recognition sites within the cells' genome, resulting in death. This process has been termed "post-segregational killing". Interestingly, Rocha et al. showed (Rocha et al., 2001) that RMSs are a burden to both bacteria and phage as evidenced by the avoidance of recognition sites for restriction endonucleases within the genomes of phage and bacteria. It should also be noted that post-segregational killing has been observed after the loss of genes encoding type II RMSs, however, alleles encoding type I RMSs are able to be lost or replaced by other RMSs with different sequence specificities (O'Neill et al., 1997). Therefore not all RMSs are selfish genetic elements, suggesting that some restriction endonucleases provide selective advantages through utility in defense or some other function.

The *Eco*RI RMS has been shown to be effective in defense against phage infection (Yoshimori et al., 1972), and in post-segregational killing (Handa et al., 1999). Regardless of the true "function" of the *Eco*RI RMS, the system must have evolved to avoid cleavage of cellular DNA in order for bacteria harboring this RMS to remain viable and competitive with other *E. coli* cells vying for the same niche. The *E. coli* K12 chromosome (GenBank accession number U00096) contains 645 GAATTC sites, 20,833 miscognate sites and about 9.2 million non-specific sites (every base-pair of the genome represents the beginning of a non-specific binding site). It is imperative that the endonuclease of a RMS avoids cleavage of cellular DNA at the specific recognition site or related sites, as restriction of the host genome can be a lethal event. Lesser et al. in studying the *Eco*RI endonuclease, proposed that type II restriction endonucleases

evolved multiple fail-safes in order to avoid genomic DNA cleavage *in vivo* (Lesser et al., 1990). It has long been appreciated that restriction endonuclease expression is lethal in the absence of expression of the companion methylase (Oelgeschlager et al., 1990). This is due to the fact that the endonuclease will cleave GAATTC (applicable to the specific case of the *Eco*RI restriction endonuclease) within the host genome if these sites are not modified by the sister methylase. Indeed, the probability of cleavage of the methylated cognate site is reduced about 10^9-fold relative to the unmodified site (Jen-Jacobson et al., 1996). Second, binding affinities for nonspecific and miscognate DNA sites are comparable, but nonspecific DNA sites are about 450-fold more numerous than *Eco*RI* sites in the *E. coli* genome. Consequently, nonspecific DNA competes effectively with *Eco*RI* sites for endonuclease binding *in vivo*. Third, the stringent discrimination against binding to *Eco*RI* sites and the lower cleavage rate constants combine to vastly reduce the probability of initial cleavage at such a site. Fourth, in the unlikely event that a nick occurs in one strand of an *Eco*RI* site (most likely in the normal half-site), cleavage of the second DNA strand is even more unlikely, since the dissociation rate constant for *Eco*RI* complexes is $>10^5$-fold faster than the rate of strand cleavage in the incorrect half-site (Lesser et al., 1990). Thus, potentially lethal double-strand cuts are avoided by this "kinetic proofreading" mechanism (Figure 4.1), and the single strand nicks can be sealed by DNA ligase (Heitman et al., 1989).

In this chapter, I present data which characterize three promiscuous mutant variants of the *Eco*RI restriction endonuclease, A138T, E192K and H114Y, with respect to their discrimination against miscognate and non-specific DNAs at both the levels of binding and cleavage. I explain how the mutant enzymes manage to partially bypass the multiple failsafe mechanisms that were evolved to ensure protection against cleavage of genomic DNA in cells carrying the wild-type *Eco*RI restriction modification system. The ideas presented in this chapter provide a basis for resolving the apparently paradoxical observation that mutant restriction endonucleases which bind more tightly than the wild-type enzyme to the specific recognition sequence *in vitro* (as demonstrated in Chapter 2), nevertheless exhibit relaxed specificity *in vivo*. In addition, an analysis of the effects of flanking context on miscognate DNA binding and cleavage by the *Eco*RI enzyme reveals that modulation of miscognate DNA cleavage rates by the sequences flanking the miscognate site provides an additional level of protection against restriction of cellular DNA at miscognate sites. I show that the A138T mutant enzyme is expected to

circumvent this hitherto unappreciated failsafe. Finally, I discuss potential barriers and pathways for the design and/or evolution of restriction endonucleases with altered specificities in the context of the body of structural, biochemical and genetic data on the *Eco*RI restriction modification system.

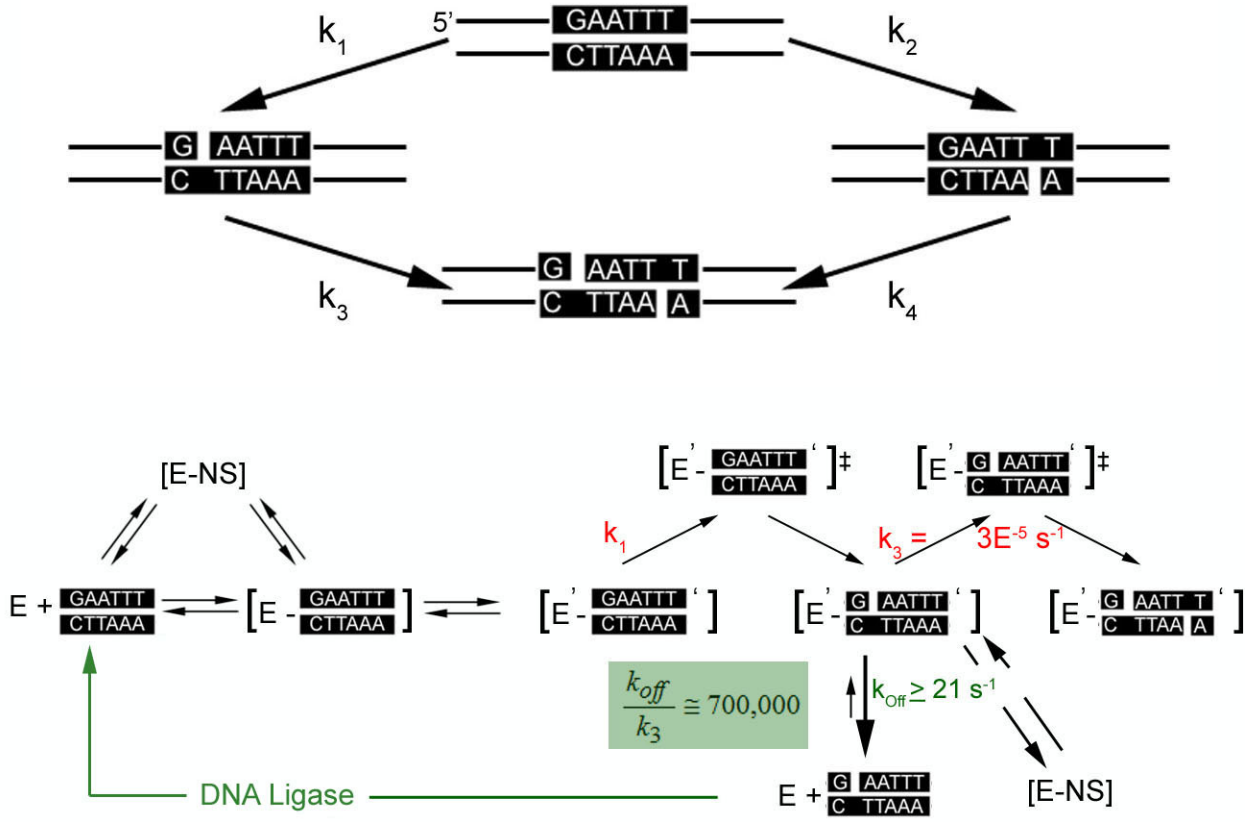


Figure 4.1 Kinetic proofreading prevents double stranded DNA cleavage by the *Eco*RI endonuclease *in vivo*.

Top panel) Illustration of the parallel sequential pathway for cleavage of DNA substrates. The miscognate site, AAATTC is used as the example in this figure. k_1 and k_2 refer to the rate constants for cleavage of the cognate and non-cognate half-sites of the AAATTC miscognate site respectively. k_3 and k_4 refer to the rate constants for cleavage of the nicked substrates.

Bottom panel) Illustration of the kinetic proofreading mechanism that is believed to protect against double stranded cleavage of miscognate sites *in vivo*. In this figure, E represents the *Eco*RI endonuclease, NS is non-specific DNA, entities in brackets indicate enzyme-DNA complexes, and primed superscripts denote complexes in which conformational changes of the DNA and enzyme have likely occurred. Data support a predominating pathway in which miscognate sites are nicked at the cognate half-site, after which the enzyme dissociates from the DNA before the second nick occurs; these nicked substrates are then healed by DNA ligase.

4.2. Results and Discussion

4.2.1. Mutant proteins maintain binding discrimination against miscognate sites

Binding specificity is defined as a simple ratio of equilibrium association constants for specific and non-cognate DNA sites (non-cognate = miscognate *or* non-specific).

$$Specificity\ Ratio = \frac{K_{Specific}}{K_{Non-cognate}} \qquad 4.1$$

One obvious potential explanation for the promiscuity of the mutant proteins is that they no longer bind selectively to the cognate GAATTC site, but instead bind preferentially to miscognate DNA sites. Because binding to GAATTC is improved by the mutations, a reduction in the specificity ratio would require that binding to miscognate sites is even more improved.

To test this possibility, we measured equilibrium binding of wild type and mutant proteins to a subset of the nine possible *Eco*RI* sites. Among the miscognate sites tested is a representative site with a base-pair change at each position within the recognition sequence. The representative *Eco*RI* site chosen at each position was the site that showed the most favorable binding and fastest cleavage rate in studies with the wild-type *Eco*RI endonuclease (Lesser et al., 1990). Two sites were chosen with substitutions at the first position; TAATTC by virtue of its being cleaved the fastest of all the *Eco*RI* sites, and AAATTC because the rate constants are symmetrical for cleavage of both DNA strands of this site, making AAATTC one of the most likely miscognate sites to undergo double stranded cleavage in a single binding event.

Wild-type binding to *Eco*RI* sites entails penalties of 6.2-6.7 kcal/mol relative to specific binding (Table 4.1). (These penalties are larger than those reported in previous studies (Lesser et al., 1990), because we now use the most favored CGC flanking context which improves specific but not *Eco*RI* binding relative to the GCA context used previously.) Mutant proteins bind more favorably to miscognate DNA sites by 1.4-fold to 48-fold ($\Delta\Delta G°_{bind}$ = –0.2 to –2.3 kcal/mol, Table 4.1, Figure 4.2), but for each mutant protein, this is less than the improvement in binding to the cognate GAATTC site. Thus, using the customary metric of the ratio of

equilibrium association constants (the "specificity ratio", Equation 4.1), the mutant enzymes have a higher degree of binding preference for cognate over miscognate sites than does the wild-type enzyme. This was completely unexpected, and compelled us to look elsewhere for the source of promiscuity.

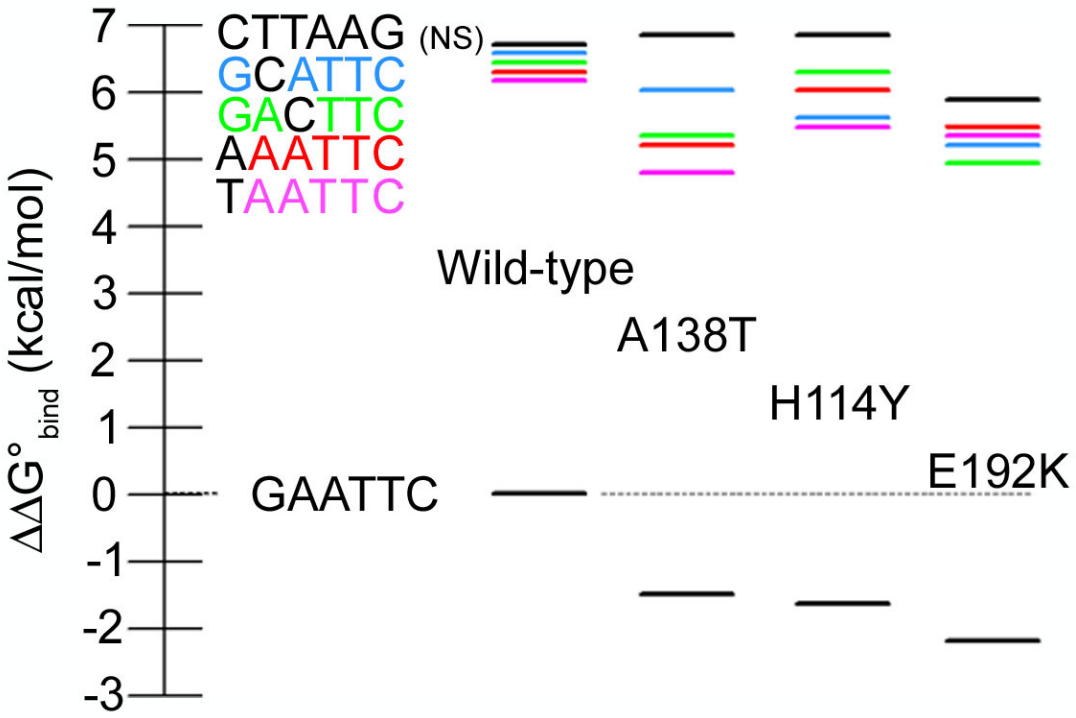


Figure 4.2 Binding discrimination by the *Eco*RI endonuclease against miscognate and non-specific DNAs.

Difference in the free energy change for binding to 24 base pair oligomers containing *Eco*RI miscognate (5/6 bp match with GAATTC) or non-specific (≤4/6 bp match with GAATTC) sites. Data in all columns are referenced ($\Delta\Delta G°_{bind} = 0$) to the wild-type interaction with a 24mer containing GAATTC embedded in a symmetrical 5' CGC flanking sequence. The non-specific 24mer contains an 'inverted' site, CTTAAG in the CGC-flanking context. Equilibrium association constants were measured in binding buffer plus 0.22M KCl, pH 7.3, 21°.

Table 4.1 Effect of *Eco*RI endonuclease mutations on non-cognate DNA binding

Site [a]	K_A [b] (M^{-1})	Specificity Ratio [c]	$\Delta\Delta G°_{bind}$ (Site) [d] (kcal/mol)	$K_{A,\ mutant}/K_{A,\ wild\text{-}type}$	$\Delta\Delta G°_{Bind}$ (Mutant) [e] (kcal/mol)
Wild-type					
GAATTC	$9.6\ (\pm 0.3)\times10^{9}$	1	0	-	0
AAATTC	$1.9\ (\pm 0.1)\times10^{5}$	51,000	6.3 ± 0.1	-	0
TAATTC	$2.3\ (\pm 0.1)\times10^{5}$	42,000	6.2 ± 0.1	-	0
GCATTC	$1.0\ (\pm 0.8)\times10^{5}$	96,000	6.7 ± 0.5	-	0
GACTTC	$1.2\ (\pm 0.1)\times10^{5}$	80,000	6.6 ± 0.1	-	0
CTTAAG	$1.0\ (\pm 0.1)\times10^{5}$	96,000	6.7 ± 0.1	-	0
A138T					
GAATTC	$1.3\ (\pm 0.1)\times10^{11}$	1	0	14	-1.5 ± 0.1
AAATTC	$1.4\ (\pm 0.1)\times10^{6}$	93,000	6.7 ± 0.1	7	-1.2 ± 0.1
TAATTC	$2.1\ (\pm 0.4)\times10^{6}$	62,000	6.5 ± 0.1	9	-1.3 ± 0.1
GCATTC	$3.3\ (\pm 0.6)\times10^{5}$	390,000	7.5 ± 0.5	3	-0.7 ± 0.1
GACTTC	$8.7\ (\pm 0.6)\times10^{5}$	150,000	7.0 ± 0.1	7	-1.2 ± 0.1
CTTAAG	$1.0\ (\pm 0.1)\times10^{5}$	1,300,000	8.3 ± 0.1	1	0.1 ± 0.1
H114Y					
GAATTC	$1.6\ (\pm 0.1)\times10^{11}$	1	0	17	-1.7 ± 0.1
AAATTC	$3.0\ (\pm 1.4)\times10^{5}$	530,000	7.7 ± 0.3	2	-0.3 ± 0.3
TAATTC	$8.1\ (\pm 2.0)\times10^{5}$	200,000	7.1 ± 0.1	4	-0.7 ± 0.1
GCATTC	$5.9\ (\pm 0.7)\times10^{5}$	270,000	7.3 ± 0.1	6	-1.0 ± 0.5
GACTTC	$1.7\ (\pm 0.1)\times10^{5}$	940,000	8.1 ± 0.1	1	-0.2 ± 0.1
CTTAAG	$8.9\ (\pm 0.1)\times10^{4}$	1,800,000	8.4 ± 0.1	0.9	0.1 ± 0.1
E192K					
GAATTC	$4.5\ (\pm 0.1)\times10^{11}$	1	0	47	-2.3 ± 0.1
AAATTC	$8.8\ (\pm 0.3)\times10^{5}$	510,00	7.7 ± 0.1	5	-0.9 ± 0.1
TAATTC	$1.1\ (\pm 0.3)\times10^{6}$	410,000	7.5 ± 0.1	5	-0.9 ± 0.1
GCATTC	$1.4\ (\pm 0.2)\times10^{6}$	320,000	7.4 ± 0.1	14	-1.5 ± 0.1
GACTTC	$1.8\ (\pm 0.4)\times10^{6}$	250,000	7.3 ± 0.1	15	-1.6 ± 0.2
CTTAAG	$5.0\ (\pm 0.3)\times10^{5}$	900,000	8.0 ± 0.1	5	-0.9 ± 0.1

[a] Each site is embedded in a double stranded 24 base-pair oligodeoxynucleotide, 5'-GGGCGGGCGC<u>XXXXXX</u>GCGGGCGC, where <u>XXXXXX</u> represents the site listed in column 1.

[b] Equilibrium association constants were measured in binding buffer plus 0.22 M KCl (pH 7.3, 21°C); means ± std. dev. of ≥ 3 determinations.

[c] For each enzyme, the specificity ratio is calculated as $K_{GAATTC}/K_{variant}$

[d] For each enzyme, the observed difference in standard binding free energy between each variant and specific site is calculated as

$$\Delta\Delta G°_{bind\ (Site)} = -RT \ln (K_A^{variat}/K_A^{specific}) \text{ at } 294 \text{ K}.$$

[e] For each sequence, the observed difference in standard binding free energy between the mutant and wild-type enzymes is calculated as:

$$\Delta\Delta G°_{bind\ (Mutant)} = -RT \ln (K_A^{mutant}/K_A^{wild\text{-}type}) \text{ at } 294 \text{ K}.$$

4.2.2. Mutant enzymes exhibit increased discrimination between non-specific and miscognate sites

Our results show that the relaxed specificity mutants have increased specificity ratios when discrimination between specific and miscognate DNAs are considered (Table 4.1). If one bears in mind the context in which these mutants were isolated, the *E. coli* cell, we can gain some understanding as to how the mutants exhibit more stringent binding discrimination *in vitro*, yet have relaxed specificity *in vivo*. In the *E. coli* cell, cleavage of the 645 GAATTC sites is prevented by the activity of the *Eco*RI methylase; thus the specific recognition site, and hence the specificity ratio discussed above is irrelevant to *in vivo* specificity. An important aspect of discrimination by the wild-type enzyme is that it has similar affinities for miscognate and nonspecific DNA sites (Table 4.1); thus a change in the ratio of *miscognate* to *non-specific* DNA binding affinities is expected to have a profound effect on *in vivo* specificity.

The wild-type enzyme prefers binding to the most favored *Eco*RI* site (TAATTC) by only a factor of two ($\Delta\Delta G°_{bind}$ = –0.5 kcal/mol) over nonspecific DNA. By contrast, the A138T and H114Y enzymes have improved binding to TAATTC but an unaltered binding affinity for nonspecific DNA, and thus prefer TAATTC over nonspecific DNA by 21-fold (A138T) and 9-fold (H114Y). Smaller changes of the same kind have occurred for the other *Eco*RI* sites tested. In Figure 4.3, the binding data (Table 4.1) for miscognate and nonspecific DNA sites are plotted on a free-energy scale relative to the wild-type endonuclease interaction with nonspecific DNA. Because the A138T and H114Y mutations improve miscognate binding but do not affect nonspecific binding, it follows (Figure 4.3) that these two mutant enzymes discriminate more stringently than the wild-type enzyme between miscognate and nonspecific DNA sites. In consequence, nonspecific DNA (which cannot be cleaved) competes less effectively for the mutant endonucleases than it does for wild-type endonuclease.

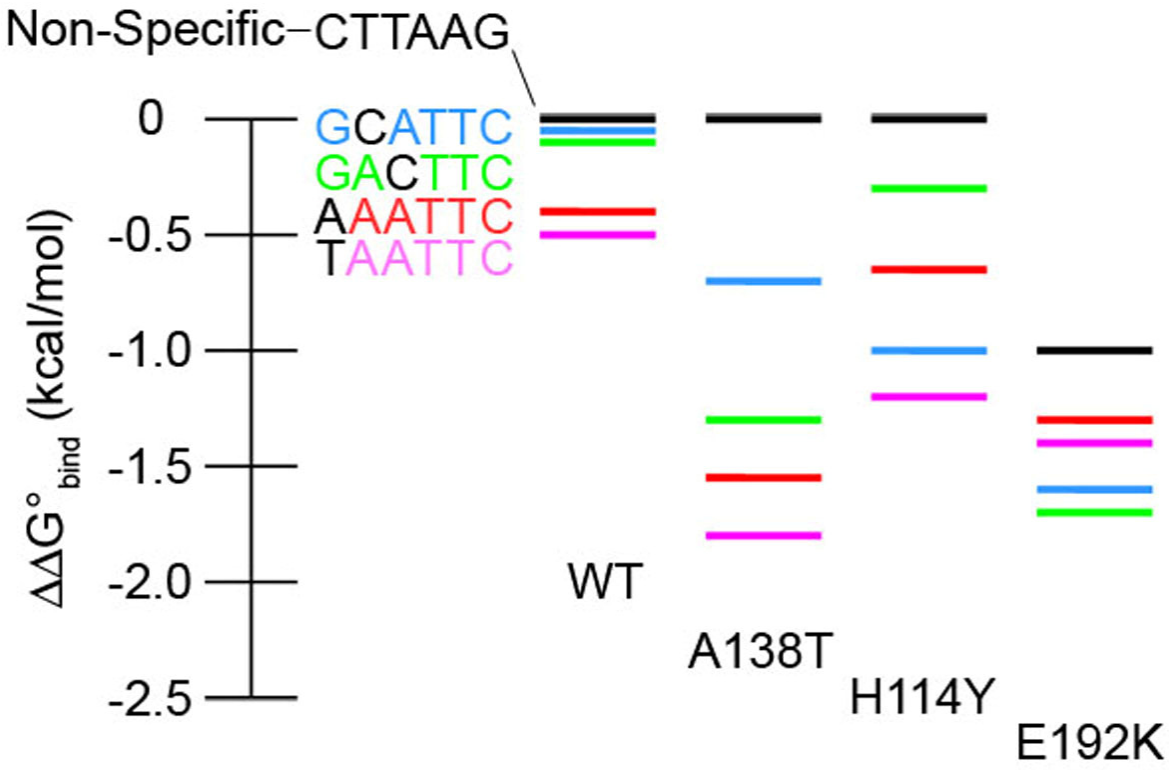


Figure 4.3 Differences in *Eco*RI endonuclease binding affinities between miscognate and non-specific DNAs.

Data are referenced ($\Delta\Delta G°_{Bind}=0$) to the wild-type interaction with non-specific DNA. Negative values of $\Delta\Delta G°_{bind}$ signify more favorable binding relative to the wild-type non-specific DNA interaction. Raw data are reported in Table 4.1.

To show how these changes in relative affinities would affect partitioning of endonuclease molecules between the two classes of sites in the cell, we calculated a partitioning ratio (Equation 4.2), which takes into account both the binding constants and the abundance of *Eco*RI* and nonspecific sites in the *E. coli* genome. This equilibrium view is useful in discussing the in v*ivo* specificity of the *Eco*RI endonuclease, as the association and dissociation rates for miscognate complexes are much faster than miscognate DNA cleavage rates (Lesser et al., 1990).

$$Partitioning\ Ratio = \frac{K_{A,Miscognate} \times No.\ Miscognate\ Sites}{K_{A,Non-specific} \times No.\ Non-specific\ Sites} \qquad 4.2$$

In these calculations, we took each of the 4.6 x 10^6 bp in the *E. coli* genome to mark the beginning of a potential nonspecific binding site. The number of each of the miscognate sites in the genome was obtained from the nucleotide sequence of the *E. coli* K12 genome (GenBank accession no. U00096). Considering the TAATTC site as an example, the wild-type enzyme has a partitioning ratio of 1:140, meaning that only one wild-type endonuclease molecule is expected to bind to the *Eco*RI* site for every 140 molecules bound to nonspecific DNA on the *E. coli* chromosome (Table 4.2). By contrast, the A138T and H114Y mutant enzymes have increased affinity for *Eco*RI* but not nonspecific DNA, resulting in partitioning ratios of 1:9 (A138T) and 1:22 (H114Y) for the TAATTC site. Similar trends in changes of the A138T and H114Y partitioning ratios relative to the wild-type partitioning ratios are observed for the other *Eco*RI* sites tested. These data indicate that nonspecific DNA would be less effective at sequestering the A138T and H114Y endonucleases away from miscognate sites *in vivo*, thereby reducing the protective capacity of nonspecific DNA in cells expressing one of these mutant endonucleases.

The E192K mutant is the only enzyme that binds more strongly to all three classes of DNA sites (specific, miscognate and nonspecific) with nonspecific affinity being 5-fold better than that of the wild-type enzyme ($\Delta\Delta G°_{bind}$ = –0.9 kcal/mol). Tighter binding of the E192K protein might reflect a more favorable electrostatic component to the binding free energy, resulting from both the removal of a negative charge and the addition of a positive charge near the phosphate backbone of DNA. This more favorable electrostatic interaction would lead to tighter binding to all classes of DNA, but perhaps to differing degrees. We have shown that the E192K enzyme has higher affinity than wild-type enzyme for all three types of DNA sites, but with gradations in the degree of enhancement among the three kinds of sites. Relative to wild-type enzyme, E192K enzyme binds 20- to 50-fold better to specific DNA (depending on flanking context-see Table 2.1), 5- to 15-fold better to various miscognate DNA sites and 5-fold better to nonspecific DNA (Table 4.1). Given that the strength of electrostatic interactions depends on the distances between charges, and residue 192 lies in the outer segment of the *Eco*RI arm, the range in E192K binding enhancements observed for the different classes of DNA sites might reflect differences in the conformation and/or dynamics of the arms in the three types of complexes. Also, the fact that the Y193H mutant is promiscuous (Heitman and Model, 1990) suggests that increased electrostatic interaction between this region of the arm and the DNA may be sufficient to cause promiscuous behavior.

Taking the partitioning ratios into account, the E192K enzyme is promiscuous for only two of the four EcoRI* sites tested, GACTTC and GCATTC, (Table 4.2) but would partition similarly to wild-type enzyme between nonspecific DNA and the AAATTC or TAATTC miscognate sites.

It has long been appreciated that binding to non-specific DNA accelerates the search of DNA binding proteins for their cognate binding sites by reducing a random three dimensional search for a target DNA site to a linear search by sliding along the DNA helix in a non-specific binding mode (Berg et al., 1981, Winter et al., 1981). Indeed, it has been shown that the *Eco*RI endonuclease undergoes linear diffusion along DNA to accelerate the search for target sites within the double helix (Jack et al., 1982, Ehbrecht et al., 1985, Terry et al., 1985). However, the partitioning ratios presented here underscore another critical role of non-specific DNA in the function of the *Eco*RI restriction endonuclease; a protective role whereby non-specific DNA competes with miscognate DNA for endonuclease binding *in vivo*.

Table 4.2 Partitioning of *Eco*RI endonucleases among miscognate and non-specific DNA sites *in vivo*

Enzyme	AAATTC:NS [a]	TAATTC:NS	GCATTC:NS	GACTTC:NS
Wild-type	1:120 [b]	1:140	1:3300	1:980
A138T	1:8	1:9	1:58	1:17
H114Y	1:44	1:22	1:24	1:230
E192K	1:140	1:140	1:72	1:80

[a] Partitioning ratios were calculated using equation 4.2 (see text), using intrinsic equilibrium association constants to correct for the contribution of non-specific binding to the apparent miscognate DNA binding constant. The values of K_{Int} were calculated by:

$$K_{Int.} = K_{App.} - N_{NS} * K_{NS}$$

A partitioning ratio of 1:120 means that one endonuclease molecule is bound to the miscognate site for every 120 bound to non-specific DNA. This calculation assumes the most favored symmetrical CGC flanking triplets, but actual miscognate sites *in vivo* will have other triplets, that are usually asymmetric. For the wild-type protein, flanking context affects neither *Eco*RI* nor non-specific binding. However, for the A138T protein, contexts affects *Eco*RI*, but not non-specific binding, so these are maximum estimates of the fraction of A138T bound to *Eco*RI* sites. The effect of context on *Eco*RI* binding by H114Y and E192K is unknown.

4.2.3. Mutant proteins stabilize the transition state in miscognate complexes

To determine if the relaxed specificity phenotype of the promiscuous mutants is due in part to relaxed specificity at the cleavage step, we measured the first-order rate constants for cleavage of four different *Eco*RI* sites. The results (Table 4.3) show that each of the mutant enzymes still prefers the specific recognition site in cleavage as well as in binding (Table 4.1). For example, the A138T enzyme cleaves the GAA half-sites of *Eco*RI* sites 5- to 380-fold slower ($\Delta\Delta G^{\circ\ddagger}$ = +0.9 to +3.5 kcal/mol) than GAATTC, and the miscognate half-site 11- to 41,000-fold slower ($\Delta\Delta G^{\circ\ddagger}$ = +1.4 to +4.9 kcal/mol) than GAATTC (Table 4.3). The H114Y and E192K proteins also have roughly similar preferences for GAATTC over *Eco*RI* sites. Interestingly, the mutant proteins retain the same hierarchy of cleavage rates among *Eco*RI* sites as the wild-type enzyme, as well as the same qualitative relationships between k_1 and k_2 for cleavage of each *Eco*RI* site. In addition, many idiosyncratic features of the endonuclease-*Eco*RI* cleavage reactions are preserved by the mutant proteins. For example, the GACTTC site is the only one for which the wild-type enzyme cleaves more rapidly in the incorrect than in the correct half-site, and this is also true for all three mutant proteins. Despite these conserved features, all three mutant enzymes are observed to cleave miscognate *Eco*RI* sites from 2- to 120-fold faster ($\Delta\Delta G^{\circ\ddagger}$ = –0.3 to –2.8 kcal/mol) than does the wild-type enzyme, depending on the mutant and the *Eco*RI* site (Table 4.3). Also, there is an hierarchy of *Eco*RI* site cleavage rate constants among the various enzymes, such that the rate constants for any given site generally fall in the order A138T > H114Y > E192K > wild-type (Figure 4.4). It follows that the mutant proteins, although they have improvements over the wild-type enzyme in binding to both GAATTC and *Eco*RI* sites in the ground state, are even more improved in stabilizing the transition states of miscognate complexes. This preferential interaction with the transition states of *Eco*RI* complexes is likely a significant factor in the promiscuity of these mutant enzymes.

Table 4.3 Effect of *Eco*RI endonuclease mutations on miscognate DNA cleavage

Site [a]	$k_{1,cleave}$ [b] (s-1)	$\Delta\Delta G^{\circ}_{1}{}^{\ddagger}$ (Site) [c] (kcal/mol)	$\Delta\Delta G^{\circ}_{1}{}^{\ddagger}$ (Mutant) [d] (kcal/mol)	$k_{2,cleave}$ [e] (s^{-1})	$\Delta\Delta G^{\circ}_{1}{}^{\ddagger}$ (Site) [c] (kcal/mol)	$\Delta\Delta G^{\circ}_{1}{}^{\ddagger}$ (Mutant) [d] (kcal/mol)
Wild-type						
GAATTC	0.27 ±0.01	0	0	0.29 ±0.04	0	0
AAATTC	1.7 (± 0.6)x10^{-3}	3.0 ± 0.2	0	1.1 (± 0.1)x10^{-3}	3.2 ± 0.1	0
TAATTC	6.6 (± 3.0)x10^{-3}	2.2 ± 0.3	0	5.1 (± 1.1)x10^{-5}	5.1 ± 0.2	0
GCATTC	5.6 (± 0.5)x10^{-5}	5.0 ± 0.1	0	4.8 (± 0.6)x10^{-6}	6.4 ± 0.1	0
GACTTC	1.7 (± 0.2)x10^{-3}	3.0 ± 0.1	0	2.3 (± 0.2)x10^{-3}	2.8 ± 0.1	0
CTTAAG	No Cleavage	∞	∞	No Cleavage	∞	∞
A138T						
GAATTC	0.42 ±0.02	0	-0.3 ± 0.1	0.41 ±0.02	0	-0.2 ± 0.1
AAATTC	2.3 (± 0.4)x10^{-2}	1.7 ± 0.1	-1.5 ± 0.2	1.7 (± 0.2)x10^{-2}	1.9 ± 0.1	-1.6 ± 0.1
TAATTC	9.4 (± 3.3)x10^{-2}	0.9 ± 0.2	-1.6 ± 0.3	3.6 (± 0.1)x10^{-4}	4.1 ± 0.1	-1.1 ± 0.1
GCATTC	1.1 (± 0.1)x10^{-3}	3.5 ± 0.1	-1.8 ± 0.1	1.0 (± 0.2)x10^{-5}	4.9 ± 0.1	-1.8 ± 0.1
GACTTC	3.3 (± 0.2)x10^{-2}	1.5 ± 0.1	-1.7 ± 0.1	3.7 (± 0.3)x10^{-2}	1.4 ± 0.1	-1.6 ± 0.1
CTTAAG	No Cleavage	∞	∞	No Cleavage	∞	∞
H114Y						
GAATTC	0.30 ±0.02	0	-0.1 ± 0.1	0.31 ±0.02	0	0.0 ± 0.1
AAATTC	1.8 (± 0.1)x10^{-2}	1.6 ± 0.1	-1.4 ± 0.2	1.4 (± 0.1)x10^{-2}	1.8 ± 0.1	-1.5 ± 0.1
TAATTC	1.1 (± 0.1)x10^{-2}	1.9 ± 0.1	-0.3 ± 0.3	6.7 (± 0.3)x10^{-4}	3.6 ± 0.1	-1.5 ± 0.1
GCATTC	2.1 (± 0.3)x10^{-3}	2.9 ± 0.1	-2.0 ± 0.1	5.8 (± 0.2)x10^{-4}	3.7 ± 0.1	-2.8 ± 0.1
GACTTC	1.7 (± 0.1)x10^{-2}	1.7 ± 0.1	-1.3 ± 0.1	3.4 (± 0.3)x10^{-2}	1.3 ± 0.1	-1.6 ± 0.1
CTTAAG	No Cleavage	∞	∞	No Cleavage	∞	∞
E192K						
GAATTC	0.22 ±0.01	0	0.1 ± 0.1	0.26 ±0.03	0	0.1 ± 0.1
AAATTC	1.2 (± 0.6)x10^{-2}	1.7 ± 0.3	-1.1 ± 0.4	8.1 (± 0.6)x10^{-3}	2.0 ± 0.1	-1.2 ± 0.1
TAATTC	1.7 (± 0.2)x10^{-2}	1.5 ± 0.1	-0.6 ± 0.3	1.2 (± 0.2)x10^{-4}	4.5 ± 0.2	-0.5 ± 0.2
GCATTC	6.3 (± 0.3)x10^{-4}	3.4 ± 0.1	-1.4 ± 0.1	1.5 (± 0.6)x10^{-4}	4.4 ± 0.1	-2.0 ± 0.1
GACTTC	9.3 (± 3.3)x10^{-3}	1.9 ± 0.2	-1.0 ± 0.2	1.4 (± 0.1)x10^{-2}	1.7 ± 0.1	-1.1 ± 0.1
CTTAAG	No Cleavage	∞	∞	No Cleavage	∞	∞

[a] Each site is embedded in a double stranded 24 base-pair oligonucleotide: 5'-GGGCGGGCGC<u>XXXXXX</u>GCGGGCGC, where <u>XXXXXX</u> represents the site listed in column 1.

[b] Means ± SD of ≥ 3 determinations; first-order strand scission constants were obtained under single-turnover conditions using 8 mM Mg2+ to initiate the reaction; k_1 represents cleavage in the cognate half-site, GAA and k_2 represents cleavage of the non-cognate half site (see figure 4.1).

[c] For each *enzyme*, the difference in activation free energy for cleavage between each miscognate site and the specific site (as reference) is calculated as:

$$\Delta\Delta G^{\circ}_{1}{}^{\ddagger}\text{ (Site)} = -RT \ln (k_1{}^{miscognate}/k_1{}^{cognate}) \text{ or } \Delta\Delta G^{\circ}_{2}{}^{\ddagger} = -RT \ln (k_2{}^{miscognate}/k_2{}^{cognate}) \text{ at 294 K.}$$

[d] For each *site*, the difference in activation free energy for cleavage between the mutant and wild-type is calculated as:

$$\Delta\Delta G^{\circ}_{1}{}^{\ddagger}\text{ (Mutant)} = -RT \ln (k_1{}^{mutant}/k_1{}^{wild\text{-}type}) \text{ or } \Delta\Delta G^{\circ}_{2}{}^{\ddagger} = -RT \ln (k_2{}^{mutant}/k_2{}^{wild\text{-}type}) \text{ at 294 K.}$$

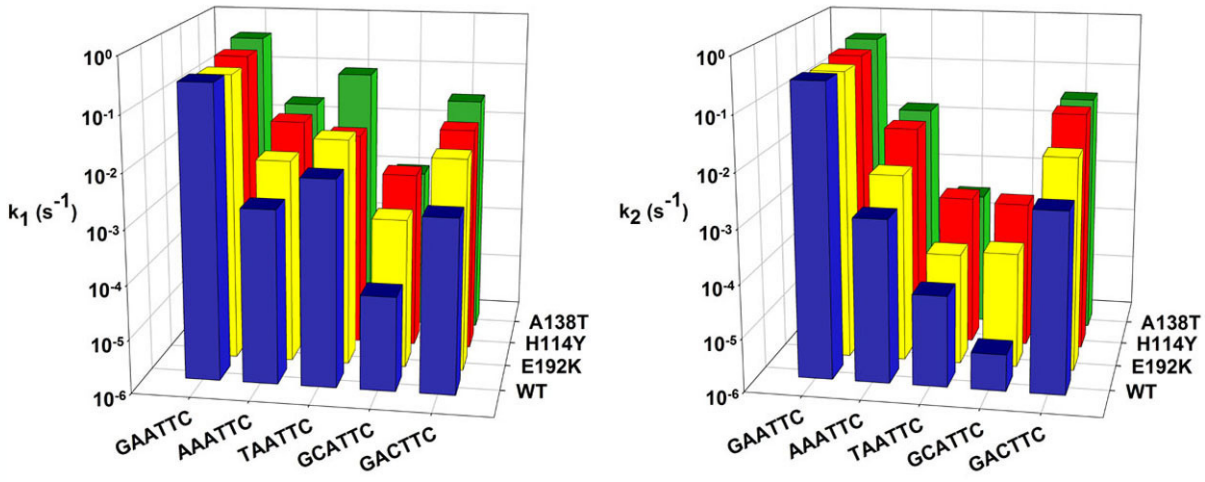


Figure 4.4 Hierarchies of cleavage rate constants among *Eco*RI endonucleases.

Rate constants for cleavage of the cognate half-site of miscognate DNAs, k_1, are shown in panel A, and rate constants for cleavage of the non-cognate half-site, k_2, are shown in panel B. Bars in panels A and B represent means of at least three trials per sequence. Note that the vertical axis is on a logarithmic scale. Raw data are from Table 4.3. Key: wild-type (blue), E192K (yellow), H114Y (red), A138T (green).

4.2.4. Flanking sequence modulates miscognate DNA cleavage

To determine if flanking context influences the *Eco*RI endonuclease-miscognate DNA interaction, we measured the equilibrium association constants and first order cleavage rate constants for the AAATTC miscognate site flanked by different base-pair triplets. We chose a subset of 9 out of 64 of the possible flanking triplets with the aim of separating effects on binding and cleavage as a result of changing the assortment of functional groups in the major groove of flanking DNA from effects originating from changes in sequence dependent DNA flexibility (results will be discussed in this context in chapter 5). In chapter 2 we reported data for the wild-type and A138T *Eco*RI endonucleases that show specific DNA binding affinity varies over a range of 4 kcal/mol, depending on the flanking context, yet first order rates of DNA cleavage are unaffected by altering the GAATTC flanking sequence (section 2.2.2). Data in table 4.4 support the general conclusion that the effects of changing the flanking sequence of specific and miscognate DNAs on binding and cleavage are reversed. That is, whereas changing

the GAATTC flanking context results in large variability in binding affinities and no differences in cleavage rate constants, changing the AAATTC flanking context results in only minor effects on binding affinity and major effects on cleavage rate. The wild-type enzyme showed only a 2-fold (0.4 kcal/mol) range in binding affinities for the AAATTC site in the subset of flanking contexts. However, in changing the sequence flanking AAATTC, I observed a 300 – fold range in cleavage rate constants. Flanking sequence has a greater effect on A138T miscognate DNA binding affinity (15 – fold range) compared to the effect on wild-type binding, but the cleavage rates are much more dramatically affected (350 – fold range) by the context of the AAATTC site.

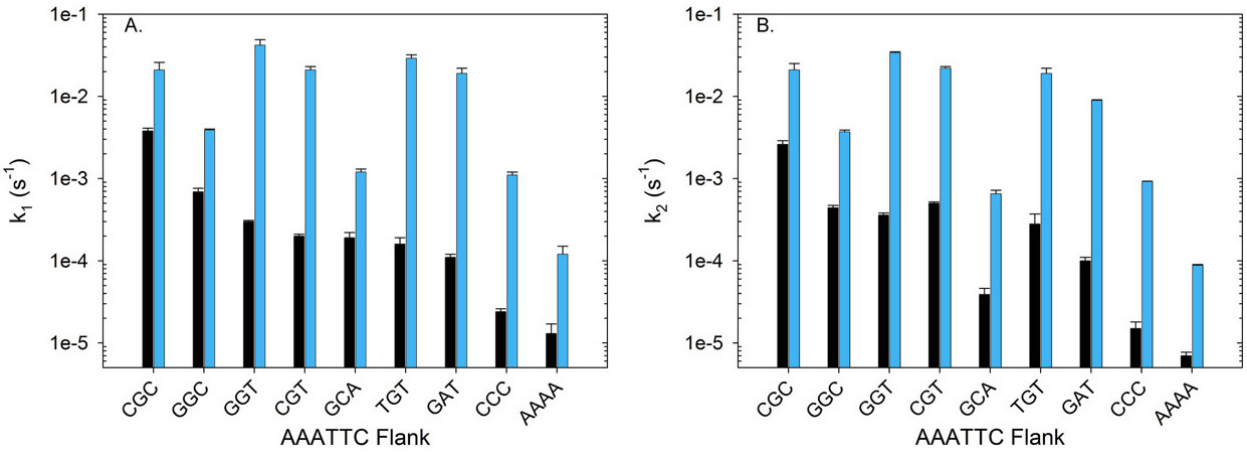


Figure 4.5 Effect of AAATTC flanking context on cleavage rates.

First order rate constants, k_1 (Panel A) and k_2 (Panel B) for cleavage of the AAATTC site in different symmetrical flanking contexts are shown. Wild-type data are black and A138T data aqua. The Y-axis is on a logarithmic scale. Raw data are given in Table 4.4.

Table 4.4 Effect of flanking context on miscognate DNA binding and cleavage

Sequence [a]	k_1 (s^{-1}) [b]	k_2 (s^{-1})	$k_{1,Mut.}/k_{1,WT}$ [c]	$\Delta\Delta G^{o}_{bind}$ (kcal/mol) [d]
Wild-type				ND[e]
CGCaaattcGCG	$3.8 \pm 0.3 \times 10^{-3}$	$2.6 \pm 0.3 \times 10^{-3}$	-	6.6 ± 0.1
GGCaaattcGCC	$6.9 \pm 0.7 \times 10^{-4}$	$4.4 \pm 0.3 \times 10^{-4}$	-	6.9 ± 0.1
GGTaaattcACC	$3.0 \pm 0.1 \times 10^{-4}$	$3.6 \pm 0.2 \times 10^{-4}$	-	7.0 ± 0.1
CGTaaattcACG	$2.0 \pm 0.1 \times 10^{-4}$	$5.0 \pm 0.2 \times 10^{-4}$	-	ND
GCAaaattcTGC	$1.9 \pm 0.3 \times 10^{-4}$	$3.9 \pm 0.7 \times 10^{-5}$	-	6.9 ± 0.1
TGTaaattcATA	$1.6 \pm 0.3 \times 10^{-4}$	$2.8 \pm 0.9 \times 10^{-4}$	-	ND
GATaaattcATC	$1.1 \pm 0.1 \times 10^{-4}$	$1.0 \pm 0.1 \times 10^{-4}$	-	6.9 ± 0.1
CCCaaattcGGG	$2.4 \pm 0.2 \times 10^{-5}$	$1.5 \pm 0.3 \times 10^{-5}$	-	ND
AAAAaaattcTTTT	$1.3 \pm 0.4 \times 10^{-5}$	$7.0 \pm 0.7 \times 10^{-6}$	-	6.8 ± 0.1
A138T				ND
GGTaaattcACC	$4.2 \pm 0.7 \times 10^{-2}$	$3.4 \pm 1.0 \times 10^{-2}$	140	7.2 ± 0.2
TGTaaattcATA	$2.9 \pm 0.3 \times 10^{-2}$	$1.9 \pm 0.3 \times 10^{-2}$	180	ND
CGCaaattcGCG	$2.1 \pm 0.5 \times 10^{-2}$	$2.1 \pm 0.4 \times 10^{-2}$	6	6.3 ± 0.1
CGTaaattcACG	$2.1 \pm 0.2 \times 10^{-2}$	$2.2 \pm 0.1 \times 10^{-2}$	110	ND
GATaaattcATC	$1.9 \pm 0.3 \times 10^{-2}$	$8.9 \pm 0.2 \times 10^{-3}$	170	7.3 ± 0.1
GGCaaattcGCC	$3.9 \pm 0.1 \times 10^{-3}$	$3.7 \pm 0.2 \times 10^{-3}$	6	7.9 ± 0.2
GCAaaattcTGC	$1.2 \pm 0.1 \times 10^{-3}$	$6.5 \pm 0.7 \times 10^{-4}$	6	7.2 ± 0.1
CCCaaattcGGG	$1.1 \pm 0.1 \times 10^{-3}$	$9.2 \pm 0.1 \times 10^{-4}$	46	ND
AAAAaaattcTTTT	$1.2 \pm 0.3 \times 10^{-4}$	$8.8 \pm 0.2 \times 10^{-5}$	9	7.9 ± 0.1

[a] Sequence of double stranded 24 base-pair oligonucleotide is 5'-GGGCGGGXXXAAATTCYYYGGCGC; where the underlined sequence is entered into column 1. Sequences are entered in order according to the magnitude of the cleavage rate constants for each individual enzyme.

[b] Means ± SD of ≥ 3 determinations; first-order strand scission constants were obtained under single-turnover conditions using 8 mM Mg2+ to initiate the reaction; k_1 represents cleavage in the cognate half-site, GAA and k_2 represents cleavage of the non-cognate half site (see figure 4.1).

[c] Ratio of cleavage rate constants for cognate half-site cleavage

[d] For each enzyme, the difference in the observed standard binding free energy between the AAATTC site in different flanking contexts and the specific site is calculated as

$\Delta\Delta G°_{bind} = -RT \ln (K_A^{miscognate}/K_A^{cognate})$ at 294 K., where the reference cognate site is in the CGC flanking context

[e] ND = not determined

Despite the fact that changing the AAATTC flanking context results in a large range in first order cleavage rate constants for both the wild-type and A138T endonucleases, closer inspection of the AAATTC flanking context cleavage rate profiles of the wild-type and A138T enzymes sheds light on one of the factors causing relaxed *in vivo* specificity of the mutant based on the following considerations. The wild-type AAATTC cleavage rate constants decrease significantly (6 –fold) from the best context (CGC) to the second (GGC) context in the hierarchy (Table 4.4). Further, wild-type cleavage rate constants for AAATTC drop 20-fold from the most favored CGC context to the fifth triplet (GCA) in the hierarchy. We propose that this steep gradation in cleavage rate constants between the best miscognate DNA flanking context and other flanking sequences represents an additional level of protection against *in vivo* cleavage of genomic miscognate sites by the *Eco*RI restriction endonuclease. This is due to the fact that of the 3386 AAATTC sites in the *E. coli* K12 genome, only 68 (2.0%) of the non-cognate half-sites, and 76 (2.2%) of the cognate half-sites are abutted in one strand by CGC, the context in which AAATTC is most likely to be cleaved. That is the vast majority of AAATTC sites in the *E. coli* genome are in contexts where the site is cleaved much more slowly by the wild-type enzyme than the sites in the optimum (CGC) context.

Might the cleavage preferences of the *Eco*RI endonuclease play a role in the evolution of the *E. coli* genome by providing a selective pressure for the avoidance of miscognate sites in flanking contexts where DNA cleavage is most likely? To properly address this question, one would need to ascertain whether or not the *E. coli* chromosome and the *Eco*RI RMS co-segregated for any significant length of evolutionary time. Given that the *Eco*RI RM genes are plasmid encoded (Betlach et al., 1976), it seems unlikely that this RM system was a permanent fixture in *E. coli* lineages. Nonetheless, a cursory glance at the distribution of flanking contexts of AAATTC sites in the *E. coli* genome shows that the number of sites (68 or 76 depending on the half-site) in the optimal context is *greater* than the number expected by chance (53). Furthermore, in general, there is no correlation between the abundance of AAATTC sites and flanking sequence cleavage preference by the *Eco*RI endonuclease. (These calculations assume a GC content of 50% in the *E. coli* genome. A proper analysis would take into account not only the actual GC content, but also the di- and tri-nucleotide frequencies in the *E. coli* genome.)

By contrast to the wild-type, the A138T enzyme cleaves the AAATTC miscognate site ‘rapidly’ in a large subset of flanking contexts (Figure 4.5). For the A138T enzyme the rate

constants decrease only 1.4-fold between its best (GGT) and second (TGT) flanking triplet, and decrease by a factor of two from the most favored (GGT) context to the fifth (GAT) triplet in the hierarchy (Table 4.4). This lack of gradation at the top of the hierarchy is expected to increase the probability of AAATTC cleavage in cells expressing the A138T enzyme relative to cells expressing the wild type.

It is difficult to quantify this *in vivo* effect for the following reasons: It has been shown that the base-pair triplet is the dominant modulator of *specific* DNA binding affinity for the *Eco*RI (David Chi, Michael Kurpiewski, and Linda Jen-Jacobson, *unpublished*), and *Bam*HI (Daniel Lai, Lisa Engler and Linda Jen-Jacobson, *unpublished*) endonucleases. It is reasonable to assume that the fourth base-pair removed from a miscognate site will also have little or no effect on DNA binding or cleavage. Nonetheless, we have currently only examined 9/64 possible flanking triplets, therefore our characterization of 'flanking sequence space' is not complete.

Further, we have measured the cleavage rates exclusively for DNAs with symmetrical flanking sequences, yet miscognate sites in the *E. coli* chromosome are almost always embedded in asymmetric flanking contexts. (For example, only four of the 144 AAATTC sites flanked in one half-site by CGC are also flanked in the other half-site by CGC). I cannot predict how an asymmetric flanking sequence will affect cleavage rates in the two strands of miscognate DNAs, which are them selves asymmetric. It is possible that the two flanking sequences abutting the two ends of a single miscognate site will independently affect the transition state for cleavage of the 'local' DNA half-site. Indeed, many subtle perturbations made to one strand of the specific site have no effect on cleavage of the unmodified strand. However, single base-pair substitutions (*Eco*RI* sites) are accompanied by a reduction in the cleavage rates of both the invariant and miscognate half-sites, supports the idea that the effects of this type of perturbation are communicated between the two DNA half-sites. Therefore, I believe it is more likely that the flanking sequence will not only affect cleavage of the local half-site, but may influence the transition state for cleavage of the distal half-site. The issue of miscognate site flanking sequence symmetry might also be important when considering miscognate sites for which the rates of cleavage are asymmetric. (For example, for TAATTC in the symmetrical CGC context, the k_1/k_2 ratio is 130 for the wild-type.) In such cases, what would be the effect of embedding the TAATTC in a DNA where the cognate half-site is flanked by an unfavorable triplet, and the

non-cognate half-site is flanked by a favorable triplet? This arrangement might result in increased symmetry of the two rate constants, which in turn could undermine the kinetic proofreading mechanism (Figure 4.1), and result in an increased probability of double stranded cleavage of the site. Questions regarding the effects of asymmetric flanking sequences on miscognate DNA binding and cleavage will be very labor intensive to untangle due to the large number of sequence permutations involved.

At present we have only completed this analysis of miscognate DNA flanking sequences using the AAATTC miscognate site with the wild-type and A138T endonucleases. Are equilibrium binding and cleavage rate constants for the other miscognate sites affected in a similar manner to that observed for AAATTC? Further, it will be interesting to ask if miscognate DNA cleavage by the remaining promiscuous mutants is also less affected by flanking context relative to that of wild type enzyme such that this may be a common factor contributing to the relaxed specificity in all of the promiscuous *Eco*RI enzymes.

4.2.5. Estimates of in vivo error rates correlate with phenotypes of mutants

The *in vivo* phenotypes produced by the three promiscuous mutations differed in severity, as judged by the DNA-damage assay of Heitman and Model (Heitman and Model, 1990) in which the *dinD* promoter drove *lacZ* expression (Heitman et al., 1987). Colonies containing the wild-type endonuclease were white, E192K colonies light blue, H114Y colonies medium blue, and A138T colonies dark blue.

To ask whether the *in vitro* thermodynamic and kinetic parameters of the mutant endonucleases are related to the *in vivo* phenotypes, we used our binding and cleavage data to make estimates of how many cuts at miscognate sites would occur in cellular DNA due to the presence of each of the *Eco*RI endonucleases. We used an *E. coli* generation time of 30 minutes, assumed equal expression of each enzyme, and assumed that miscognate DNA cleavage proceeds as a second order reaction *in vivo*. (The partitioning ratios in Table 4.2 support the supposition that the enzyme is not saturated with miscognate DNA.). We calculated the reaction velocity *V* for *in vivo* cleavage of miscognate DNA using equation 4.3:

$$V = [Enzyme]_{Cell} \times [Site]_{Cell} \times k_{cleave} \times K_{A,app,Cell} \qquad 4.3$$

where k_{cleave} is the first order *Eco*RI* site cleavage rate constant (raw data in Table 4.3), $[Site]_{t,Cell}$ is the *in vivo* DNA concentration for a given miscognate site, and $K_{A,\ App.,Cell}$ (see legend to Table 4.5 for equation used to determine this parameter) is the apparent miscognate DNA binding constant *in vivo*, taking into account competitive binding by non-specific DNA. These calculations do not account for the presence of polyamines and other co-solutes in the E. coli cytoplasm that affect DNA binding affinity (Record et al., 1998). Further, these calculations make the assumption that each of the miscognate sites in the *E. coli* genome is embedded in a symmetrical CGC flanking context. Thus we ignore the (large and probably important, see section 4.2.4) effect of flanking context on *in vivo* miscognate cleavage velocity. Therefore these calculations should be appreciated for their ability to make predictions on the relative number of *in vivo* cuts made by the four enzymes, not the absolute number of cuts. Estimates (number of cuts/cell/generation) obtained using equation 4.3 are given in Table 4.5. The calculated values show that all three mutant enzymes have increased probability of cutting at miscognate sites, with hierarchical order A138T > H114Y > E192K > wild-type. For example, the calculations in Table 4.5 suggest that A138T might cut all the 3386 AAATTC sites in the *E. coli* genome at least once in every generation, in contrast to the wild-type enzyme that would nick less than 1% of such sites. The discrepancy between enzymes is actually much greater than this, because for the wild-type enzyme, dissociation from an *Eco*RI* site after first-strand cutting is virtually certain at physiological salt concentration (Lesser et al., 1990), so double-strand cuts in a single binding event are extremely improbable. Double-strand cuts as a result of two independent binding events would occur at the vanishingly low frequency given by the product of the two very small single-strand nicking frequencies.

The nicking frequency at miscognate sites in cells expressing the wild-type endonuclease must be low enough that the single-strand nicks are repaired by DNA ligase, since the presence of the wild-type enzyme does not elicit the SOS response *in vivo* if the *Eco*RI methylase is present (Heitman and Model, 1990). By contrast, it is evident from the mutant phenotypes that the increased cutting capacity of the A138T and H114Y enzymes exceeds the “healing” capacity of DNA ligase, resulting in the conversion of nicks to double strand breaks, which are more difficult to repair (Heitman et al., 1999). The predictions of cutting capacity based on our data

thus correlate qualitatively with the severity of the promiscuous phenotypes (Table 4.5) of the endonuclease alleles observed by Heitman and Model (Heitman and Model, 1990).

The A138T and H114Y enzymes have cutting capacities for genomic miscognate DNA that far exceed that of the wild-type, but the E192K and wild-type enzymes are predicted to cleave some kinds of miscognate DNA sites at similar rates in vivo. (In this respect the E192K mutation is roughly analogous to the E34K mutant of the bacteriophage λ cI repressor (Nelson and Sauer, 1985), which shows improved binding to both operator and non-operator DNA, but has unaltered *in vivo* specificity relative to the wild-type repressor.) Why, then, does the E192K enzyme elicit an SOS response in *E. coli*? First, we have demonstrated that although the E192K mutation does improve nonspecific DNA binding, the partitioning ratios are nevertheless shifted towards miscognate binding for some *Eco*RI* sites (Table 4.2). Second, we only examined four of the nine possible *Eco*RI* sites, and it remains possible that the E192K mutation will show more dramatic deviations from the wild-type enzyme in binding and cleavage characteristics with the untested sites.

Table 4.5 Predicted number of cuts in E. coli chromosomal DNA made by EcoRI endonuclease in vivo

Sequence	Wild-type[a]	E192K	H114Y	A138T
5′-G↓AATT T[b]	23	72	650	4400
C TTAA↑A[c]-5′	15	39	480	3400
5′-G↓AATT A	76	190	770	16000
C TTAA↑T-5′	1	1	50	60
5′-G↓AATG C	<1	14	140	30
C TTAC↑G-5′	<1	3	38	3
5′-G↓AAGT C	3	190	110	2000
C TTCA↑G-5′	4	200	230	2200
LacZ Phenotype[d]	White	Light blue	Medium blue	Dark blue

[a] Number of miscognate DNA cuts per cell per generation, calculated with equation (2) in the text. Cleavage rate constants used are from Table 4, measured at 8 mM Mg^{2+}. The free Mg^{2+} concentration *in vivo* is likely in the range 2−4 mM.[55,56] The $K_D^{Mg^{2+}}$ is not well-characterized for EcoRI* sites, for either mutant or wild-type enzyme, but probably lies within the bounds of the wild-type/GAATTC value (3 mM)[2] and the wild-type/AAATTC value (~5 mM; our preliminary data). These considerations may imply that our calculations slightly overestimate all cutting rates. The *in vivo* endonuclease concentration is estimated to be 1 μM. The concentration of each individual miscognate site in the *E. coli* genome was obtained from analysis of the *E. coli* K12 genome sequence, (GenBank accession no. U00096). Site concentrations used were AAATTC, 3.5 μM; TAATTC, 2.1 μM; GCATTC, 2.8 μM; GACTTC, 1.9 μM. The calculation takes into account competitive binding by non-specific DNA:

$$K_{A,app,cell} = \frac{K_{A,int,miscognate}}{1 + [DNA]_t \times K_{A,int,NS}}$$

Binding and cleavage-rate constants used are for symmetrical CGC flanking triplets, but EcoRI* sites *in vivo* will be flanked by other triplets, usually asymmetrically. For wild-type enzyme, the CGC flanking triplet is most favored for cleavage rates, so actual cutting rates at sites in other flanking contexts will be lower. For A138T, context affects cleavage rates to a much lesser degree. Effects of context on H114Y and E192K are unknown.

[b] Cuts in unaltered half-site of miscognate site.

[c] Cuts in altered half-site of miscognate site.

[d] Data from Heitman & Model[23] referring to phenotypes of alleles of EcoRI endonuclease in the context of a strain containing a DNA damage inducible (*din*D) promoter fused to the *lac*Z gene. The strain also expressed the EcoRI methylase.

4.2.6. Barriers and pathways in the evolution of sequence specificity

The "promiscuity" of the mutant *Eco*RI endonucleases is perhaps best understood not as a loss of specific interaction with the canonical GAATTC site, but as a relaxation of discrimination against cleavage of *Eco*RI* sites. This is achieved by the dual expedients of increasing the binding affinity for *Eco*RI* sites, and using this additional binding energy efficiently on the path to the transition state for cleavage of the *Eco*RI* site. Our data show that mutations at several different sites can produce the same set of effects, albeit with somewhat different degrees of relaxation of discrimination. Could these kinds of mutations be exploited to engineer endonucleases with novel recognition sites? Heitman suggested (Heitman, 1993) that promiscuous mutants of the *Eco*RI endonuclease might represent a snapshot of evolution along the path to the evolution of a RMS with a new sequence specificity. It was hypothesized that the

hyper-mutable state in which the SOS response is activated would accelerate the generation of different endonuclease alleles, and these alleles could then be paired with pre-existing, newly acquired, or mutagenized methylase genes until a specificity "match" is found. The idea is intriguing; however, it seems to me that deleterious effects (slow growth, hyper-mutation of genome) associated with persistent DNA damage would put cells expressing relaxed specificity mutant enzymes at an extreme disadvantage. Therefore it seems more likely that catalytic inactivation of the endonuclease, or acquisition of methylase with altered or promiscuous specificity would precede the accumulation of alleles encoding endonucleases with changed or promiscuous specificity.

Nature has succeeded in creating a large diversity in DNA sequence specificities of type II restriction-modification systems (Roberts et al., 2003). Despite the weak sequence homologies among them, it has been proposed (Bujnicki, 2003), based on structural similarity (Johnson et al., 1990) that many of the type II restriction endonucleases arose from a common ancestor. It was suggested that the similarity of folds of two different proteins is insufficient to conclude common ancestry between proteins (Murzin, 1998), as there is believed to be a finite number of ways to compactly fold a chain such that the physicochemical constraints inherent in any polypeptide are satisfied. However, the fact that all type II restriction endonucleases carry out similar reactions, and have active sites of similar composition and geometry (Deva et al., 2001) lends additional support to the idea that type II restriction endonucleases indeed share a common ancestor. Alternatively or additionally, some members of this class of enzymes may have arisen by convergent evolution of structure (Wilson, 1991), for example, through the acquisition of catalytic power by DNA-binding proteins (Heitman, 1993). Thus, it remains uncertain whether or not mutational pathways exist by which the sequence specificities of type II restriction endonucleases can be altered during evolution.

The barriers to such alteration of recognition specificity can be appreciated by considering both the *in vivo* function of the restriction-modification system and the design principles of the protein-DNA interface that ensure this function is carried out efficiently and with minimal risk to the host cell. Any mutation that causes the sequence specificity of the endonuclease to diverge from that of the companion methylase would be deleterious at best, and lethal at worst. One option is that before an endonuclease gene can set forth upon the path of specificity modification through mutation, its synthesis and/or catalytic activity must be

eliminated or greatly attenuated. In any event, before the modified endonuclease can become fully active towards a novel target site, the cell must generate or acquire a methylase to protect the new target sequence.

Single amino acid changes in *Eco*RI endonuclease have been sufficiently well studied (Yanofsky et al., 1987, Alves et al., 1989, Needels et al., 1989, Hager et al., 1990, Fritz et al., 1998) so we can be confident that single mutations cannot cause a true change of DNA sequence specificity. Recognition is not so simple as a one-to-one interaction between a single protein side-chain and a DNA base. While we cannot predict definitively whether there exists some particular combination of a few mutations that might cause the endonuclease to recognize a different DNA sequence with high specificity, the obstacles to producing such a mutational combination are well understood and must be regarded as formidable. First, restriction endonucleases bind tightly to their specific recognition sequences, and use the same intermolecular contacts to stabilize the ground-state specific DNA complex and to stabilize the transition state, with the consequence that DNA sequence recognition is tightly coupled to cleavage. Therefore, the (hypothetical) mutational changes to the endonuclease must not only impart increased binding affinity for the new target DNA site, but also result in tight binding to the transition state. This set of requirements is further constrained because the cleavage site lies within the recognition sequence, so each individual substitution must be simultaneously well suited both to the ground state (recognition) and the transition state (cleavage). Furthermore, some individual side-chains participate not only in recognition but also in the catalytic mechanism itself (Jen-Jacobson, 1997, Kurpiewski et al., 2004), and thus cannot be altered without severe reduction of catalytic activity (Wolfes et al., 1986). This stringent set of requirements would serve to narrow the range of choices dramatically.

Second, a significant contribution to sequence specificity is provided by “indirect readout” of the specific DNA sequence (Otwinowski et al., 1988) through the sequence-dependent conformational proclivities of the DNA itself. For the many endonucleases that distort their specific DNA recognition sites, DNA distortion is critical for achieving the juxtaposition of protein and DNA groups that promotes efficient catalysis (Kurpiewski et al., 2004). It has been postulated that the energetic cost of DNA distortion is smallest for the specific sequence and that type II restriction endonucleases discriminate against miscognate sites at least in part due to the prohibitive cost of distorting miscognate DNA sites into catalytically

competent conformations (Lesser et al., 1990, Jen-Jacobson, 1997). Support for this view is provided by the strikingly similar DNA "kinks" found in the specific complexes of *Eco*RI endonuclease with GAATTC and of *Mun*I endonuclease (Deibert et al., 1999) with CAATTG, suggesting that some intrinsic mechanical property of the AATT module makes it amenable to this particular DNA distortion. Thus, a change to recognition of a different DNA site would require amino acid substitutions that stabilize a new DNA conformation, resulting in optimum protein-DNA complementarity in both the ground and transition states. To maintain stringent discrimination, these substitutions must simultaneously destabilize complexes with other sequences. We still know far too little about the physics of DNA conformations to predict how this set of requirements might be met.

Third, it has been shown that a conformational transition of the *Eco*RI endonuclease is coupled to specific binding (Grigorescu, 2003). This conformational change is required to assemble the catalytic sites (Grigorescu, 2003, Kurpiewski et al., 2004) and we have suggested that a change with these particular, precise features does not occur upon binding to miscognate or nonspecific DNA sites. Hence, a specificity change would also require amino acid substitutions that uncouple this conformational change from GAATTC and couple the conformational change to binding of the new recognition sequence.

Changing the specificity of the *Eco*RI restriction endonuclease from GAATTC to some other site would therefore require at least three types of amino acid changes: (a) those affecting direct readout of the bases in the target sequence; (b) changes that stabilize sequence-dependent conformational features of the bound DNA that are required for efficient catalysis of the new recognition sequence; (c) changes that couple conformational transitions in the protein to recognition of the new sequence. These classes of residues are no more than partially overlapping, so that three or more substitutions are likely to be required before a true novel specificity is generated. This set of requirements is certainly not unique to the *Eco*RI system, as existing information on other type II endonucleases suggests the same kinds of constraints (Engler, 1998, Deibert et al., 1999, Martin et al., 1999).

The degree to which the cooperative interface constrains sequence versatility can be appreciated by contrast with the Type I restriction-modification systems, which are much better suited for diversification of sequence specificity (Murray, 2000). First, type I endonucleases function as multi-protein complexes consisting of five polypeptides, where a single target

recognition subunit confers specificity to the two methylase and endonuclease subunits of the holoenzyme. Thus, the specificity of the endonuclease and methylase can be changed concomitantly, without subjecting the genomic DNA of the bacterial cell to restriction. Second, there are two target recognition domains (TRDs) within the target recognition subunit. Each of these TRDs is responsible for recognizing one half of a bipartite recognition sequence. This modular design of the recognition subunit has allowed investigators (and presumably Nature) to engineer novel specificities by creating chimeric recognition subunits containing a single TRD from different enzymes with different specificities (Nagaraja et al., 1985, Murray, 2000). In addition, it has been shown that changing the spacing between the two TRDs in a specificity subunit results in predictable changes in the optimal spacing between the two DNA hemi-recognition sites (Gubler et al., 1992). Finally, bacteria have evolved mechanisms to ensure that acquisition of a new type I system, or a change in specificity of an existing system, does not result in restriction of the host genome (O'Neill et al., 1997). Although the plasticity of the specificity of type I restriction systems gives insight into the evolution of these systems and the arms race between the bacterial cell and foreign DNA, researchers cannot take advantage of this plasticity to create new tools for genetic engineering, because the type I restriction endonucleases cleave DNA neither close to their recognition sites nor at a fixed distance away.

The characterization of these promiscuous mutant forms of *Eco*RI endonuclease defines a common molecular basis by which single-base mutations at diverse locations in a type II restriction endonuclease can have phenotypic consequences that are sub lethal to lethal. The fact that the occurrence of such lethal mutations is implicit in the protein structure itself suggests that type II restriction-modification systems may be “selfish” not only in the sense that they resist loss of the plasmids that carry them (Kulakauskas et al., 1995, Naito et al., 1995) but also in the sense that they strongly resist alteration of their own “identities” as defined by their sequence recognition specificities.

4.3. Conclusions

In this chapter, I presented data that quantify binding and cleavage discrimination by the wild-type and promiscuous mutant endonucleases and provided an explanation for the relaxed *in*

vivo specificity exhibited by the mutant enzymes. Our *in vitro* measurements confirm that the promiscuous mutants do not represent cases of changed specificity, but rather cases of relaxed specificity as the mutant enzymes still prefer the GAATTC site over any other site in both the binding and cleavage steps of the reaction. However, we have found that the promiscuous mutants manage to by-pass many of the fail-safes in place to protect genomic *Eco*RI* sites from cleavage by the wild-type endonuclease in the cell. The fact that mutant enzymes maintain a specific activity comparable to the wild-type endonuclease and cells expressing the promiscuous mutants are alive (though sick) demonstrates that the *Eco*RI methyltransferase protects genomic GAATTC sites from cleavage by the mutant endonucleases.

Each of the mutant enzymes tested binds more favorably to miscognate sites and cleaves miscognate DNAs more rapidly than the wild-type endonuclease. Tighter miscognate binding and faster miscognate DNA cleavage collectively undermine the 'kinetic proofreading' failsafe (Figure 4.1) which contributes to the protection of miscognate sites against double stranded breaks in cells expressing the wild-type endonuclease. It has been shown that the *Eco*RI endonuclease pauses when it encounters a miscognate site while undergoing linear diffusion along the double helix (Jeltsch et al., 1994). The basic tenet of the kinetic proofreading hypothesis states that for miscognate DNAs, the length of this pause (complex life-time) is much shorter than k_3 (rate constant for cleavage of the non-cognate half-site of previously nicked miscognate DNA site see figure 4.1) such that the endonuclease will dissociate from the nicked substrate prior to cleavage within the miscognate half-site. Thus, the worst outcome of a single miscognate DNA binding event is nicking of the cognate half-site, which can be repaired by DNA ligase. Tighter miscognate DNA binding by the mutants most likely translates into higher kinetic stability of these complexes, although this has yet to be experimentally verified. Longer mutant miscognate complex lifetimes coupled with the faster cleavage rates of miscognate half-sites of *Eco*RI* sites would result in a greater likelihood of both single and double-stranded cleavage of miscognate DNAs by the mutant enzymes compared to the wild-type enzyme. In addition, the A138T and H114Y mutations increase the amount of cellular endonuclease predicted to be bound to miscognate DNA relative to non-specific DNA, thereby reducing the ability of non-specific DNA to compete with miscognate sites for endonuclease binding. Also, in section 4.2.4 we showed that a steep reduction in DNA cleavage rate occurs when the AAATTC flanking context departs from the most favored CGC flanking triplet. The fact that

only 2% of the AAATTC sites in the *E. coli* genome are flanked by this triplet in one strand lead us to propose that modulation of miscognate DNA cleavage rate by flanking sequence represents an unappreciated level of protection against cleavage of miscognate DNAs in the cell. The A138T enzyme is predicted to partially circumvent this fail-safe since mutant AAATTC cleavage rates are unchanged in many flanking contexts. Finally, calculations which utilize the *in vitro* measurements presented in this chapter to predict the number of nicks at miscognate sites within the *E. coli* chromosome correlate well with the level of SOS induction of the mutant alleles. Thus we feel we have a good understanding of the factors leading to the increased probability of miscognate DNA cleavage in cells expressing the mutant enzymes.

Our approach of studying three different mutants has resulted in a greater understanding of factors leading to relaxation of *in vivo* specificity in the *Eco*RI endonuclease than had we chosen to study a single mutant. For example, the E192K enzyme is only enzyme whose non-specific binding mode is of higher affinity than that of the wild-type. This observation underscores the role on non-specific DNA endonuclease specificity in the cell, and is likely responsible for why the E192K mutant is the least promiscuous of the three enzymes studied. In addition, each of the three mutants has a different cleavage preference among the subset of miscognate DNAs examined (Table 4.3). Obtaining structural and/or energetic correlates to these differences represents an opportunity to gain additional insights into the specificity of the *Eco*RI endonuclease.

5. TOWARDS A MOLECULAR UNDERSTANDING OF RELAXED SPECIFICITY EXHIBITED BY THE PROMISCUOUS EcoRI ENDONUCLEASE MUTANTS

5.1. Introduction

The *Eco*RI-specific DNA complex in the absence of divalent metal cofactor is termed a pre-transition state complex, in which the enzyme-DNA contacts, as well as the conformational properties of the enzyme and DNA are similar to those characteristic of the transition state complex. This is supported by the observation that a comprehensive set of perturbations of the GAATTC DNA site (e.g. synthetic base analog substitutions that remove single or multiple functional groups) alter binding free energy but do not affect the free energy of activation for the chemical step; that is there is little or no effect on the rate constant for cleavage (Jen-Jacobson, 1995, Jen-Jacobson, 1997) (Figure 5.1A). This implies that binding free energy is conserved in the transition state, and that many energetic factors contribute equally to the specific recognition complex and the transition state complex. By contrast, single base-pair changes to the GAATTC site not only exact large binding free energy penalties, but also produce severe reductions in cleavage rates relative to that of the specific site (Figure 5.1B). The magnitudes of these binding penalties are too large to originate solely from disruption of one or a few enzyme-DNA base contacts. Therefore miscognate complexes are inferred to differ from the specific complex not only in protein-base contacts, but also with respect to phosphate contacts, DNA conformation, and enzyme conformation (Lesser et al., 1990). These structural deviations serve to minimize free energy for binding to miscognate DNA, but the adaptations are off the path to the transition state. This is manifest as a reduction in miscognate DNA cleavage rates relative to the GAATTC site.

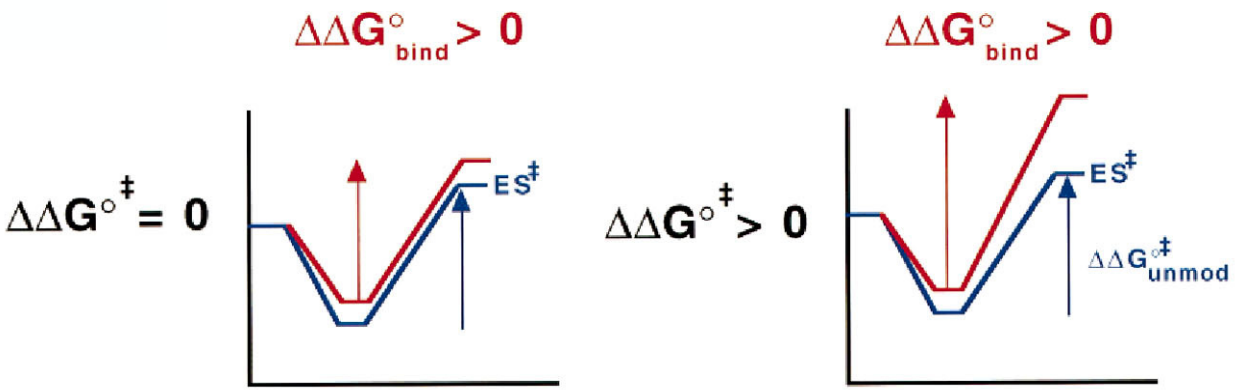


Figure 5.1 Reaction coordinate diagrams showing how perturbations in the protein or DNA may affect endonuclease binding ($\Delta G°_{bind}$) and cleavage ($\Delta G°^{‡}$).

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, ($\Delta G°^{‡}$) for the unmodified substrate is indicated by the blue arrow. The red reaction coordinate on the left is representative of minor perturbations to the GAATTC site where binding affinity, the activation free energy change is unaffected (cleavage rates are the same). In the right panel, the blue and red coordinates represent specific and miscognate reactions respectively. Binding to miscognate DNA entails not only binding penalties, but an increase in activation energy (cleavage rates are reduced). The free energy scales in the right and left panels are not meant to be the same. Figure taken directly from (Jen-Jacobson, 1997).

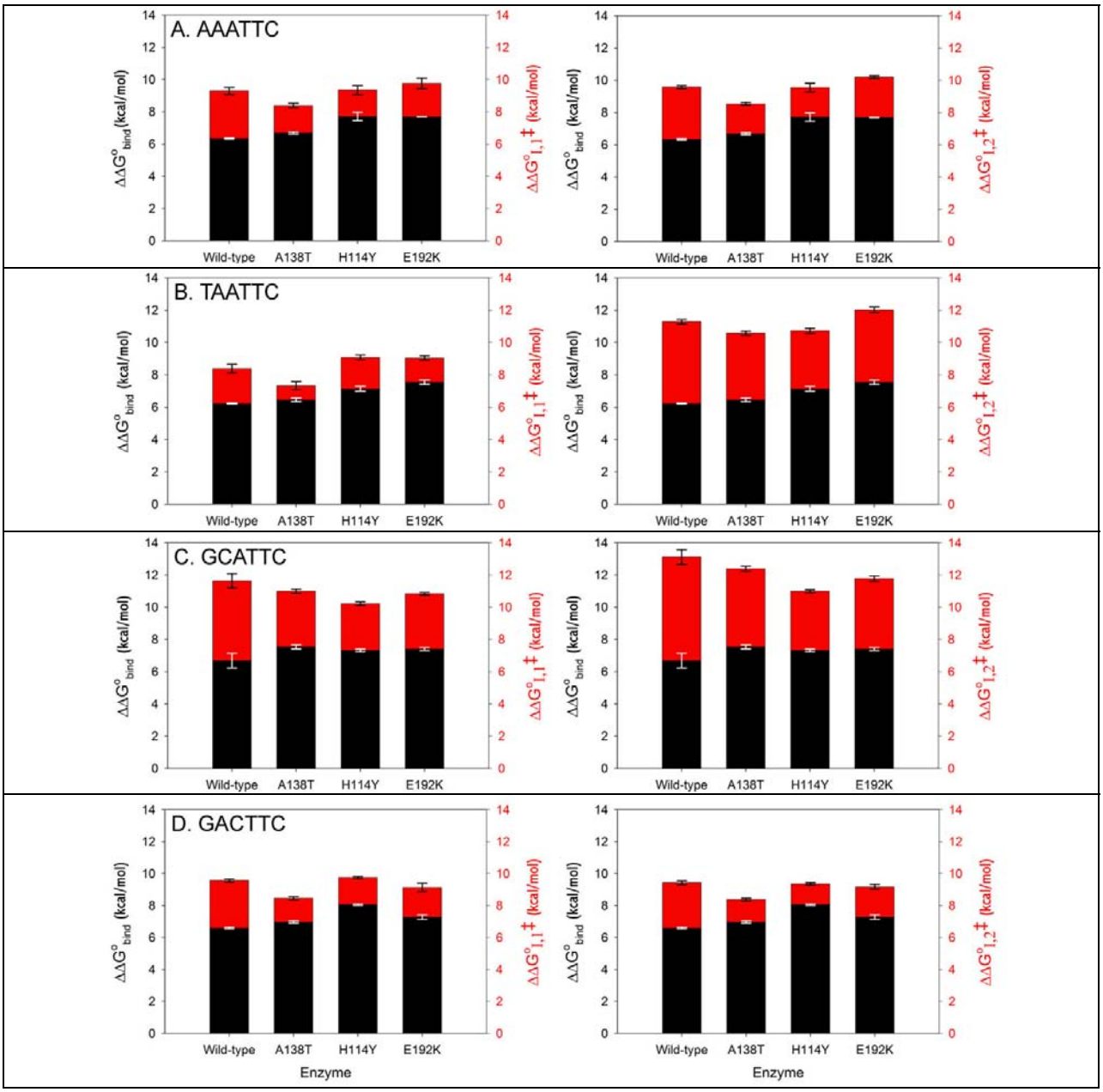


Figure 5.2 Promiscuous mutant enzymes preferentially stabilize transition states for cleavage of miscognate DNA sites.

The left hand axis and the corresponding black bars refer to binding discrimination, where data for each enzyme are referenced ($\Delta\Delta G°_{bind} = 0$) to the interaction with a specific DNA (CGC flanking context) (raw data in table 4.1). The right handed axis (bars in red) refers to the transition state interaction free energy discrimination, which takes into account both binding and cleavage (see equations 1.1 – 1.4). Note that the red bars also have their origin at 0. Each row presents data for a single miscognate site. Within each row, the left and right panels plot transition state interaction free energy discrimination for cleavage of the cognate and non-cognate half-site of miscognate DNAs respectively (raw cleavage data in table 4.3). Within each row, $\Delta\Delta G°_{bind}$ data (black bars) are the same in both panels.

In chapter 4, I showed that the miscognate complexes involving the promiscuous mutants share many characteristics with wild-type *Eco*RI-miscognate complexes (i.e. binding penalties and reduction in cleavage rates relative to specific DNA). However, while the mutant enzymes incur greater penalties (relative to GAATTC) for binding to miscognate DNA (Figure 5.2), the mutants cleave miscognate DNA sites more rapidly than the wild type enzyme. Taken together, these observations show that the mutant miscognate DNA complexes, like their wild-type counterparts, are structurally adaptive, yet the binding free energy in mutant complexes is more efficiently utilized in the transition state. What then are the molecular features of promiscuous *Eco*RI-miscognate DNA complexes that allows for their preferential transition state stabilization? In this chapter I present data using a battery of different experimental probes: a) the effect of Ca^{2+} on miscognate binding; b) the affinity for Mg^{2+} in miscognate complexes; c) the effect of flanking sequence on miscognate DNA binding and cleavage; and d) ethylation interference footprints on miscognate DNA, that together provide insight into the molecular basis for the observed differences in the miscognate DNA binding modes of the wild-type and the A138T promiscuous mutant enzymes.

5.2. Active sites of A138T-miscognate DNA complexes are more assembled than wild-type counterparts

X-ray crystal structures of specific complexes solved in the presence of divalent metal, have shown that Ca^{2+} ions are coordinated by the negatively charged residues constituting the active sites of restriction endonucleases (Horton et al., 1998, Viadiu and Aggarwal, 1998, Horton and Cheng, 2000). This is in accord with the findings that calcium is a strong inhibitor of DNA cleavage by the *Eco*RI (Figure 5.3), and *Bam*HI restriction endonucleases (Engler et al., 2001). Inclusion of $CaCl_2$ in the specific DNA binding reactions results in dramatic enhancements in binding affinity relative to conditions where the metal is omitted (Vipond et al., 1995, Engler et al., 1997, Engler, 1998, Kurpiewski et al., 2004), presumably due to the fact that by binding to the active site, Ca^{2+} relieves the electrostatic repulsion that is produced when the scissile phosphate is apposed to the negatively charged active site residues. In addition, the steep pH dependence (improved DNA binding with reduced solution pH) of specific binding exhibited by

restriction endonucleases is almost eliminated in the presence of Ca^{2+} (Figure 5.4A), suggesting that titration of active site residues is largely responsible for the pH dependence in the absence of divalent metal.

Interestingly, the pH dependence of *Eco*RV binding to miscognate DNA is much shallower than the pH dependence of specific binding, and the *Eco*RV enzyme shows no pH dependence for binding to non-specific DNA (Figure 5.4B) (Engler et al., 1997). These solution data suggest that the geometry of the active site, or the distance between the active site residues and DNA backbone differ among the three types of complexes. Tight, complementary interfaces that are observed in the crystal structures of specific complexes where the scissile phosphate is juxtaposed with the active site residues are consistent with a steep pH dependence of binding, and a large calcium binding enhancement. By contrast, a computational model of the *Bam*HI endonuclease in complex with non-specific DNA (Sun et al., 2003) in which the active site residues Glu113 and Glu77 are 22 Å and 27 Å away from the DNA backbone respectively is consistent with the shallow (relative to specific and miscognate DNA) pH dependence observed for non-specific binding (Lisa E. Engler and Linda Jen-Jacobson, *unpublished*). There are no high resolution structures of true miscognate complexes, but pH dependence data (Figure 5.4B), and other metrics (discussed in detail in section 1.3.3.2, Table 1.3) suggest that the miscognate complexes are intermediate between specific and non-specific ones with respect to active site assembly and protein-DNA complementarity.

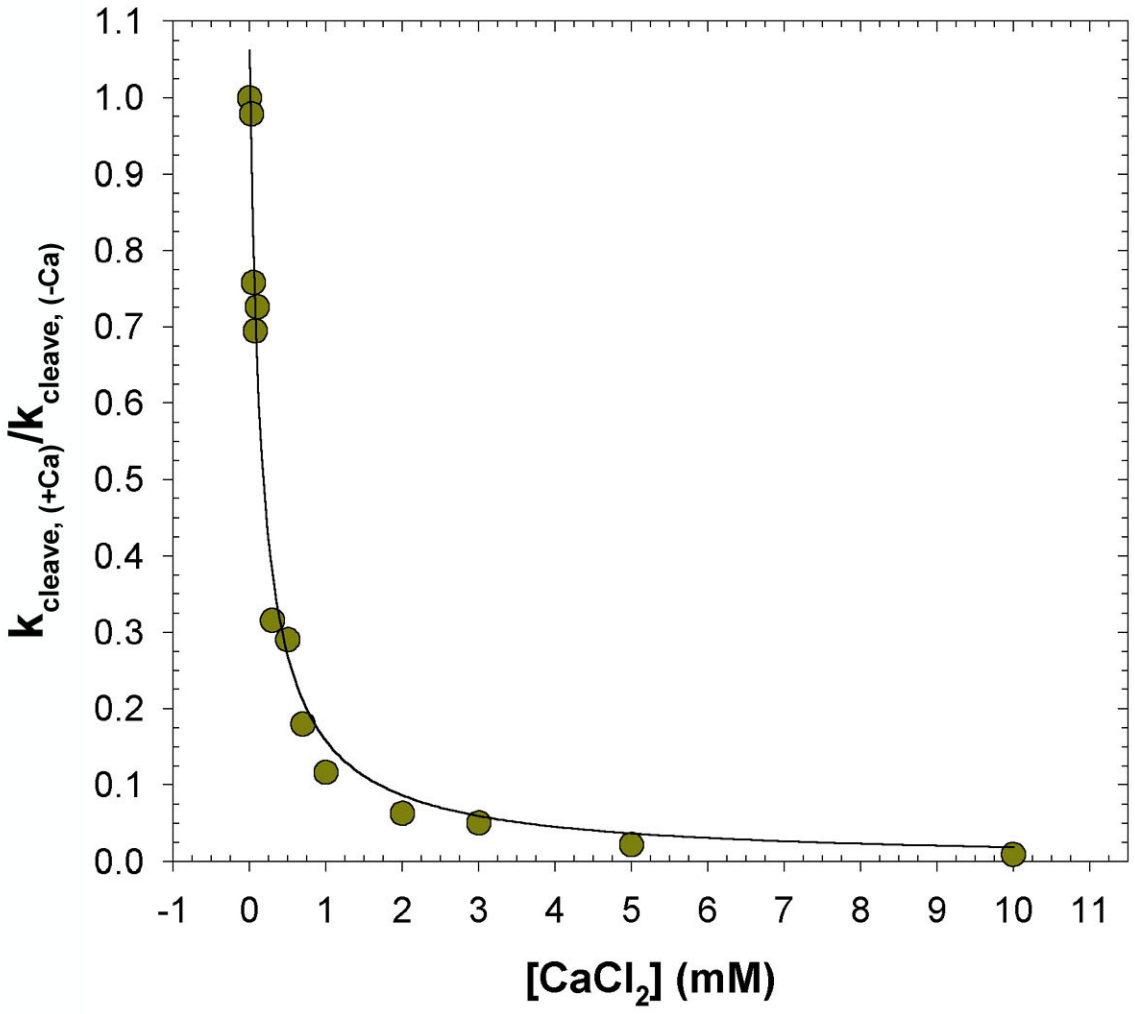


Figure 5.3 Ca2+ inhibits DNA cleavage by competing with Mg2+ for active site binding.

First order rate constants for cleavage of the double stranded DNA containing the specific recognition sequence (5'-GATGGGTGCAGAATTCTGCAGGTA) in 8 mM $MgCl_2$ plus varying concentrations of $CaCl_2$. From fitting these data, I find a K_I of 29 μM for Ca^{2+} binding at 0.1 M NaCl.

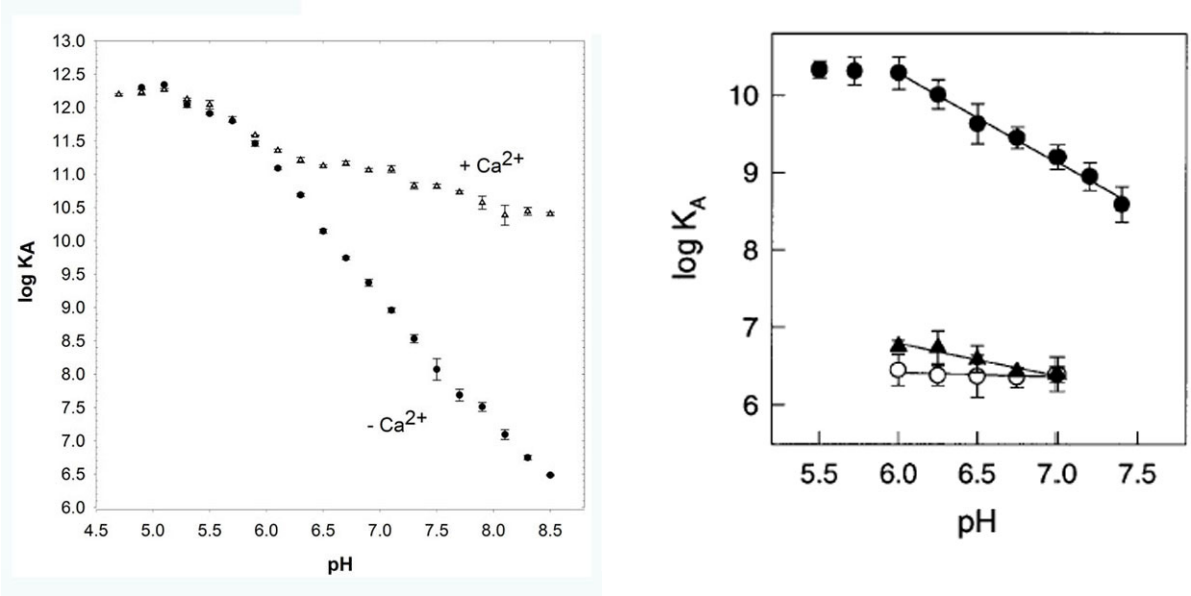


Figure 5.4 pH dependence of endonuclease binding as a probe for active site assembly.

Left) The pH dependence of the *Eco*RI endonuclease specific binding interaction is shown on the left. Inclusion of Ca^{2+}, which is coordinated by the active site, severely reduces the pH dependence specific binding, indicating titration of active site residues is the dominant contributor to the pH dependence.

Right) pH dependence of the *Eco*RV endonucleases binding to the specific (●), miscognate (▲), and non-specific (○) DNA sites. Right hand panel taken directly from (Engler et al., 1997).

Might improved active site assembly in A138T-miscognate DNA complexes lead to the faster cleavage rates observed for the mutant enzyme? To test this, we measured miscognate DNA binding affinities in the presence and absence of Ca^{2+}, and used Ca^{2+} induced binding enhancement as a probe to report on active site assembly in wild-type and A138T complexes with the different classes of sites. In the following discussion, an 'assembled active site' refers to a complex in which the proper juxtaposition of the active site residues and the scissile phosphate has been achieved.

Table 5.1 Effect of CaCl2 on DNA binding discrimination by the *Eco*RI endonuclease

Site [a]	$K_{A,\ -CaCl2}$ [b] (M^{-1})	$K_{A,\ mutant}/K_{A,\ wild\text{-}type}$ ($-CaCl_2$)	$\Delta\Delta G°_{bind}$ (site) [c] (kcal/mol)	$K_{A,\ +CaCl2}$ [d] (M^{-1})	$\Delta\Delta G°_{bind}$ (site + $CaCl_2$) (kcal/mol) [f]	$K_{A,\ mutant}/K_{A,\ wild\text{-}type}$ ($+CaCl_2$)	Fold $CaCl_2$ Enhancement [e]
Wild-type						-	
GAATTC	$9.6\ (\pm 0.3)\times10^{9}$	-	0	$5.0\ (\pm 0.3)\times10^{12}$ §	0	-	520
AAATTC	$1.9\ (\pm 0.1)\times10^{5}$	-	6.3 ± 0.1	$2.3\ (\pm 0.3)\times10^{6}$	8.5 ± 0.1	-	12
TAATTC	$2.3\ (\pm 0.1)\times10^{5}$	-	6.2 ± 0.1	$8.4\ (\pm 0.7)\times10^{5}$	9.1 ± 0.1	-	4
GCATTC	$1.0\ (\pm 0.8)\times10^{5}$	-	6.7 ± 0.5	$8.3\ (\pm 1.7)\times10^{4}$	10.5 ± 0.1	-	1
GACTTC	$1.2\ (\pm 0.1)\times10^{5}$	-	6.6 ± 0.1	$6.3\ (\pm 0.4)\times10^{5}$	9.3 ± 0.1	-	5
CTTAAG (NS)	$1.0\ (\pm 0.1)\times10^{5}$	-	6.7 ± 0.1	$5.6\ (\pm 2.0)\times10^{4}$	10.7 ± 0.2	-	1
A138T							
GAATTC	$1.3\ (\pm 0.1)\times10^{11}$	14	0	$4.7\ (\pm 1.4)\times10^{13}$ §	0	9	360
AAATTC	$1.4\ (\pm 0.1)\times10^{6}$	7	6.7 ± 0.1	$3.1\ (\pm 0.5)\times10^{7}$	8.3 ± 0.2	13	22
TAATTC	$2.1\ (\pm 0.4)\times10^{6}$	9	6.5 ± 0.1	$3.3\ (\pm 0.2)\times10^{7}$	8.3 ± 0.2	39	16
GCATTC	$3.3\ (\pm 0.6)\times10^{5}$	3	7.5 ± 0.5	$8.0\ (\pm 1.2)\times10^{5}$	10.5 ± 0.2	10	2
GACTTC	$8.7\ (\pm 0.6)\times10^{5}$	7	7.0 ± 0.1	$1.1\ (\pm 0.6)\times10^{7}$	8.9 ± 0.4	17	13
CTTAAG (NS)	$1.0\ (\pm 0.1)\times10^{5}$	1	8.3 ± 0.1	$1.3\ (\pm 0.5)\times10^{5}$	11.5 ± 0.3	2	1

[a] Each site is embedded in a double stranded 24 base-pair oligodeoxynucleotide: 5'-GGGCGGGCGC<u>XXXXXX</u>GCGGGCGC, where <u>XXXXXX</u> represents the site listed in column 1.

[b] Equilibrium association constants were measured in binding buffer plus 0.22 M KCl (pH 7.3, 21°C); means ± std. dev. of ≥ 3 determinations.

[c] For each enzyme, the observed difference in standard binding free energy between each variant and specific site is calculated as $\Delta\Delta G°_{bind\ (Site)} = -RT \ln (K_A^{variat}/K_A^{specific})$ at 294 K.

[d] Equilibrium association constants were measured in binding buffer plus 0.21 M KCl, 5 mM $CaCl_2$ (pH 7.3, 21°C); means ± std. dev. of ≥ 3 determinations.

[e] For each enzyme, the fold $CaCl_2$ enhancement is: K_{+Ca}/K_{-Ca}.

[§] Values entered into the table are extrapolated from experimentally determined values measured in binding buffer plus 0.35M KCl, 5 mM $CaCl_2$ (pH 7.3, 21°C). I used a salt dependence (log K_A vs. Log [KCl]) of -11.8 to calculate the extrapolated value. The salt dependence of binding of the wild-type and A138T enzymes are the same in the presence and absence of $CaCl_2$.

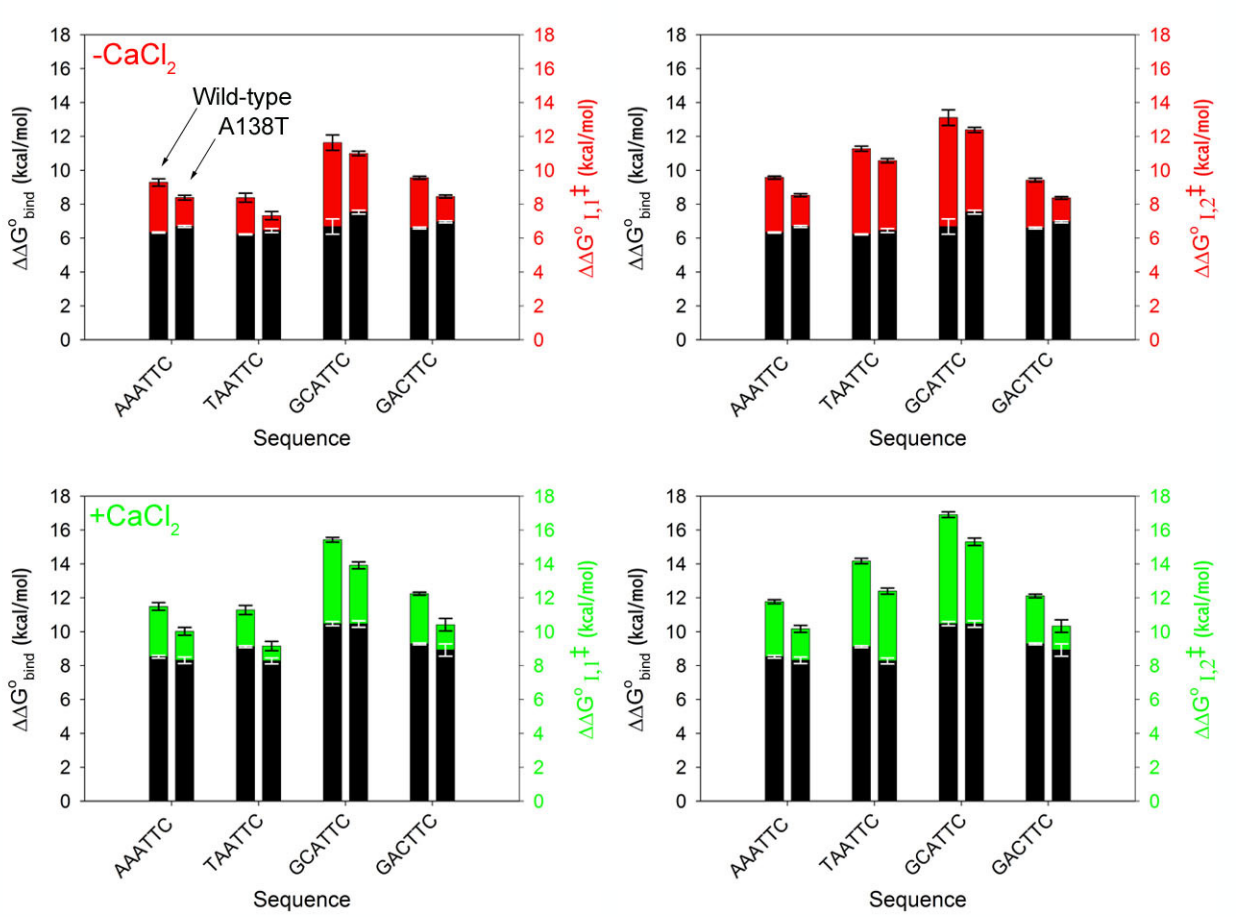


Figure 5.5 Calcium reduces discrimination by the A138T enzyme relative to the wild-type.

In all panels, the left hand axis and the corresponding black bars refer to binding discrimination, where data for each enzyme are referenced ($\Delta\Delta G°_{bind} = 0$) to the interaction with specific DNA (CGC flanking context). Note error bars for $\Delta\Delta G°_{bind}$ are shown in white. The right handed axis (bars in red or green) refers to the transition state interaction free energy discrimination, which takes into account both binding and cleavage (see equations 1.1 – 1.4). Note that the red and green bars also have their origins at 0 on the y-axis. The top two panels use equilibrium association constants measured in the absence of calcium (raw data in table 4.1), and the bottom two panels use constants measured in the presence of calcium (raw data in table 5.1). Within each row, the left and right panels plot transition state interaction free energy discrimination for cleavage of the cognate and non-cognate half-site of miscognate DNAs respectively (raw cleavage data in table 4.2). Within each row, $\Delta\Delta G°_{bind}$ data (black bars) are the same in both panels. In each panel, a sequence has a set of two bars where the left and right bars are the wild-type and A138T enzymes, respectively.

Specific binding affinity is dramatically improved for both the wild-type and A138T enzymes in the presence of Ca^{2+} relative to the condition in which divalent metal is absent (Table 5.1). Wild-type and A138T specific DNA binding affinities are increased by 520-fold ($\Delta\Delta G°_{bind}$ = -3.7 kcal/mol) and 360-fold ($\Delta\Delta G°_{bind}$ = -3.4 kcal/mol) by Ca^{2+} respectively. On the other hand, non-specific DNA binding by both enzymes is unaffected by Ca^{2+}, consistent with the inference that the active site is not assembled in a non-specific complex. Although miscognate DNA binding is improved by the inclusion of divalent metal, the binding enhancement is much less than that observed for specific binding. Therefore binding discrimination by both enzymes is *increased* in the presence of Ca^{2+} (compare columns 4 and 6 in Table 5.1, see also Figure 5.5). However, Ca^{2+} improves miscognate DNA binding by the A138T enzyme to a greater extent than by the wild-type. For example, the A138T enzyme exhibits a 16-fold binding enhancement (relative to binding in the absence of Ca^{2+}) for the TAATTC site, but Ca^{2+} induces only a 4-fold binding enhancement for the wild-type enzyme. Qualitatively similar differences between the wild-type and A138T enzymes are observed with the other miscognate sites tested (Table 5.1).

In the absence of Ca^{2+}, the A138T mutant exhibits more stringent binding discrimination against miscognate sites than the wild-type (Table 5.1, Figure 5.5). By contrast, in the presence of Ca^{2+}, the A138T enzyme shows decreased discrimination against binding to the TAATTC site relative to the wild-type; compare columns 4 and 6 (Table 5.1), and the two enzymes discriminate similarly against binding to the remaining miscognate sites. In addition, since Ca^{2+} preferentially stabilizes the ground state of mutant miscognate complexes relative to wild-type, it follows that the transition state interaction free energy discrimination ($\Delta\Delta G°_{I}^{\ddagger} = \Delta\Delta G°_{bind} + \Delta\Delta G°^{\ddagger}$) is also preferentially reduced for the mutant (Figure 5.5). Interestingly, miscognate DNA binding enhancement is correlated with cleavage rate and the symmetry of cleavage rates in the two different DNA half-sites (k_1/k_2 ratio). That is sites which are cleaved the fastest, and cleaved symmetrically in both DNA half-sites ($k_1/k_2 \cong 1$) exhibit the largest Ca^{2+} induced binding enhancements. Taken together, these observations support the ideas that the effect of Ca^{2+} on DNA binding, like pH dependence of binding is a good indicator of active site assembly, and that the active sites of A138T-miscognate complexes are more assembled than the wild-type counterparts.

When interpreting the binding data measured in the presence of Ca^{2+}, we must bear in mind that Ca^{2+} does not promote DNA cleavage, so it is possible that the differences in active

site configuration that produce greater Ca^{2+} binding enhancements for the A138T enzyme are off the path to the transition state. To test this, we asked whether A138T-miscognate DNA complexes exhibit improved binding to Mg^{2+}, the physiologically relevant co-factor which promotes DNA cleavage. By measuring the dependence of first order cleavage rate on Mg^{2+} concentration, we can obtain the apparent binding affinity of an enzyme-DNA complex for Mg^{2+}. Since DNA cleavage is the readout in this assay, it represents a functional test that can determine whether differences in active site assembly in the ground state (measured by Ca^{2+} binding enhancement) are translated to differences in binding to the Mg^{2+} co-factor.

Both half-sites of the A138T-AAATTC miscognate DNA complex have higher affinity for Mg^{2+} ($K_{D1\ App.\ Mg2+}$ = 1.5 ± 0.4 mM, $K_{D2\ App.\ Mg2+}$ = 2.8 ± 1.0 mM) than do the half-sites in the wild-type complex with the same sequence ($K_{D1\ App.\ Mg2+}$ = 10.8 ± 3.5 mM, $K_{D2\ App.\ Mg2+}$ = 8.0 ± 2.3 mM) (Figure 5.6A&B). ($K_{D1\ App.\ Mg2+}$ and $K_{D2\ App.\ Mg2+}$ refer to the Mg^{2+} dissociation constants for the cognate and non-cognate DNA half-sites of a miscognate DNA.) Note that the Mg^{2+} affinities of both half-sites within the same complex are similar, which correlates with the fact that the wild-type and A138T enzymes both cleave the AAATTC site with symmetric rates in the two half-sites ($k_1/k_2 \cong 1$). In addition the asymmetric cleavage rates for the TAATTC site are reflected by differences in Mg^{2+} affinities in the two DNA half-sites (Figure 5.6C). The more rapidly cleaved cognate half-site binds more tightly to Mg^{2+} ($K_{D1\ App.\ Mg2+}$ = 3.0 ± 0.8 mM) than the miscognate half-site ($K_{D2\ App.\ Mg2+}$ = 7.1 ± 1.4 mM). The precision of data from similar experiments with the wild-type enzyme for cleavage of TAATTC are insufficient to draw conclusions correlating Mg2+ binding affinities to cleavage rate constants (data not shown).

Engler et al. (Engler et al., 2001) suggested that differences in active site assembly in *Bam*HI complexes with the different classes of sites are important for enzyme function *in vivo*. The repulsive cluster of negative charges at the active site of the specific complex destabilizes the ground state of the enzyme-complex, and produces a high affinity binding site for Mg^{2+}, which leads to progression to the transition state. Since the active sites of non-specific and miscognate complexes are not assembled, binding to these sites is not accompanied by the same 'electrostatic drive' towards metal binding and the transition state (Engler et al., 2001). Based on these ideas, we predict that the increased active site assembly in mutant-miscognate DNA complexes also produces an increased electrostatic drive towards the transition state for DNA cleavage. Furthermore, the differences in active site assembly (Mg^{2+} affinity) will have an

impact on DNA sequence discrimination *in vivo* based on the following: The free Mg^{2+} concentration in the *E. coli* cell (1-2 mM (Kuhn et al., 1985)) is too low for the wild-type or A138T-AAATTC complexes to be saturated with metal *in vivo*. (Note that the *in vivo* apparent Mg^{2+} affinities of miscognate DNA complexes will likely be lower than those reported in figure 5.6, since the salt concentration in which I made my measurements (0.1 M NaCl) is lower than the salt concentration in the *E. coli* cytoplasm (0.23 M KCl) (Richey et al., 1987); and the Mg^{2+} binding affinity is likely to be affected by salt concentration.) Thus in the cell, where Mg^{2+} concentration is limiting, a greater fraction of A138T-miscognate complexes are predicted to be bound to the metal than wild-type complexes, which in turn will lead to a greater probability of DNA cleavage by the mutant.

As a complementary approach to metal binding, I would like to test the pH dependencies of miscognate DNA binding by the wild-type and A138T enzymes. If the active sites of A138T-miscognate complexes are more assembled than the wild-type counterparts, I expect to observe steeper pH dependence associated with A138T miscognate DNA binding compared to wild-type.

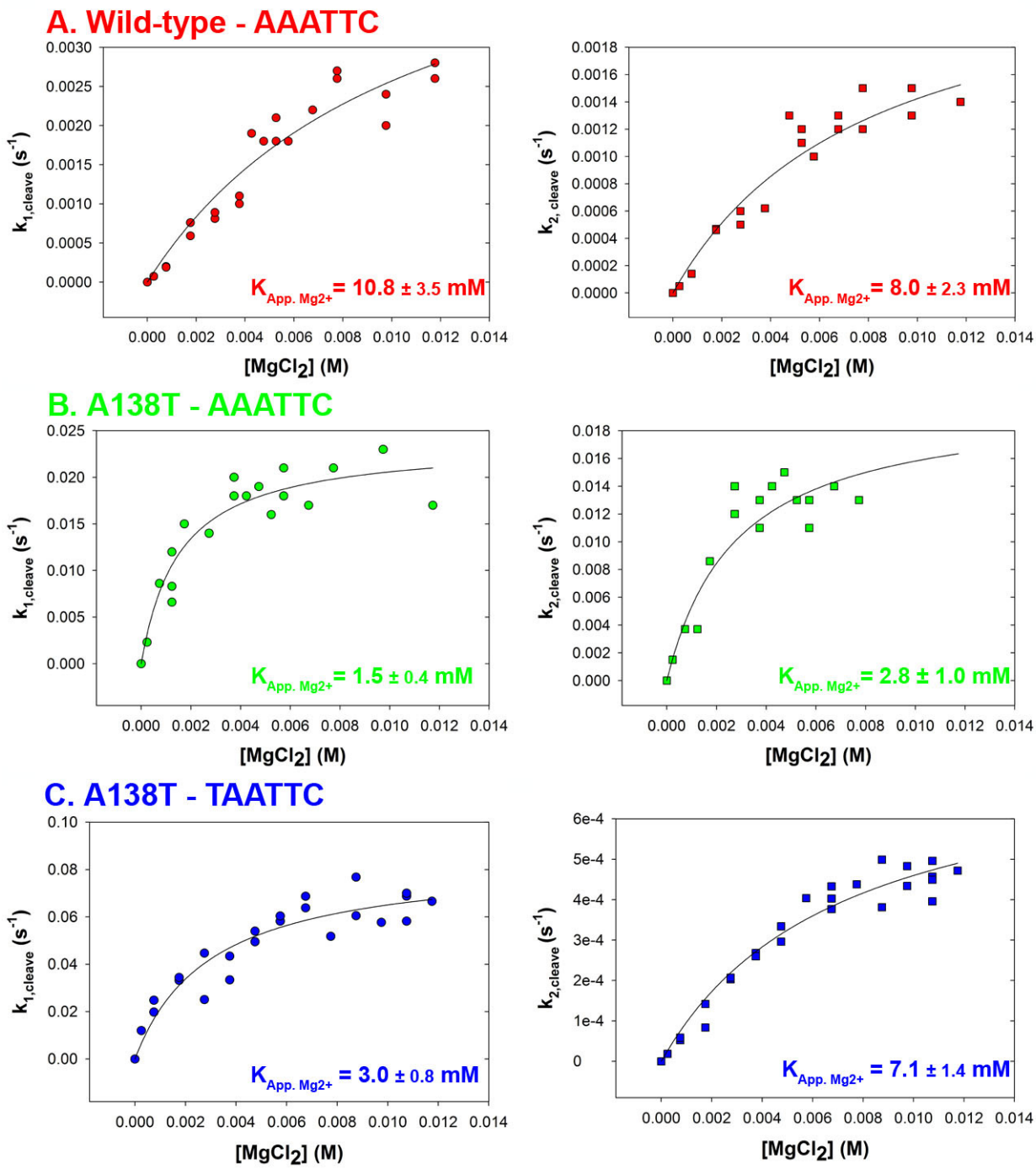


Figure 5.6 Mg2+ affinity of active sites in *Eco*RI-miscognate DNA complexes.

First order cleavage rate constants were measured in binding buffer plus 0.1 M NaCl (pH7.1, 21°C); reactions were initiated by adding varying amounts of $MgCl_2$ (concentration on the x-axis). Left and right panels show the dependence of cleavage of the cognate and non-cognate half-sites on [Mg^{2+}] respectively. Lines show non-linear least-squares regression fits to the equation: $k_{cleave} = k_{cleave\ max}\ [Mg^{2+}]/K_{A,\ App.\ Mg2+} + [Mg^{2+}]$. The binding constants recorded on the figure represent the equilibrium *dissociation* constants for Mg^{2+} binding.

5.3. Effect of flanking sequence on miscognate binding and cleavage reveals 'specific-like' properties of some A138T-miscognate complexes

I discussed the relevance of *Eco*RI miscognate DNA flanking sequence cleavage rate profiles to the *in vivo* specificity of the endonuclease in section. These data also give us insights into the molecular features of miscognate complexes, and how wild-type and A138T miscognate complexes differ with respect to these features. These results were already presented in section 4.2.4, but I will recapitulate the data here to aid in the clarity of the ensuing discussion. In contrast to specific binding (Figure 5.7), where flanking context dramatically affects binding affinity, with no effect on cleavage rates, miscognate DNA flanking context has a small effect on binding affinity, and a large effect on DNA cleavage rate (Figure 5.7 and Table 5.2). This is generally true of both enzymes, however, the A138T enzyme shows a much larger range in binding affinities (1.6 kcal/mol range) compared to that (0.4 kcal/mol range) of the wild-type enzyme (Table 5.2 and Figure 5.7). Conversely, when considering the AAATTC site in the five most rapidly cleavage flanking contexts, the A138T enzyme shows a much smaller range in cleavage rate constants (2.2 – fold) compared to wild-type enzyme (20 – fold). Therefore, the mutant profile for the interaction with *miscognate* DNA bears resemblance to the profile of the wild-type interaction with *specific* DNA. This is shown graphically in Figure 5.7B, where in A138T plots of $\Delta\Delta G^{\circ}_{bind}$ vs. $\Delta\Delta G^{\circ}_{I}{}^{\ddagger}$ (recall $\Delta\Delta G^{\circ}_{I}{}^{\ddagger} = \Delta\Delta G^{\circ}_{bind} + \Delta\Delta G^{\circ\ddagger}$), data for the AAATTC site in a subset of flanking contexts (CGC, GGT, GAT) are spread out in the $\Delta\Delta G^{\circ}_{bind}$ dimension, and the spread in the transition state interaction free energy dimension originates from the contribution from differences in binding free energy ($\Delta\Delta G^{\circ}_{bind}$) to the $\Delta\Delta G^{\circ}_{I}{}^{\ddagger}$ parameter, not differences in cleavage rate ($\Delta\Delta G^{\circ\ddagger}$). This indicates that the A138T enzyme is able to efficiently utilize binding free energy in the transition state for complexes with miscognate DNA sites in many different flanking contexts. This profile bears resemblance to the specific complex, where flanking sequence affects binding, but not the DNA cleavage rate (Figure 5.7). A plot of data for the wild-type interaction with miscognate DNAs in different flanking contexts (Figure 5.7A) illuminates differences with the A138T enzyme. In the wild-type plot, there are no differences in binding affinities (no spread of data in the $\Delta\Delta G^{\circ}_{bind}$ dimension), and differences in the transition state interaction free energy ($\Delta\Delta G^{\circ}_{I}{}^{\ddagger}$) are dominated by differences in cleavage rate ($\Delta\Delta G^{\circ\ddagger}$).

How can we interpret the differences between the wild-type and A138T miscognate DNA interaction profiles at the molecular level? Since A138T cleavage rates of the AAATTC site are unaffected by a subset of flanking contexts (GGT, TGT, CGC, CGT and GAT), it is useful to consider hypotheses to explain why *specific* DNA cleavage is unaffected by flanking sequence. Jen-Jacobson and colleagues (Jen-Jacobson, 1997, Jen-Jacobson, 2000) proposed that specific complexes with different flanking contexts differ with respect to the energetic cost of achieving the distorted DNA conformation, and the changes in enzyme and DNA dynamics that accompany differences in the optimization of protein-DNA contacts (e.g. H-bond distance and geometry) across the interface. That the GAATTC flanking context does not affect the global features of the bound states such as protein-base contacts, protein-phosphate contacts, or DNA conformation is supported by structural studies (Grigorescu, 2003). Furthermore, measurements in solution (Kurpiewski et al., 2004) suggest that the ability to achieve the distorted DNA conformation observed in the specific complex is critical to reaching the transition state. Given the importance of achieving this precise DNA geometry in reaching the transition state, the fact that the *Eco*RI enzyme cleaves the GAATTC site at equal rates regardless of context, indicates the DNA conformations in the bound states of the different complexes are the same. That is, specific complexes with different flanking sequences are *all* pre-transition state complexes poised to enter the transition state. Thus flanking sequence modulation of the differential costs for achieving the cleavage-competent DNA conformation in these pre-transition state complexes are reflected in the range of binding affinities; no addition costs must be paid in the transition state.

Interestingly, for both enzymes, the specific flanking sequence binding hierarchy and miscognate flanking sequence cleavage hierarchy are very similar (Table 5.2). This leads me to hypothesize that some of the same factors contributing to flanking sequence modulation of specific binding are operative in modulation of the probability of achieving productive cleavage of miscognate sites. Bearing this in mind, I propose that the A138T enzyme cleaves the AAATTC site in a large number of flanking contexts rapidly and with similar rates because the adaptations that produce wild type miscognate complexes that are not on the path to the transition state are not as severe and/or do not occur for as large a fraction of the A138T-AAATTC complexes.

As a consequence of the adaptations that minimize the free energy for wild type binding to miscognate sites, crucial phosphate clamp contacts (required for stabilizing the precise geometry of the cleavage-competent DNA conformation) are missing (see section 5.4 below). That is, wild type miscognate complexes are so far off the path to the transition state that flanking context [which influences the correct placement of the crucial phosphate clamps (see below)] has no effect on binding to miscognate sites. Instead, for the wild type enzyme, flanking sequence effects are observed as differential costs of distorting the AAATTC site in the transition state. By contrast, as noted above, adaptations for A138T binding to the AAATTC site embedded in a subset of flanking sequences (first five sequences in Table 5.2) are not as severe. (See ethylation interference footprinting section below). Thus these A138T complexes show a flanking context "signature" for binding and cleavage that more closely resembles that of the specific wild type complex. The adaptations in the mutant complexes with the remaining sites (AAATTC flanking contexts 6-9 in Table 5.2), share the characteristics of the wild-type miscognate complex in that the ground state complexes are significantly off of the path to the transition state and the differential energetic cost of distorting the AAATTC site is manifest in the range in cleavage rate constants.

Table 5.2 Effect of AAATTC flanking sequence on binding and cleavage

Site [a]	k_1 (s^{-1})	k_2 (s^{-1})	Relative position in *specific* flanking seq. binding hierarchy [b]	$\Delta\Delta G^o_{bind}$ (kcal/mol) [d]
Wild-type				
CGCaaattcGCG	$3.8 \pm 0.3 \times 10^{-3}$	$2.6 \pm 0.3 \times 10^{-3}$	1	6.6 ± 0.1
GGCaaattcGCC	$6.9 \pm 0.7 \times 10^{-4}$	$4.4 \pm 0.3 \times 10^{-4}$	2	6.9 ± 0.1
GGTaaattcACC	$3.0 \pm 0.1 \times 10^{-4}$	$3.6 \pm 0.2 \times 10^{-4}$	3	7.0 ± 0.1
CGTaaattcACG	$2.0 \pm 0.1 \times 10^{-4}$	$5.0 \pm 0.2 \times 10^{-4}$	ND [c]	ND
GCAaaattcTGC	$1.9 \pm 0.3 \times 10^{-4}$	$3.9 \pm 0.7 \times 10^{-5}$	5	6.9 ± 0.1
TGTaaattcATA	$1.6 \pm 0.3 \times 10^{-4}$	$2.8 \pm 0.9 \times 10^{-4}$	4	ND
GATaaattcATC	$1.1 \pm 0.1 \times 10^{-4}$	$1.0 \pm 0.1 \times 10^{-4}$	6	6.9 ± 0.1
CCCaaattcGGG	$2.4 \pm 0.2 \times 10^{-5}$	$1.5 \pm 0.3 \times 10^{-5}$	7	ND
AAAAaaattcTTTT	$1.3 \pm 0.4 \times 10^{-5}$	$7.0 \pm 0.7 \times 10^{-6}$	8	6.8 ± 0.1
A138T				
GGTaaattcACC	$4.2 \pm 0.7 \times 10^{-2}$	$3.4 \pm 1.0 \times 10^{-2}$	1	7.2 ± 0.2
TGTaaattcATA	$2.9 \pm 0.3 \times 10^{-2}$	$1.9 \pm 0.3 \times 10^{-2}$	ND	ND
CGCaaattcGCG	$2.1 \pm 0.5 \times 10^{-2}$	$2.1 \pm 0.4 \times 10^{-2}$	2	6.3 ± 0.1
CGTaaattcACG	$2.1 \pm 0.2 \times 10^{-2}$	$2.2 \pm 0.1 \times 10^{-2}$	ND	ND
GATaaattcATC	$1.9 \pm 0.3 \times 10^{-2}$	$8.9 \pm 0.2 \times 10^{-3}$	3	7.3 ± 0.1
GGCaaattcGCC	$3.9 \pm 0.1 \times 10^{-3}$	$3.7 \pm 0.2 \times 10^{-3}$	5	7.9 ± 0.2
GCAaaattcTGC	$1.2 \pm 0.1 \times 10^{-3}$	$6.5 \pm 0.7 \times 10^{-4}$	6	7.2 ± 0.1
CCCaaattcGGG	$1.1 \pm 0.1 \times 10^{-3}$	$9.2 \pm 0.1 \times 10^{-4}$	4	ND
AAAAaaattcTTTT	$1.2 \pm 0.3 \times 10^{-4}$	$8.8 \pm 0.2 \times 10^{-5}$	7	7.9 ± 0.1

[a] Each site is embedded in a double stranded 24 base-pair oligodeoxynucleotide: 5'-GGGCGGGyyyAAATTCzzzGGCGC, where the unlined site is listed in column 1. For each enzyme, sites are listed in the order of their cognate strand cleavage rates.

[b] Raw A138T specific binding data are in Table 2.1. Michael R. Kurpiewski's unpublished wild-type *Eco*RI binding data were used to determine this hierarchy since he examined more than 2/3 of the 64 possible flanking triplets.

[c] ND means that the measurement was not done for the particular sequence.

[d] $\Delta\Delta G^o_{bind}$ values were calculated using the equilibrium binding constant for the specific site (CGC flanking context) as the reference for each enzyme. Equilibrium association constants were measured in binding buffer plus 0.22M KCl (pH 7.3, 21°C).

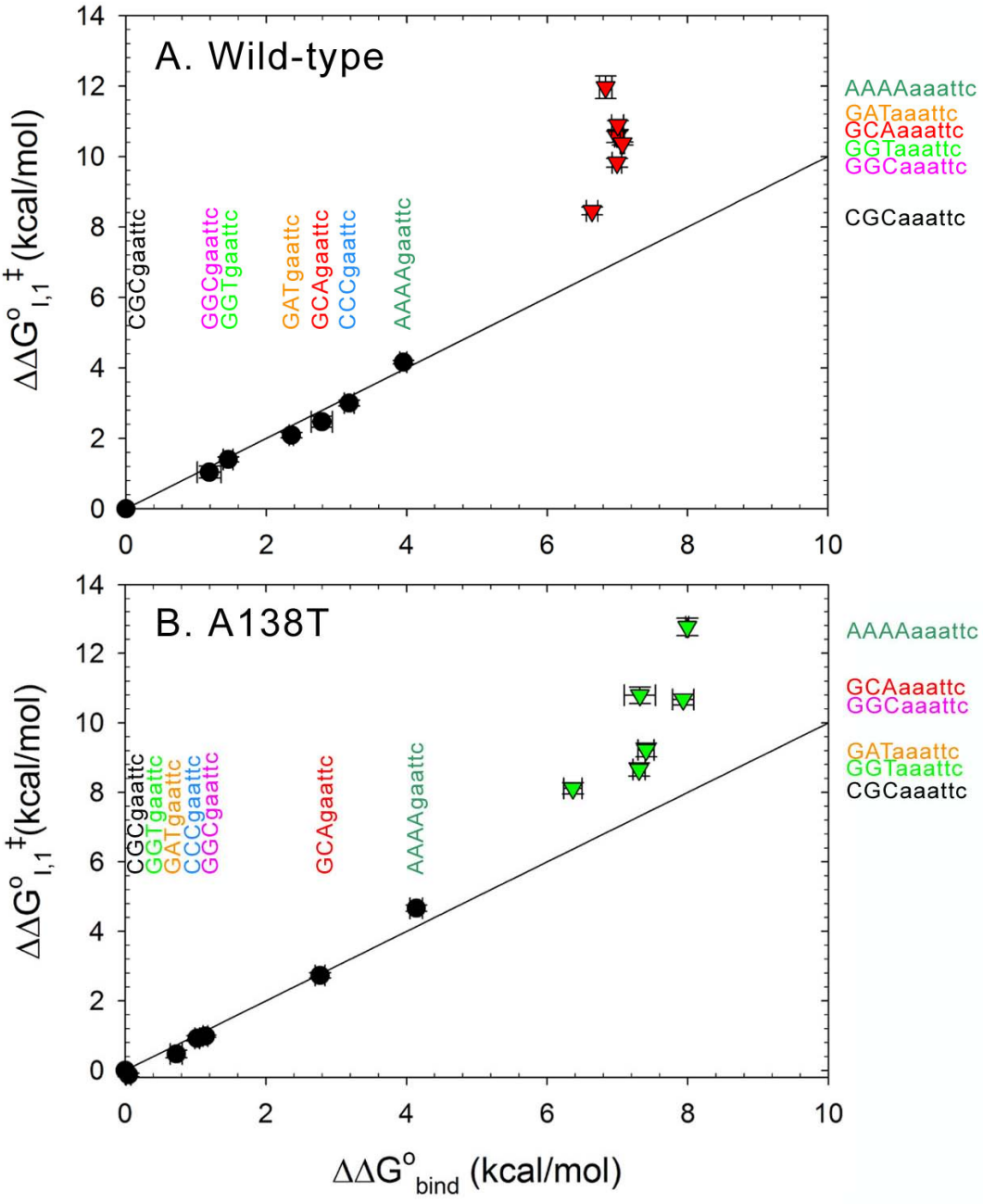


Figure 5.7 Effect of flanking sequence on the *Eco*RI specific and miscognate DNA interactions.

(A) Data for the wild-type *Eco*RI interaction with the specific site (●) and AAATTC miscognate site (▼) in different flanking contexts. (B) Data for the A138T *Eco*RI interaction with the specific site (●) and AAATTC miscognate site (▼) in different flanking contexts. In both panels, values of $\Delta\Delta G°_{bind}$ (x-axis), and $\Delta\Delta G°_{I,1}{}^{\ddagger}$ (y-axis, $\Delta\Delta G°_{I,1}{}^{\ddagger} = \Delta\Delta G°_{bind} + \Delta\Delta G°^{\ddagger}$ for the cognate strand bond breaking step) are referenced to the interaction with the specific site in the CGC flanking context (at 0,0). Data points for the specific sequences lie on the line of unit slope ($\Delta\Delta G°_{bind} = \Delta\Delta G°_{I,1}{}^{\ddagger}$), but data points for miscognate sites lie above the y = x line, indicating that $\Delta\Delta G°_{I,1}{}^{\ddagger} >> \Delta\Delta G°_{bind}$. Miscognate and specific sites in similar flanking contexts are colored the same. Note that the specific DNA binding and miscognate DNA cleavage flanking sequence hierarchies are similar.

5.4. Ethylation interference footprints reveal structural differences between wild-type and A138T-miscognate DNA complexes

Ethylation interference foot-printing (Siebenlist and Gilbert, 1980) is an experimental probe that is used to identify functionally important protein-phosphate contacts. Results from ethylation interference footprinting have identified 6 phosphate "clamps" (3 in each strand at N**p**N**p**GAA**p**TTC) that are critical to the formation of the *Eco*RI-specific DNA complex (Figure 5.8A) (Lesser et al., 1990). The functional importance of these phosphate contacts in binding is supported by the extensive networking of these phosphate contacts to key base recognition elements observed in the crystal structure of the *Eco*RI-specific DNA complex (Grigorescu, 2003). Since crystal structures of *Eco*RI-miscognate complexes are not yet available, we have relied on measurements in solution to gain insights into the molecular basis for sequence discrimination. Indeed, ethylation interference foot-printing of *Eco*RI-miscognate complexes (Lesser et al., 1990) provided valuable information concerning the role that phosphate contacts play in the discrimination mechanism of the *Eco*RI endonuclease; many of the phosphate clamp contacts observed in the specific complex are absent from the altered half-site in miscognate complexes (Figure 5.8B&C). Foot-prints on asymmetric miscognate sites are also asymmetric (Figure 5.8B&C); this provides a structural correlate to the asymmetric cleavage rates observed for miscognate sites. For example, foot prints of the wild type enzyme on miscognate sites typically show stronger interferences at phosphates in the canonical half site relative to interferences in the noncanonical half site, suggesting that the enzyme has structurally adapted towards the unaltered half site. These adaptations support the view that the structures of wild type miscognate complexes deviate significantly from the pre-transition state configuration of specific complexes; thus miscognate complexes are not on the direct path to the transition state.

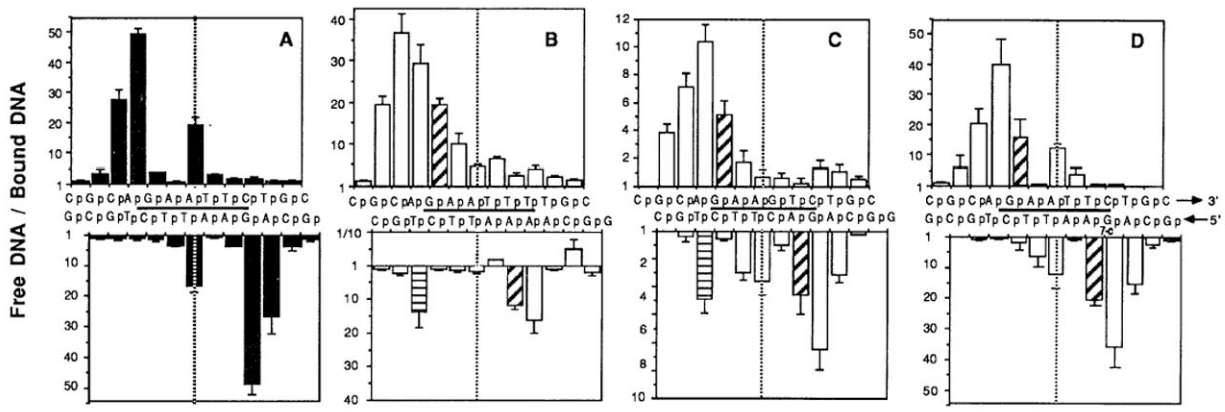


Figure 5.8 Contacts to DNA phosphates are important in discrimination against miscognate sites.

Ethylation interference footprints of the *Eco*RI endonuclease binding to the specific site (A), the miscognate sites AAATTC (B) and GACTTC (C), and the specific site with 7-deaza-guanosine substituted for **G**AATTC (D). In panels C-D, the bottom figure represents the footprint on the strand containing the modified half-site. Figure is taken from (Lesser et al., 1990).

The ability to achieve the distorted DNA conformation observed in the crystal structure appears to be critical for efficient DNA cleavage by the *Eco*RI endonuclease (Lesser et al., 1993, Kurpiewski et al., 2004). It has been proposed (Lesser et al., 1990) that the interactions between the enzyme and clamp phosphates play a crucial role in stabilizing the precise DNA conformation required for cleavage. In section 5.3, I hypothesized that the A138T enzyme is more successful than the wild-type at distorting miscognate a DNA site into a conformation that is cleavage competent. Since proper formation and positioning of the phosphate clamp contacts and DNA distortion are linked phenomena, we would predict that these crucial contacts are stronger (i.e. have more optimized H-bonding geometry) in mutant than wild type miscognate complexes and/or interactions which are missing between the enzyme and clamp phosphates in wild-type miscognate complexes are restored in mutant miscognate complexes. To test this prediction, we compared the ethylation interference foot-prints of wild-type and A138T *Eco*RI endonucleases on the AAATTC miscognate DNA site.

The footprint I obtained for wild type binding (in the absence of Ca^{2+}) to the miscognate AAATTC site embedded in a 22 bp oligomer (CGC flanking context) is much weaker than that reported by Lesser et al (Lesser et al., 1990) for this site embedded in 17 bp oligomer (GCA flanking context). The weaker footprint for the longer miscognate DNA substrate likely derives

from increased partitioning of the enzyme into the nonspecific binding mode since more nonspecific sites exist within this oligomer than in the 17 bp oligomer. This can be understood from the following equation:

$$K_{obs.} = K_{int} + K_{NS} * N_{NS} \qquad (5.1)$$

where $K_{int.}$ is the intrinsic miscognate DNA binding constant, K_{NS} the non-specific DNA binding constant, and N_{NS} the number of non-specific sites. This equation states that non-specific binding will make a significant contribution to the observed binding constant if the magnitudes of the intrinsic and non-specific binding constants are similar and/or the number of non-specific sites is large. Since I have shown that the *Eco*RI endonuclease binding affinities are similar for non-specific and miscognate DNAs (Table 4.1), the contribution of non-specific binding to the observed miscognate DNA binding affinity will be significant if the DNA oligomer is longer than the miscognate binding site size. Indeed, the number of non-specific sites in the 22 base-pair oligomer (22 non-specific sites: based on the crystallographically determined site size of 12 base-pairs) is much greater than the number in the 17 base-pair oligomer (12 non-specific sites), therefore the enzyme will be more likely to partition into the non-specific binding mode when the larger substrate is used. This would manifest itself as a weaker foot-print since the *Eco*RI endonuclease does not make localized foot prints on non-specific DNA (Lesser et al., 1990).

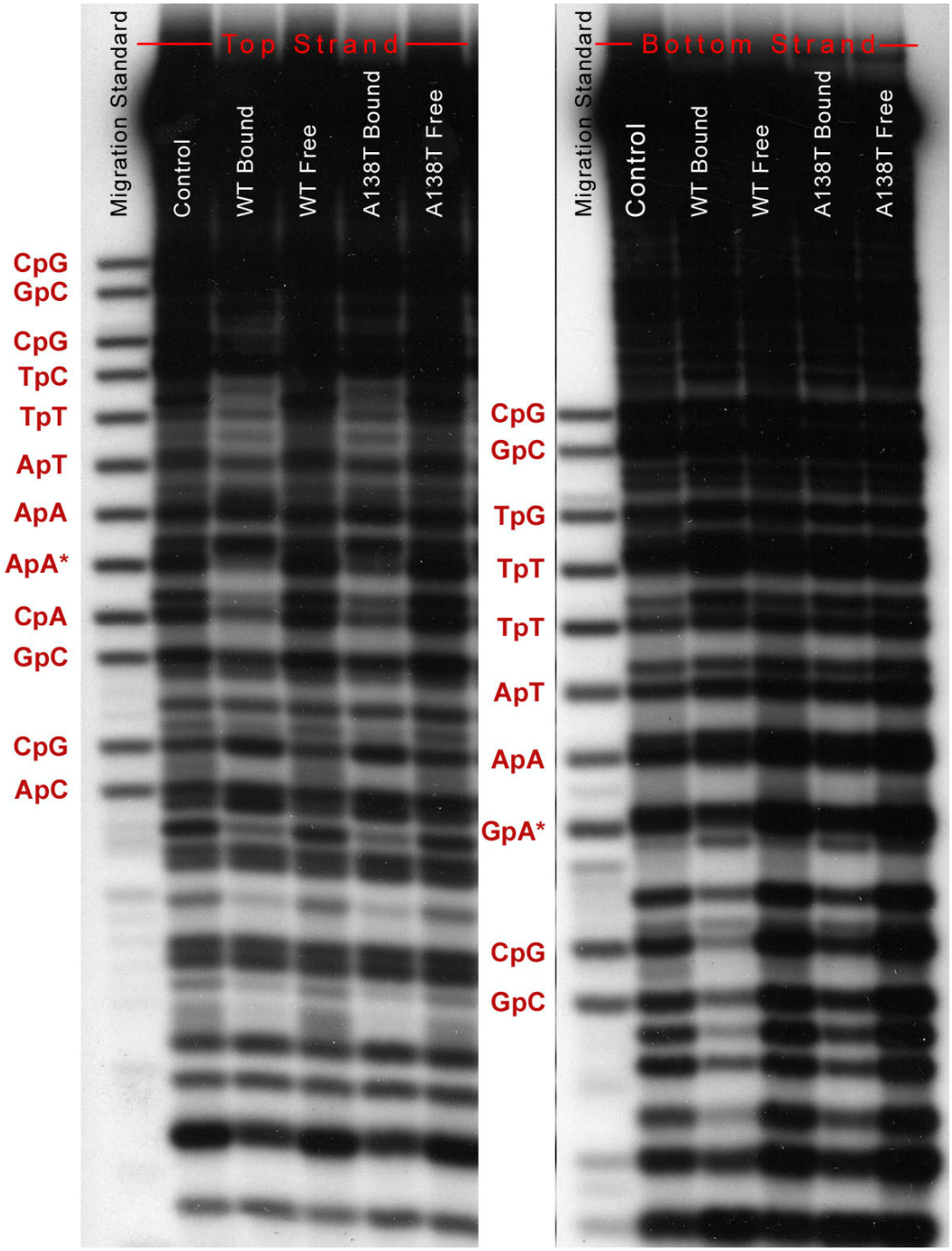


Figure 5.9 Ethylation interference footprints made by the *Eco*RI endonuclease on the cgcAAATTCgcg site in the absence of Ca^{2+}.

Complete sequence of the top strand of the 22 base oligonucleotide is given in Appendix A. Complexes were formed in binding buffer (Appendix B) plus 40 mM NaCl. Size standards are used to identify the positions of oligos terminating in various 3'-OH groups. See section 6.11 for details on the methods.

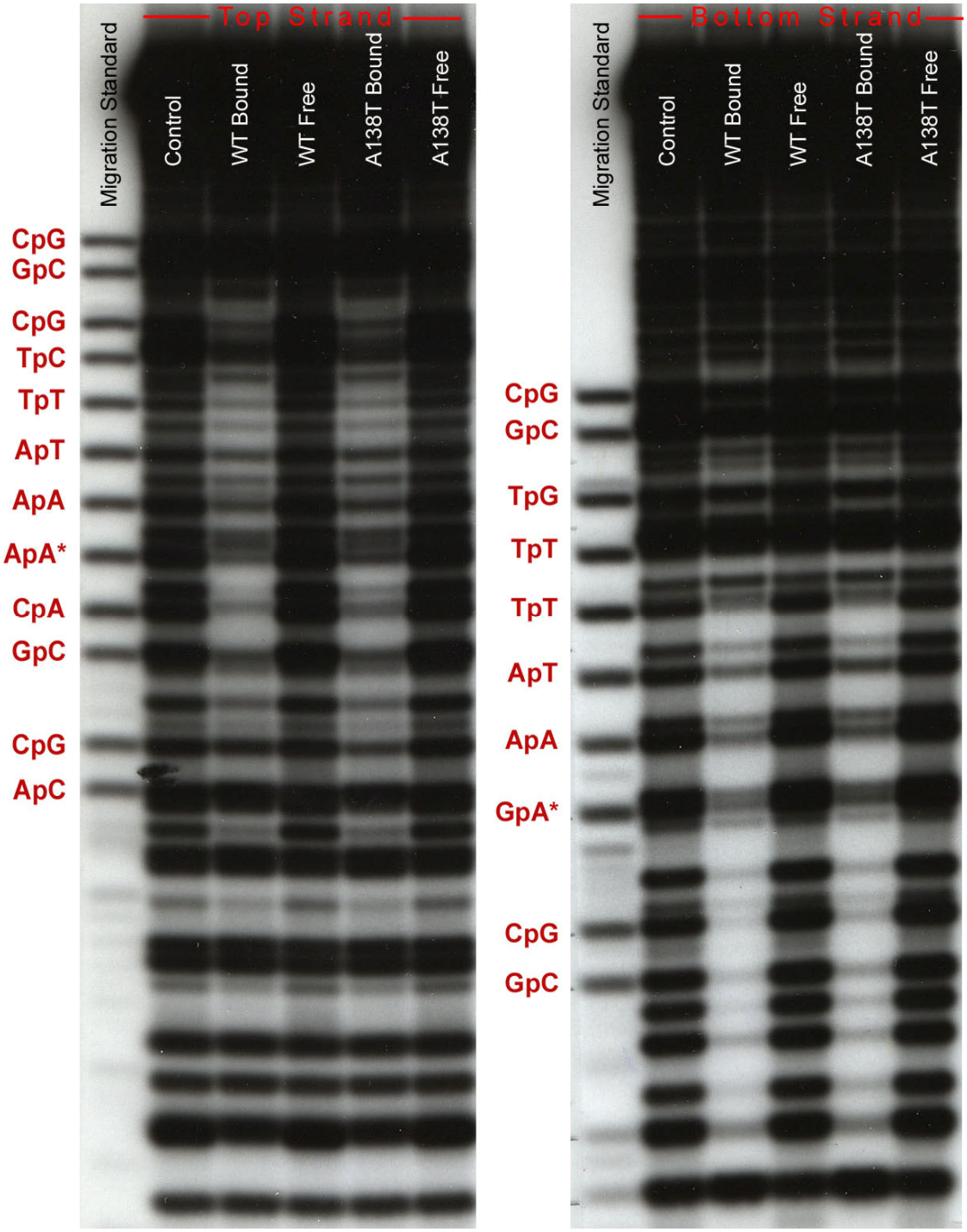


Figure 5.10 Ethylation interference footprints made by the *Eco*RI endonuclease on the cgcAAATTCgcg site in the presence of Ca^{2+}.

Complete sequence of the top strand of the 22 base oligonucleotide is given in Appendix A.. Complexes were formed in binding buffer plus 37 mM NaCl, 3mM $CaCl_2$. Size standards are used to identify the positions of oligos terminating in various 3'-OH groups. See section 6.11 for details on the methods. This footprint is stronger than the one shown in figure 5.9, which was measured in the absence of Ca^{2+}.

To circumvent this issue, I obtained footprints for miscognate complexes formed in the presence of Ca^{2+}. Ca^{2+} is expected to increase partitioning of the enzyme towards the miscognate DNA binding mode since the divalent metal improves miscognate binding, without commensurate improvement in non-specific binding (Table 5.1). Indeed, both enzymes make strong, localized foot-prints on the AAATTC site (22 base-pair oligonucleotide) when the complexes are formed in the presence of Ca^{2+} (compare figures 5.9 and Figure 5.10).

The footprints on the CGCaaattcGCG site made by the both the wild-type and A138T enzymes are the same within experimental error (Figures 5.11). There are three explanations, which are not mutually exclusive, for the observation that the two enzymes, which cleave this DNA site at different rates, have similar footprints. The first is that the A138T mutation enhances the cleavage rate of this sequence by some factor independent of changes in protein-DNA phosphate contacts. The second is that ethylation interference scores an average of the phosphate contacts over the entire ensemble of complexes, and small differences in the ensemble may not be detected by this experimental probe. The miscognate DNA binding ensemble is likely composed of non-specific complexes, other forms that may be incapable of reaching the transition state for DNA cleavage, as well as cleavage competent miscognate complexes. We hypothesize that A138T mutation causes a small shift in this ensemble, increasing the population of complexes that are on the path to the transition state. Despite the modest increases in the A138T cleavage rates for this site (5 and 8-fold increases relative to wild-type for cleavage of the cognate and non-cognate half-sites respectively), the mutant still cleaves this miscognate site 15-20-fold slower than the GAATTC site. This supports the idea that the ensemble of A138T complexes is still significantly populated by conformers that are off the path to the transition state. Therefore, although the A138T mutation may cause a shift in the miscognate DNA binding ensemble towards cleavage competent complexes, these complexes, as a fraction of the total would remain small, and any differences with the wild-type enzyme would be difficult to detect by measuring ensemble averaged properties. Lastly, we find it likely that the degree to which a miscognate DNA binding ensemble is populated by cleavage competent forms will be correlated with the profile of phosphate contacts to a given site. A qualitative measure of the partitioning of a population of miscognate complexes between productive and non-productive binding modes can be obtained by comparing the cleavage rate constant for a given miscognate site with the specific DNA cleavage rate constant, since the binding free energy in the specific

complex is efficiently utilized along the path to the transition state. The greater the decrease in miscognate cleavage rate relative to the specific DNA cleavage rate, the greater the shift of the miscognate binding ensemble into complexes that are off the path to the transition state. Studies on the effect of flanking sequence on AAATTC cleavage showed that the CGC flanking context is exceptional in its ability to promote AAATTC cleavage; this site is cleaved about 70-fold more slowly than the specific site. However, the site in the next to most favored context (GGC) is cleaved six-fold more slowly than CGCaaattcGCG, and 390-fold more slowly than the specific site. It follows that the ensemble of wild-type complexes with CGCaaattcGCG has a much higher population of forms which are on the path to the transition state than in complexes with any other AAATTC site. We are therefore least likely to observe differences between wild-type and A138T footprints on this site.

Based on this hypothesis, for the GGTaaattcACC site, which the wild-type enzyme cleaves 900-fold more slowly than GAATTC, we expect the wild-type-GGTaaattcACC binding ensemble to partition significantly into a non-productive binding mode, which would manifest itself in patterns of phosphate contacts that drastically deviate from the pattern observed on the specific site. By contrast, the A138T mutant cleaves GGTaaattcACC only 10-fold more slowly than GAATTC. Thus we predict that the A138T-GGTaaattcACC binding ensemble is significantly populated by productive complexes which in turn would lead to 'specific-like' character to the phosphate contacts in this complex.

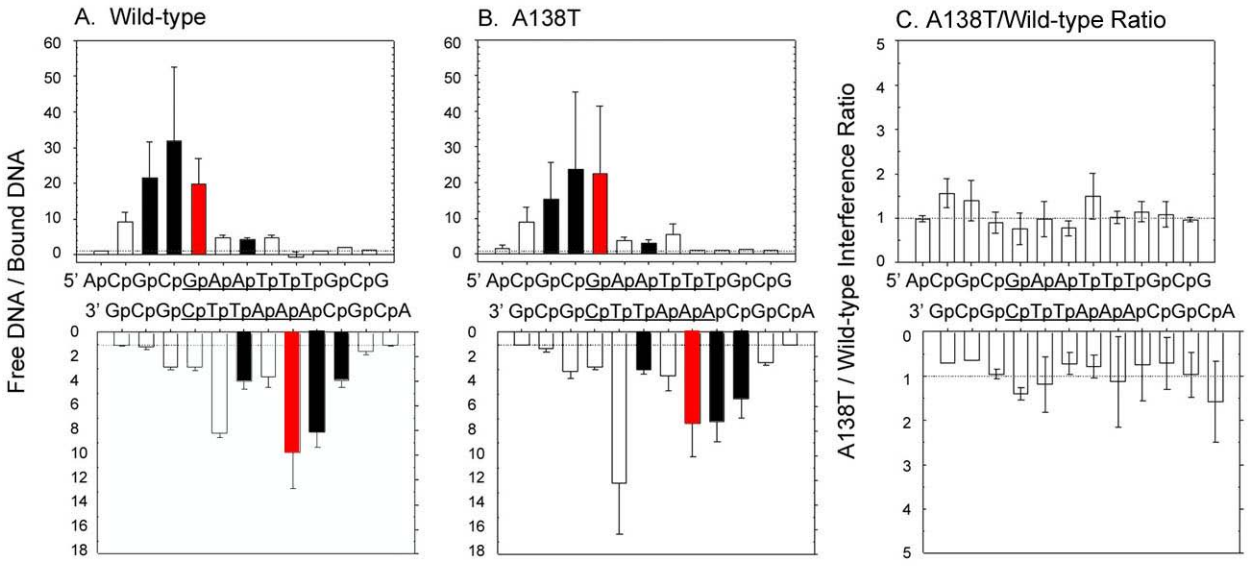


Figure 5.11 Ethylation interference footprints on the cgcAAATTCgcg site embedded in a 22 base-pair double stranded deoxyoligonucleotide.

Enzyme-DNA complexes were equilibrated in binding buffer plus 37mM, 3mM $CaCl_2$ (pH 7.3, 21ºC). See materials and methods (section 6.11) for experimental details. In panels A and B, a y-axis value > 1 indicates that an ethyl group at a given position interferes with endonuclease binding; a value <1 indicates an ethyl group enhances binding. In panel C, the ratio of the interference values for the A138T enzyme over the value for the wild-type enzyme is plotted. In panel C, a y-axis value > 1 indicates a stronger interference in the A138T complex at a given position. The top and bottom panels give the interference patterns on the strands containing the cognate and non-cognate half-sites respectively. Interference bars at the scissile phosphate are colored red, and bars at the six phosphate 'clamps' (see text) are colored black.

Indeed, I observe differences between the two enzymes with respect to their footprints on the GGTaaattcACC sequence (Figure 5.12). The most striking differences reside in the footprints on the non-canonical strand of GGTaaattcACC, where A138T exhibits 3.5 and 2-fold increases in interference relative to wild-type at the GG**p**TaaattcACC and GGT**p**aaattcACC phosphate clamp positions respectively (Figure 5.12C). In addition, ethylation at G**p**GTaaattcACC *enhances* wild-type binding, while ethylation at this same phosphate interferes with mutant binding (Figure 5.12A&B). Inspection of the autoradiogram containing the footprints (Figure 5.13) clearly shows that strands ethylated at G**p**GTaaattcACC are over-represented in the wild-type bound fraction, and underrepresented in the A138T bound fraction. This finding indicates that the wild-type complex is looser than the A138T complex in that a bulky ethyl group at G**p**GTaaattc is accommodated at the wild-type interface, yet the similar

perturbation inhibits formation of the mutant complex. Interestingly, wild-type footprints on *specific* DNA show small but reproducible interference at the analogous G**p**CAgaattcTGC phosphate position (Figure 5.8A). Finally, A138T shows a 2-fold increase in interference at the GGTaaa**p**ttcACC clamp phosphates in both strands (Figure 5.12C). The footprints on the canonical DNA site entail large interference at GAA**p**TTC clamp phosphate. The interference is likely the result of a constrained hydrogen bond between the Gly116 main-chain and a non-bridging phosphoryl oxygen. (The contact between Gly116 and the phosphate backbone anchors active site residues Glu111 and Lys113, and thus is likely important in catalysis.) This contact is severely attenuated in the wild-type GGTaaattcACC complex, in accord with the large penalty in cleavage rate for this sequence (Table 5.1). Taken together, these observations indicate a shift in the A138T-GGTaaattc binding ensemble, favoring forms where the three phosphate clamps are preferentially restored in the non-canonical half-site. Thus, unlike the wild-type-GGTaaattcACC complex in which the enzymes adapts away from the altered half-site towards the unaltered half-site, the symmetry of characteristic of the specific complex is partially restored in the mutant miscognate complex.

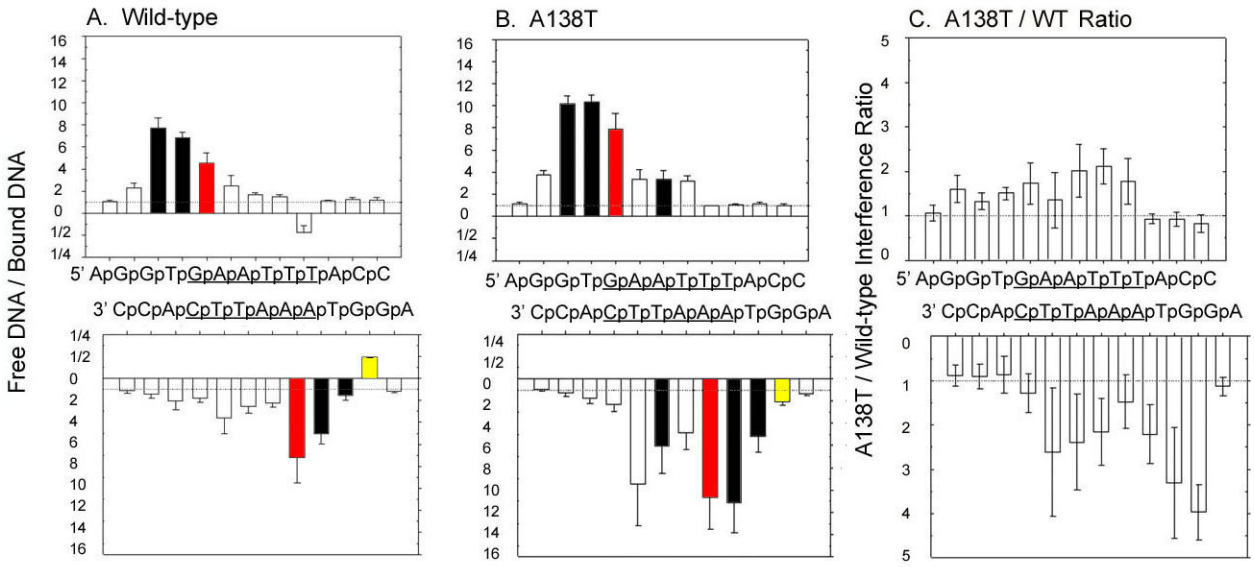


Figure 5.12 Ethylation interference footprints on the ggtAAATTCacc site embedded in a 22 base-pair double stranded deoxyoligonucleotide.

Enzyme-DNA complexes were equilibrated in binding buffer plus 37mM, 3mM $CaCl_2$ (pH 7.3, 21ºC). See materials and methods (section 6.11) for experimental details. In panels A and B, a y-axis value > 1 indicates that an ethyl group at a given position interferes with endonuclease binding; a value <1 indicates an ethyl group enhances binding. In panel C, the ratio of the interference values for the A138T enzyme over the value for the wild-type enzyme is plotted. In panel C, a y-axis value > 1 indicates a stronger interference in the A138T complex at a given position. The top and bottom panels give the interference patterns on the strands containing the cognate and non-cognate half-sites respectively. Interference bars at the scissile phosphate are colored red, and bars at the six phosphate 'clamps' (see text) are colored black. The bar colored yellow represents a phosphate where ethylation improves wild-type binding, yet interferes with mutant binding.

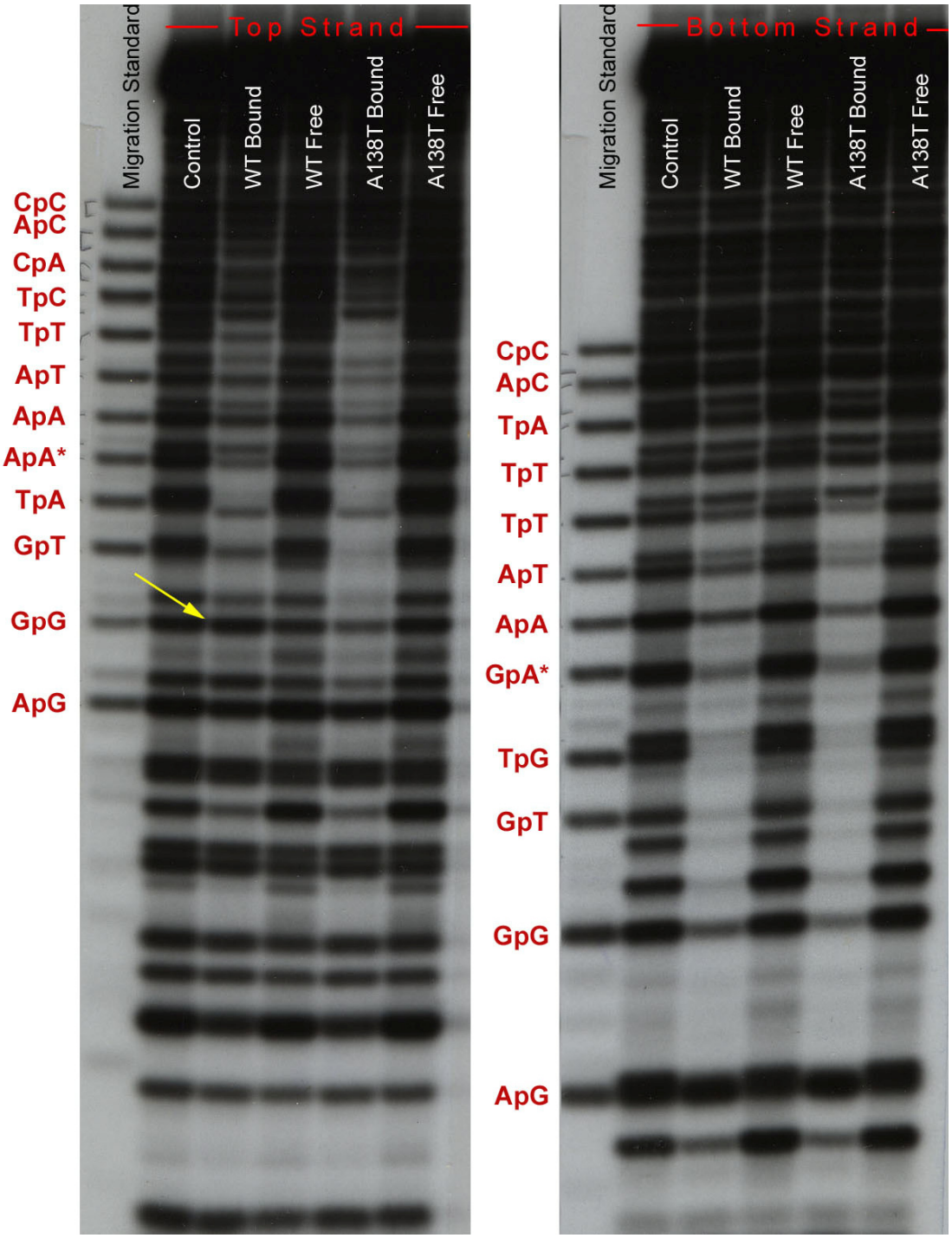


Figure 5.13 Ethylation interference footprints made by the *Eco*RI endonuclease on the ggtAAATTCacc site in the presence of Ca^{2+}.

Complete sequence of the top strand of the 22 base-pair oligonucleotide is given in Appendix A. Complexes were formed in binding buffer plus 37 mM NaCl, 3mM $CaCl_2$. Size standards are used to identify the positions of oligos terminating in various 3'-OH groups. See section 6.11 for details on the methods. The yellow arrow points to the phosphate where ethylation improves wild-type binding, yet inhibits A138T binding (see text and figure 5.12A&B).

By comparing the *wild-type* footprints on the AAATTC site in three different flanking contexts (CGC, GCA and GGT), we observe correspondence between the cleavage rates and the profile of phosphate contacts to miscognate DNAs. The footprints on the canonical half-sites of the three AAATTC sites (Figure 5.14B-D) are similar in that relative to footprints on the specific site (Figure 5.14A), the central, AAA**p**TTC phosphate clamp is absent, the NN**p**N**p**AAATTC clamps are maintained, and interference is observed at the scissile phosphate (where very little interference in specific DNA footprints). However, the footprints on the modified half-site of AAATTC in the CGC flanking context are distinct from footprints on the same site embedded in the GCA and GGT flanking triplets. The footprints on the miscognate half-site of CGCaaattcGCG show interferences at the N**p**N**p**NAAATTC phosphates; footprints on the specific DNA site exhibit interferences at analogous positions. By contrast, footprints on the modified half-site of AAATTC in the GCA and GGT flanking context show a lack of interference at the NN**p**NAAATTC clamp phosphate, and an improvement in DNA binding that is associated with ethylation of the N**p**NNAAATTC phosphate (Figure 5.14C&D). The differences in footprints on these three sites show that the wild-type complexes with the CGCaaattcGCG site is the most 'specific-like' with respect to phosphate contacts, and complexes with GCAaaattcTGC and GGTaaattcACC site are more adaptive, and further off the path to the transition state. Indeed, the differences in footprints correlate with the observation that the wild-type enzyme cleaves the AAATTC site in the CGC flanking context 13 and 20-fold faster than the same site in the GGT and GCA flanking contexts respectively. That is the complex that retains the most "specific-like" ethylation interference pattern is more likely to reach the transition state. We can therefore ask the question: Is CGC an exceptional flanking context in that it can facilitate formation of the most 'specific-like' phosphate contacts compared to complexes with the AAATTC in other flanking contexts? The fact that the CGC flanking triplet is at the top of the specific DNA binding, and miscognate DNA cleavage flanking sequence hierarchies (Figure 5.7) might suggest that the anisotropic distortability of the CGC triplet is especially suited to facilitate *Eco*RI-DNA kink formation in the specific complex, or a similar DNA conformation in the miscognate complex to be stabilized by *Eco*RI endonuclease binding. We need to determine the footprints on AAATTC in a greater number of flanking contexts to determine if CGC is indeed an exception, if there is a continuum of footprint strength

and character that correlates with DNA cleavage rate, or if some other unforeseeable pattern emerges.

In contrast to the wild-type, the A138T enzymes' footprints on the AAATTC site in two different flanking contexts are similar to one another (Figure 5.14E&F). The mutant enzyme cleaves both of these sites with approximately equal rates ($k_{1,GGT}/k_{1,CGC} = 2$) (Table 5.2). In the previous section, I proposed that the A138T enzyme cleaves the A138T site rapidly in many flanking contexts (GGT, TGT, CGC, CGT and GAT-see Table 5.2), because the enzyme is able to stabilize DNA conformations in these sites that allows for progression to the transition state. Based on the mutant footprints, I extend the above hypothesis by suggesting that stabilization of the distorted DNA conformation is achieved by the formation of 'specific-like' phosphate contacts in the ground state, leading to a high probability of reaching the transition state. This is similar to the wild-type pre-transition state complex in which DNA distortion and contacts to the phosphate backbone are inextricably linked (see section 1.3.2.2), and the ability to achieve the *Eco*RI DNA kink is likely affected by both sequence dependent conformational propensities, as well as the ability of the enzyme to stabilize the distorted conformation (The latter is accomplished in large part by non-covalent interactions between the *Eco*RI endonuclease and the phosphate backbone of DNA.)

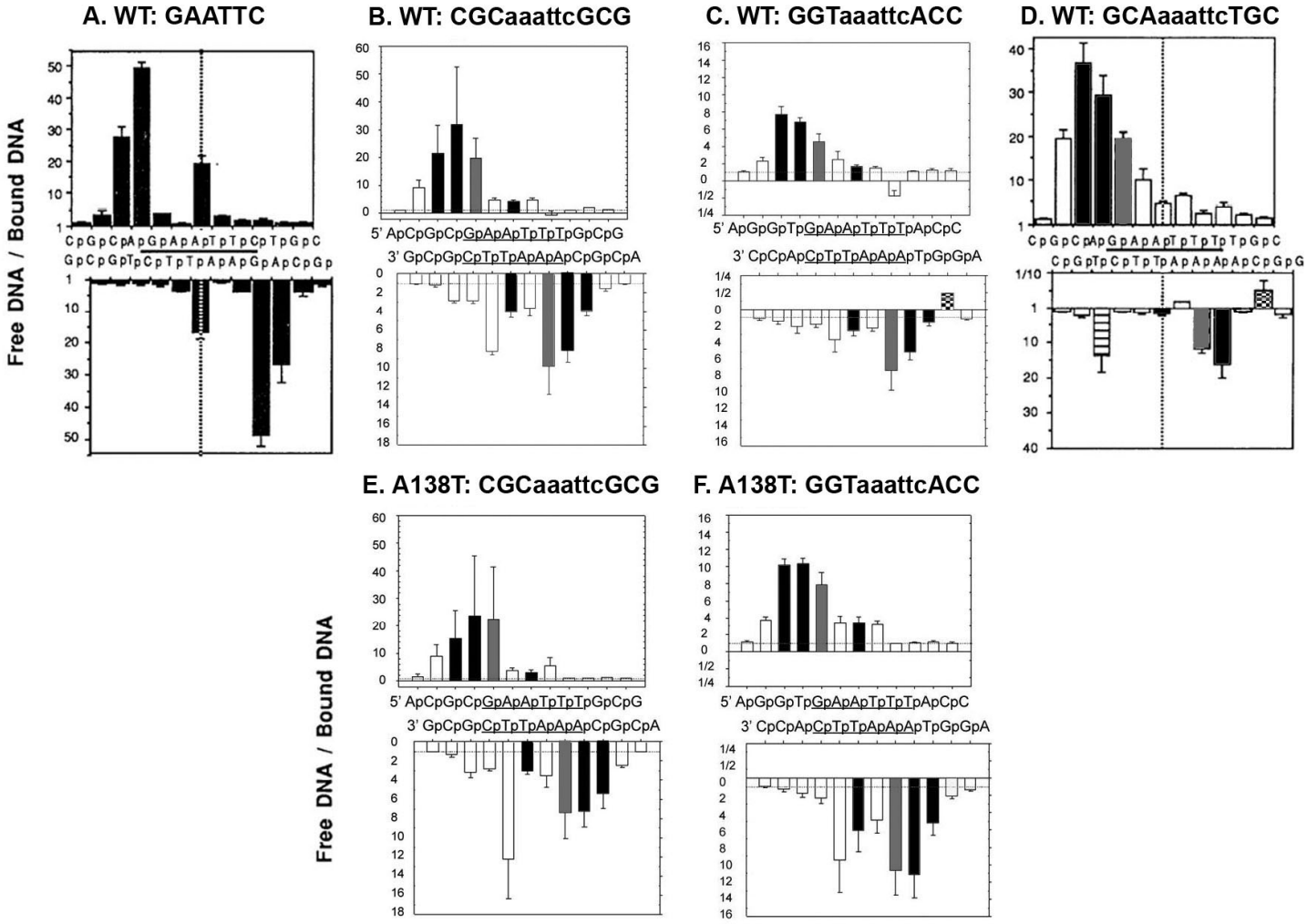


Figure 5.14 Effect of flanking sequence on AAATTC foot prints.

Top row: Ethylation interference footprints of the wild-type enzyme of the GAATTC site (A), and the AAATTC site in the CGC (B), GGT (C), and GCA (D) flanking contexts are shown. Footprints in panels A and D are taken from (Lesser et al., 1990).
Bottom row: Ethylation interference footprints of the A138T enzyme on the AAATTC site in the CGC (E) and GGT (F) flanking contexts.

What atomic factors are responsible for the differences in miscognate DNA footprints made by the two enzymes? In the absence of high resolution structural information on *Eco*RI-miscognate DNA complexes, we can use the coordinates of the specific complex as a guide, and operate under the assumption that the protein-phosphate contacts responsible for the specific footprint are responsible for interferences at analogous phosphates in miscognate DNA footprints. In the specific complex, the N**p**NNgaattc phosphate is contacted by the amide nitrogen of Gly196, which lies at the base of the outer segment of the enfolding arm. Furthermore, Arg203, which contacts the NN**p**Ngaattc clamp phosphate, is also located at the base of the arm. Therefore, an intriguing hypothesis to explain the increased interference at the G**p**GTaaattc (ethylation of this phosphate improves wild-type binding) and GG**p**Taaattc phosphates in the A138T footprint (Figure 5.12) is that the A138T mutation permits the outer arm to adapt or adjust in the miscognate complex such that contacts to these phosphates are restored. The location of the Thr138 residue in the specific structure is consistent with the A138T amino acid change having an effect on the conformation of the outer arm element (contains Gly196 which contacts to N**p**NNgaattc), as Gly196 is networked to Thr138 via main-chain hydrogen bonds between Ala139 and Ile197 (Figure 5.15). Thr138 is also indirectly "wired" to Arg203 (contact to NN**p**Ngaattc) via a solvent mediated network involving Ala139 and Arg200 (Figure 5.15). Interestingly, the I197A mutation is an intragenic suppressor of the A138T *Eco*RI relaxed specificity phenotype (Ivanenko et al., 1998). This is consistent with changes in the conformation or dynamics of the outer arm playing a role in the mechanism of A138T relaxed specificity.

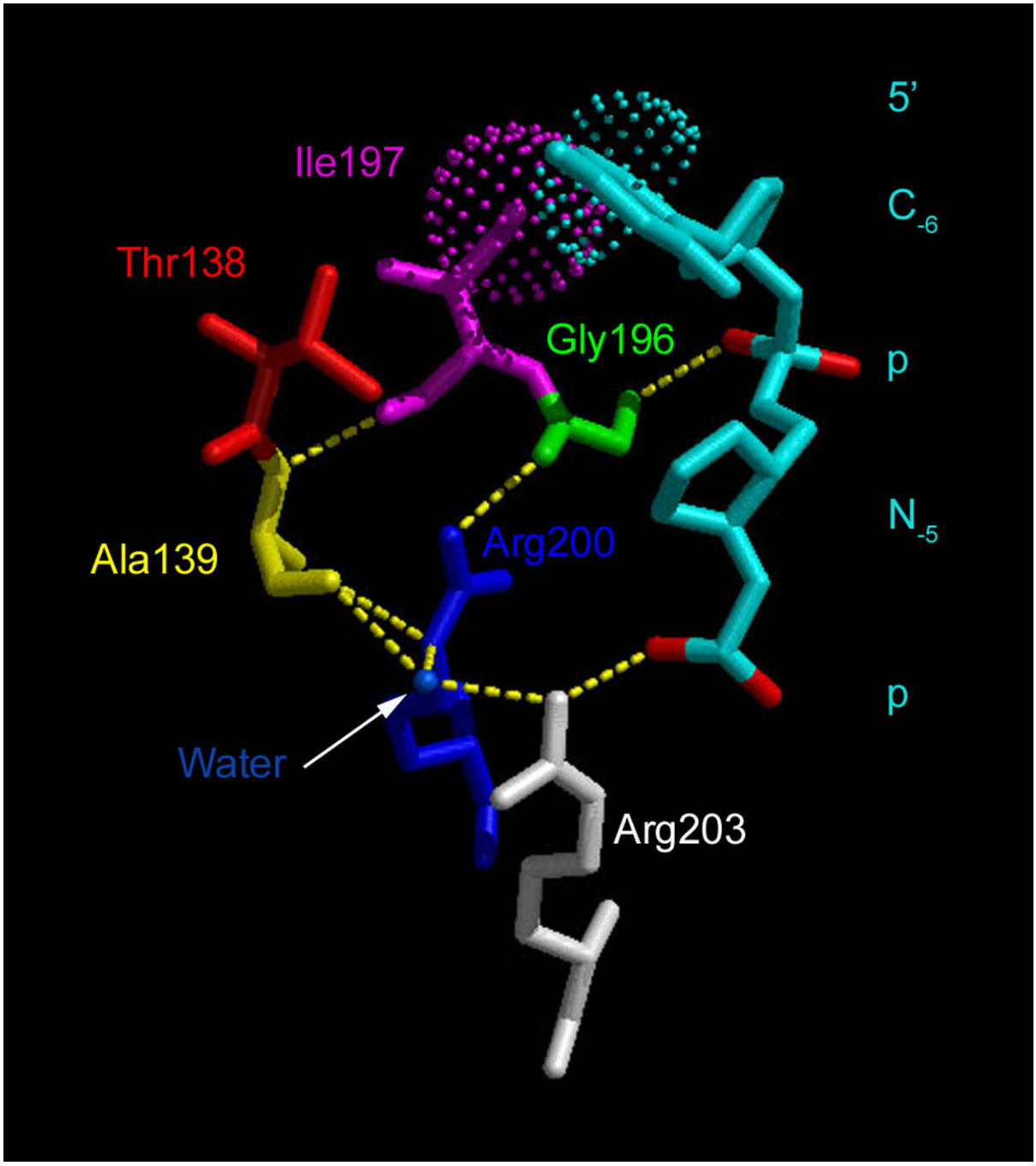


Figure 5.15 Network involving Thr138 and residues in the outer arm segment which contact the phosphate backbone.

Model coordinates are from the x-ray crystal structure of the A138T-d(TCGCGAATTCGCG) complex (presented in Chapter 3 of this work). Ethylation interference data show that the A138T mutation affects phosphate contacts to N_{-6}**p**N_{-5}**p**NAAATTC; Thr138 is networked to residues which contact the analogous phosphates in the specific complex. Hydrogen bonds are depicted as yellow dashed lines and van der walls interactions as dotted spheres.

5.5. Probes to address differences in conformation and/or dynamics of *Eco*RI arms in different complexes

Upon determining that the arms of the *Eco*RI endonuclease undergo a disorder to order transition coupled to specific DNA binding (Grigorescu, 2003), Dr. Arabela Grigorescu hypothesized that the conformational change is important to the mechanism of sequence discrimination. One of the reasons for this hypothesis is that each of promiscuous mutations involves an amino acid change within the arms, or at the junction between the arms and the main domain of the enzyme. In my characterization of the mutants, I have made several observations that are consistent with the idea that the promiscuous mutations affect the structure and or dynamics of the arms either in the free or the bound forms of the enzyme. First, the H114Y mutation affects the rate limiting step of specific DNA association and dissociation (section 2.2.5.2). Since the folding of the arms is the only known conformational change coupled to specific DNA binding, it is likely that the H114Y amino acid change perturbs this structural transition. Secondly, I find that the crystallographic B factors for localized regions of the arms are lower in the A138T-specific DNA complex than they are in the wild-type complex (section 3.2.5, Figure 3.11), consistent with an effect on the dynamics of the arms resulting from the A138T mutation. Finally, in the previous section, I demonstrated that the patterns of phosphate contacts are different in wild-type and A138T-misocgnate, and that these differences likely involve contacts made by amino acids within the outer arm segment. It would be of great interest to apply an experimental tool that can address questions regarding how the structure and dynamics of the arms differ between the free *Eco*RI endonuclease and the three classes of complexes. We would then like to ask if, how, and where the structure and dynamics of the arms are perturbed by the promiscuous mutations.

Nuclear magnetic resonance (NMR) spectroscopy would be the method of choice to gain insight into structural and dynamic characteristics of the *Eco*RI system. Unfortunately, the *Eco*RI endonuclease will precipitate under the solution conditions required to perform NMR experiments (Modrich et al., 1976). We must therefore consider other tools to test the hypothesis.

Fluorescence studies by Watrob et. al showed differences in the decay and time resolved emission anisotropy of Trp246 fluorescence between the free and specifically bound EcoRI

endonucleases (Watrob et al., 2001). In summary these data show differences in the arrangement of fluorescence quenching groups around Trp246 in the free and bound enzyme, and a reduction in the motion of the flourophore in the protein-DNA complex. Interestingly, His114 and Glu170, the groups likely to quench the fluorescence of Trp246, lie at the junctions between the main domain and the inner and outer segments of the arms respectively. This observation suggests that Trp246 fluorescence can be used to report on the conformational transition of the arms.

We propose using time resolved fluorescence resonance energy transfer (FRET) to ask how the distribution of distances between the Trp246 fluorescence donor at the base of the arm and an IAEDANS fluorescence acceptor at the tip of the arm changes between the free and bound enzymes. Further we will ask if the conformational distributions are different between the wild-type and relaxed specificity mutein(s) in the free and or bound enzymes. Although crystallographic studies have shown that the arms are folded in the specific complex, crystallographic B-factors (Grigorescu, 2003) and MD simulations (Lisa E. Engler and Linda Jen-Jacobson, *unpublished*) indicate that the arms maintain significant mobility in the bound protein. Thus we expect to see a distribution of donor to acceptor distances even in the complex. We can use this approach to ask if the promiscuous mutations affect the distribution of donor-acceptor distances in the free enzymes, and the specific, miscognate, and non-specific complexes.

To ask these questions, we will be required to engineer enzymes that have the W104Y mutation as Trp104 fluorescence dominates the emission spectrum of the wild-type enzyme (Watrob et al., 2001). In addition, we need to locate amino acids within the arm that are suitable for cysteine substitution. (A solvent accessible cysteine is required for efficient attachment of the IAEDANS acceptor group). There are no exposed cysteines in the *Eco*RI endonuclease, so specific labeling with IAEDANS at the position of our choice should be possible. Asp185 is a likely candidate for substitution since this amino acid is located at the tip of the outer arm segment, is located about 15Å from the DNA, exhibits significant mobility in the complex (average B-factor = 65 $Å^2$) and is solvent exposed. The geometry of the Trp246-Cys185 donor-acceptor pair might permit us to distinguish between hinge motion and complete disorder to order transition as potential mechanisms for arm folding. Cysteine substitution within the inner segment of the arm might be more problematic due to the proximity of this element to the DNA.

Although residues at the tip of this arm, Lys130 and Arg131 are flexible in the complex, they are located within 5 Å of the DNA backbone, and are therefore more likely to interfere with enzyme function.

Measuring the time resolved emission anisotropy decay of a fluorophore located within the arm offers a complementary approach to FRET designed to ask if there are differences in segmental motions of an enzyme region containing a flourophore. In these studies we propose to use the same IAEDANS labeled D185C muteins (in W104Y/promiscuous mutant background), and ask if there are differences anisotropy decay parameters of the IAEDANS fluorophore between the free and bound species, as well as between the wild-type and mutant proteins.

We are aware of two major caveats that may compromise these studies. The first being that the presence of the IAEDANS adduct due to its large size may interfere with DNA binding. Naturally, we will ask how the binding affinity of the labeled protein compares to wild-type binding affinity. Secondly, both methods proposed above are limited to probing processes that occur at rates comparable with the decay rates of the fluorophores involved. The decay rates for both tryptophan and IAEDANS are on the order of 1-10 ns. Significant motions of the inner and outer arm segments have been observed in a 1ns MD simulation of the *Eco*RI-GAATTC complex (Lisa Engler, and Linda Jen-Jacobson, *unpublished*), which give us reason to believe that the fluorescence experiments proposed above can report on the molecular processes in which we are interested. Finally, it might be instructive to repeat the experiments of Watrob et al. (Watrob et al., 2001) with the H114Y mutant, and ask how the decay associated spectra and anisotropy decay spectra of W246 differ between the free and bound protein as well as between the wild-type and mutant. However, we feel we can learn more about the conformational and dynamical properties of the arms by introducing a flourophore into the arms themselves.

Measuring the electron paramagnetic spectra of proteins labeled with site directed spin groups is another method that is well suited to provide inter-label distance constraints, and information on the dynamics of the enzyme segment in which the label is situated (Altenbach et al., 2001, Columbus et al., 2002). This method may be superior to fluorescence based methods due to the fact that spin labels are smaller than the fluorescent tags, and thus less likely to interfere with the folding of the arms and/or DNA binding. In addition, the range of inter-nitroxide distances that can be measured through site directed spin labeling is greater than the inter-fluorophore distances measurable by FRET. I have identified several amino acids in the

arms whose side-chains are flexible and solvent exposed, and thus are candidates for mutation to cysteine and subsequent labeling with a nitroxide spin label; they are Arg123, Arg131, Ser180, and Asp185.

X-ray crystallography remains unparalleled in its ability to provide high resolution information on the structures of macromolecules and their assemblies. A crystal structure of a miscognate DNA complex would significantly advance our understanding of the molecular basis of *Eco*RI endonuclease sequence discrimination. Much of the difficulty in obtaining crystals composed of true miscognate DNA complexes may derive from the fact that the binding affinities of miscognate and non-specific complexes are very similar, which results in partitioning of the enzyme towards the non-specific binding mode. Indeed, the preliminary structure of the *Eco*RI endonuclease in complex with the GACTTC miscognate site was solved from a crystal containing both miscognate and non-specific complexes within the lattice (Wilkosz, 1993). Non-specific complexes have been shown to be composed of a structurally diverse binding ensemble (complexes undergo large amplitude motions) (Kalodimos et al., 2004); this structural heterogeneity within the lattice manifests itself in disorder, which will in turn reduce the diffraction resolution limit of the crystal. We believe that promiscuous mutant enzyme-miscognate DNA complexes represent promising targets for crystallization, since the mutants bind more favorably to miscognate DNAs relative to non-specific DNA. Our data also show that crystallization of wild-type and mutant miscognate complexes might be more successful in the presence of calcium, which also increases the binding affinity for miscognate DNA relative to non-specific DNA. Using the promiscuous mutations and the inclusion of Ca^{2+} in the crystallization solution may reduce partitioning into the non-specific DNA binding mode, and potentially could increase the resolution of the diffraction data. A comparison of miscognate DNA footprints of complexes formed in the presence and absence of Ca^{2+}, demonstrates the ability of the ion to promote the miscognate DNA binding mode.

6. CHAPTER SIX: METHODS

6.1. Site directed mutagenesis

The *Eco*RI promiscuous mutants were originally isolated in the context of a temperature sensitive mutation, R56Q (Heitman and Model, 1990). Although this mutant showed similar *in vivo* activity to the wild-type enzyme at the permissive temperature (Muir et al., 1997), we did not want a secondary mutation to confound subsequent biochemical experiments on the promiscuous mutants. We therefore amplified the wild-type EcoRI endonuclease gene was from the plasmid pMB3 (Betlach et al., 1976), and inserted it into pET24a (Novagen) between the NheI and BamHI sites to create pPS12, in which transcription of the endonuclease gene is directed by the T7 RNA polymerase promoter. The A138T mutation (DNA specifying codon 138 changed from GCT to ACC) was engineered by sub-cloning a synthetic DNA fragment encoding the mutation into the pPS12 plasmid digested with BseRI and BglII (two endogenous restriction sites within the *Eco*RI endonuclease coding region), creating pPS13. These manipulations were carried out in strain DH5α (F-φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169 *deoR rec*A1 *endA1 hsdR17*(rk– mk+) *phoA supE44* λ– *thi-1 gyrA96 relA1*), which lacks T7 RNA polymerase and thus suffers only minimal toxic effects from the cloned promiscuous endonuclease gene.

Codons 114 and 192 did not have suitable flanking restriction sites, so the H114Y and E192K mutations were introduced by modifying the mutagenesis protocol of Deng and Nickoloff (Deng et al., 1992). We failed in multiple attempts to produce these mutants using the canonical protocol, in which two independent primer pairs are used to remove a restriction site (for selection against parental DNA) and to create a potentially toxic mutation. We hypothesized that this two-primer-pair mutagenesis strategy might lead to the enrichment of plasmids that incorporated only the product from the selection primers, because cells containing these plasmids

would be expected to have a growth advantage over cells containing plasmids that incorporated the products directed by both primer pairs. We therefore adopted a strategy by which both removal of a restriction site and introduction of the desired mutation in the EcoRI gene are accomplished by incorporation of a single primer.

To obtain the E192K mutant, we first introduced a XhoI site by silently changing the L191 codon from 5'-CTT to 5'- CTC, creating pWM1. The subsequent round of mutagenesis used a single primer to remove the XhoI site from pWM1 and introduce the E192K mutation by changing codon 192 from GAG to AAG, creating pWM2. The H114Y mutant was introduced in a similar manner, where a SpeI site was added by silently changing the L108 codon from 5'-CTA to 5'-CTT, creating pML1. This site was then removed and the H114Y mutation introduced by changing codon 114 from CAC to TAC, thereby creating pCD5. These manipulations were carried out in strain Xl mutS of genotype Δ(mcrA)183 Δ(mcrCBlsdSMR-mrr)173 end A1 supE44 thi-l gyrA96 relA1 lac mutS::Tn10(Tetr) (F' proAB lac IqZΔM15 Tn5). The DNA sequence of the entire *Eco*RI endonuclease coding region was confirmed for each of the plasmids encoding a promiscuous mutation.

6.2. Three "extra" amino acids at the N-terminus of *Eco*RI endonucleases used in this study

When sub cloning the amplified *Eco*RI endonuclease coding region into the pET24a vector between the NheI and *Bam*HI restriction sites (start codon of the gene is proximal to the NheI site), I failed to notice that there was an *Nde*I restriction site immediately upstream of the aforementioned *Nhe*I site. Therefore, translation is expected to begin at the start codon within the *Nde*I site rather than from the canonical ATG codon within the *EcoRI* endonuclease gene. The amino-acid sequence of the N-terminus of the wild-type protein is (Met)Ser-Asn-$(x)_{273}$, however, the sequence of the endonucleases used in this work is (Met)Ala-Ser-Met-Ser-Asn-$(x)_{273}$. The amino acid sequences of the endonucleases were confirmed by Edman degradation (performed by Dr. John Hempel). I tested a few of the biochemical properties of the 'wild-type' endonuclease with the three extra amino acids and found that this enzyme is indistinguishable

from the wild-type with respect to specific DNA binding affinity, specific DNA cleavage rate and cleavage rate of the AAATTC miscognate DNA.

6.3. Enzyme expression, purification, and characterization

Expression of the various mutant endonuclease genes produces effects that range for toxic (E192K) to lethal (A138T), so it was necessary to limit basal expression rigorously while still allowing for high-level expression when desired for enzyme production. Plasmids encoding the *Eco*RI endonucleases were transformed into the *E. coli* strain ER2566 (*=fhuA2 [lon] ompT lacZ::T7 genel gal sulAll Δ(mcrC-mrr)l 14::IS10 R(mcr-73::miniTn10--Tet^S)2 R(zgb-210::Tn10--Tet^S) end A1 [dcm]*) (New England Biolabs), already containing a plasmid (pAXU22-8, gift from New England Biolabs) expressing the *Eco*RI methylase. In this strain, catabolite repression by glucose can be used to reduce basal expression of the T7 RNA polymerase from the *lac* promoter. Nine liter cultures were grown to mid-log phase in LB medium (10g/L tryptone, 5g/L yeast extract, 10g/L NaCl) supplemented with 20g/L glucose (to provide catabolite repression) and the appropriate antibiotics, at which time the temperature was reduced from 37 °C to 21°C, and IPTG (final concentration of 1mM) was added to induce expression. Induction at the reduced temperature permitted good expression while slowing down cell death (as measured by levels of cellular proteins in the culture supernatant). The endonucleases were purified as described previously (Cheng et al., 1984). The enzymes were assessed to be greater than 99% homogeneous by SDS polyacrylamide gel electrophoresis.

Absolute protein concentrations were determined by loading SDS gels with known amounts of a protein whose concentration was determined by direct amino acid analysis using norleucine as an internal standard (Jen-Jacobson et al., 1983), and known volumes of protein preparations of unknown concentrations. The gels were then stained with Coomassie blue, scanned, and the optical densities of bands from the sample of known concentration were used to generate a standard curve, from which the concentration of the unknown samples could be determined.

6.4. Purification of oligonucleotide substrates

Oligonucleotides were synthesized on the Expedite 8909 synthesizer at the University of Pittsburgh DNA synthesis facility. The DNAs were purified by electrophoresis on 16% polyacrylamide gels (1x TBE, pH 8.3, 8.3 M Urea). Gels were run at constant current (75 milliamps) until the bromophenol blue dye-front traveled 25-30 cm. DNA was visualized by UV shadowing (Hassur et al., 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 then loaded onto an activated NEN-sorb 20 or Alltech C-18 column. Columns were washed with 3-5 volumes of water (relative to sample volume). The DNA was eluted from the column with 30% ethanol.

6.5. 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 et al., 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 0.22M 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 DNAs. As a third method 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%.

6.6. 5'-End Labeling of DNA

Single stranded or duplex DNAs were 5'-phosphorylated with γ-^{32}P-ATP and T4 polynucleotide kinase (PNK) as previously described (Jen-Jacobson et al., 1983), with the following modification. Usually 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). For equilibrium binding experiments, duplex DNA was labeled, whereas for cleavage experiments, single strands were labeled and then mixed such that the number of counts from the two labeled strands was the same. Reactions were terminated by adding 35 volumes of buffer A (appendix x) and loaded onto NENsorb-20 or Alltech C-18 columns to separate labeled DNA from unincorporated nucleotide. Purified labeled DNA was eluted from the column with 30% ethanol and dried under vacuum or using the TurboVap (Zymark Inc.). Dried DNA was then dissolved in filter buffer plus 50mM KCl and annealed as described above. Specific activity of labeled DNA was typically 5 x 10^{18} to 2 x 10^{19} cpm/mole.

6.7. 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., 1970, Lin et al., 1972), which are described in the next two sections.

6.7.1. 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 a set of

binding reactions had reached equilibrium, sets of 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 0.5ml/10 sec with vacuum applied from below), through a 25mm nitrocellulose filter that had been soaked in FB, followed by a wash with 350µl 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:

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

where $[D]_f$ and $[E]_f$ are the respective free concentrations of DNA and enzyme, [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 $0.2*K_D$ ($K_D=1/K_A$), and the total enzyme concentration [E] is much higher than the DNA concentration [D], all of the enzyme can be treated as unbound, and equation 6.1 can be simplified into:

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

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.

6.7.2. 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 specific sequences or solution conditions (poor flanking context, high salt, or high pH), all miscognate DNAs, and non-specific DNA 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 reference specific DNA 'probe' is much greater than the filtering time ($t_{1/2}$ >>> 6 sec.). Reactions contained the radio labeled reference probe (usually at a concentration of $6*K_D$), 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 and incubated on ice for 5 minutes to chill the solution; enzyme was then added, and the reaction was incubated on ice for another 5 minutes. The reaction was then transferred to a 21°C water bath and equilibrated for 30 minutes. Usually 10 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 were 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:

$$[ED_1] = \frac{-b - \sqrt{b^2 - 4ac}}{2} \qquad (6.3)$$

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

$$b = -\frac{K_1 + [D_2]_t * K_1}{K_2 + [E]_t + [D_1]_t},$$

$$a = 1,$$

$$c = [D_1]_t * [E]_t,$$

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 6.7.1) reactions were included to determine K_1 on the same day as the equilibrium competition experiments were performed. For all sets of competition experiments with miscognate and specific sites, a reaction series using an unlabeled specific site as competitor (D_2), whose binding constant could also be determined by direct equilibrium binding was usually included as an internal control.

6.8. DNA Cleavage Kinetics

All cleavage rate constants were obtained under single-turnover conditions. Binding buffer, pH 7.3 containing radio-labeled DNA (0.5 μM) plus 0.1M NaCl was chilled on ice for about 5 minutes, at which time, *Eco*RI endonuclease (1.2 μM) was added. Pre-equilibration proceeded on ice for 5 minutes, after which time the reaction was transferred to a 21°C water bath, and equilibrated for an additional 20 minutes. Cleavage reactions were initiated by the addition of the Mg^{2+} cofactor such that the final $MgCl_2$ concentration was 8mM. Small aliquots of the reaction were removed at various times and cleavage was quenched 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 for times tailored to the number of cpms loaded per lane. Band intensities were determined with a 3XC flat-bed scanner; scanning was conducted on the transmissive setting. The linear range of optical density for this scanner is 0.05 to 3.05, and the intensity of all bands were within this range if the band was to be included in the analysis. 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):

$$P1 = \frac{S_0}{2}\left[1 - e^{-(k_1+k_2)t} - \frac{k_2}{k_4 - k_1 - k_2}\left(e^{-(k_1+k_2)t} - e^{-k_4 t}\right)\right]$$
$$P2 = \frac{S_0}{2}\left[1 - e^{-(k_1+k_2)t} - \frac{k_2}{k_3 - k_1 - k_2}\left(e^{-(k_1+k_2)t} - e^{-k3t}\right)\right] \qquad (6.4)$$

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).

6.9. Kinetics of dissociation

Dissociation rate constants (k_d) for *Eco*RI endonuclease-DNA complexes were determined as described previously (Riggs et al., 1970, Jen-Jacobson et al., 1986). Briefly, protein (at concentrations 500 to 1000-fold over the K_D) was incubated with a 1.5 to 2-fold excess of ^{32}P-γ-ATP labeled DNA in binding buffer (Appendix B) for 5 minutes on ice, followed by 20 minutes at 21°C. Dissociation was initiated by the addition of 50 to 100-fold excess of cold DNA whose binding constant is as good or better than the labeled substrate. Aliquots were removed at desired time points and filtered through nitrocellulose discs. The dissociation constant k_d was determined by linear regression analysis, using the following equation (Riggs et al., 1970):

$$\ln[ED] - \ln[ED]_0 = k_d * t \qquad (6.5)$$

where $[ED]_0$ is the concentration of the enzyme-labeled DNA complex at the zero time point, and [ED] is the concentration of the enzyme-labeled DNA complex at time t.

6.10. van't Hoff Analysis

The effect of temperature on the equilibrium association constant was measured by direct filter binding (section 6.7.1) in a temperature controlled room over a temperature range of 3 to 37°C ± 0.2°C. Binding reactions were conducted in Binding Buffer (10 mM Cacodylate, 0.27M KCl, 1mM EDTA, 0.01% NaN_3, 100μM dithiothreitol, 100μg/ml bovine serum albumin, pH 7.3). The pH of the buffer was checked at each individual temperature, thus we were observing an "open system" according to Horn et al. (Horn et al., 2002); 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, binding reactions were incubated at the extreme hot and cold temperatures included in the analysis, and subsequently transferred to 21°C, equilibrated at this temperature and filtered. The binding constants determined from these reactions were the same as those from determined from reactions incubated at 21°C only.

The observed equilibrium binding constant is related to the heat capacity of the system by the following equation:

$$\ln K_{obs.} = \left(\frac{\Delta C^{\circ}_{p}}{R}\right)\left[\left(\frac{T_H}{T}\right) - \ln\left(\frac{T_S}{T}\right) - 1\right] \qquad (6.6)$$

where ΔC°_{p}, T_H, T_S, and R represent the observed heat capacity change for the binding reaction at constant pressure, the temperature at which $\Delta H^{\circ}=0$, the temperature at which $\Delta S^{\circ}=0$, and the universal gas constant respectively. A multiparametric fit to the curve generated by plotting ln $K_{obs.}$ versus 1/T yields the ΔC°_{p}, T_H, T_S parameters using the Sigmaplot computer program (Sigma Plot, Jandel Inc.; San Rafael, CA). The free energy (DG), enthalpy (DH) and entropy (DS) changes for the binding reaction can then be calculated from the following equations which assume that the heat capacity change is independent of temperature over the experimental temperature range (2° to 37°C):

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ = \Delta C^\circ_p \left[\left(T - T_H\right) - T \ln\left(\frac{T}{T_S}\right) \right] \qquad (6.7)$$

$$\Delta H^\circ = \Delta C^\circ_p \left(T - T_H\right) \qquad (6.8)$$

$$\Delta S^\circ = \Delta C^\circ_p \ln\left(\frac{T}{T_S}\right) \qquad (6.9)$$

6.11. Ethylation Interference Footprinting

*Eco*RI endonuclease footprints on miscognate DNAs were determined by the ethylation interference method (Siebenlist and Gilbert, 1980). To ethylate the DNA, 6.25 x 10^{-9} moles of a single stranded oligonucleotide was mixed with 100 µl of N-ethyl-N-nitrosourea (ENU, 100mg/ml dissolved in 100% ethanol); 25 µl of 200mM cacodylate (pH 8.0) was then added, and the volume of the reaction was brought to 200 µl with ddH_2O. The elthylation reaction was carried out at 50°C. In order to determine the duration of the reaction that results in one ethylation event per DNA molecule, the aliquots of the reaction mixture were removed at various time points, ethylated DNA was labeled at the 5' end with ^{32}P, and then cleaved at ethylated phosphates using cleavage buffer (10mM Na_2HPO_4, 1mM Na_2EDTA, 150mM NaOH) at 90°C for 30 minutes. The cleavage reaction is stopped by the addition of 10M urea containing enough bromo-phenol blue to visual the sample as it is loaded onto a gel. DNAs were then separated using denaturing PAGE (22% acrylamide, 8.3mM Urea, 1x TBE, pH8.3), and the gels were exposed to film. The densitometric analysis of bands within the autoradiogram allowed me to sum up the optical density of the bands arising from alkaline cleavage, and express this sum as a fraction of the optical density of the band corresponding to the full length DNA (un-cleaved, therefore un-ethylated). The Poisson distribution says that the time point that resulted in a 30%:70% cleaved:un-cleaved ratio corresponds to the reaction duration where one ethylation event per molecule occurs. For the 22 base oligonucleotides used in these studies, a 90 minute ethylation reaction resulted in one ethylation event per molecule.

Labeled, ethylated DNAs were annealed to their un-labeled, un-ethylated complementary strand. This duplex DNA was used to make *Eco*RI complexes in binding buffer (37mM NaCl, 3mM $CaCl_2$, pH 7.3). Concentrations of the DNA and enzyme were adjusted such that 30% of the DNA was enzyme bound at equilibrium. Free and bound DNAs were then separated by native PAGE [10% 19:1 acrylamide, 0.5X TA+$CaCl_2$ (see appendix B for buffer recipe)], where the free and bound species were located within the gel using the autoradiogram as a template. The free and bound bands were excised from the gel, and the DNAs were eluted from the gel slices in buffer A, and re-purified over Alltech C-18 columns.

The protein bound, and free DNA fractions, as well as a control reaction (DNA that was never exposed to the protein) are alkaline treated (see conditions above) to cleave the phosphotriestester bonds that occurs at the point of DNA ethylation. Equal counts from each reaction are then run in separate lanes on a 22% denaturing polyacrylamide gel (see above for precise gel conditions). Alkaline cleavage of the phosphotriester bond results in two 5'-radiolabeled products (two bands on autoradiogram of urea-PAGE representing the 3'OH or 3'-OPOEt cleavage products). The identity of one band of the doublet was determined by comparison with the migration of set of size standards (5'-radiolabled oligonucleotides that terminate in 3'-OH). Interference patterns are obtained from the autoradiograms using methods of analysis described in (Lesser et al., 1990).

6.12. Crystallization, structure determination and structural analysis of the A138T *Eco*RI-d(TCGCGAATTCGCG) complex

6.12.1. Preparation of macromolecules and crystal growth

The A138T mutant EcoRI endonuclease was purified as described in section 6.3. The purified enzyme was then dialyzed into a buffer solution containing 40 mM Bis Tris-Propane, 0.5 M NH_4Cl_2, 1 mM EDTA, 15% v/v Dioxane, pH 8.5, and concentrated to 6 mg/ml using Aquacide II. The self-complementary oligonucleotide, d(TCGCGAATTCGCG) was synthesized at the University of Pittsburgh DNA Sequence Facility and purified by reverse phase HPLC, using a Hamilton PRP-1 column. The purification was carried out at 90°C. The oligonucleotide

was eluted from the column with a linear gradient from 0-50% acetonitrile in 0.050 M elthylenediamine, 0.065 M formic acid. Peak fractions were combined, dried down and re-suspended in a suitable volume of de-ionized water. The DNA was placed in a 95°C water bath which was then turned off to allow annealing of the DNA strands. The DNA duplex concentration was adjusted to 10 mg/ml using the convention: 1 OD_{260} = 50μg/ml of double stranded DNA.

The protein-DNA co-crystals were grown by vapor diffusion at 4°C. 1 ml of reservoir solution (40 mM BTP), 16% v/v PEG 400, pH 6.5) was added to the reservoir compartment. 7μl DNA (10mg/ml), 6 μl enzyme (6mg/ml) and 3μl of PPT solution (40mM BTP, 40% v/v PEG 400, pH 7.0) were added to the sitting drop in the order listed. Crystals of approximate dimensions of 400 x 200 x 100 microns were obtained after about a month.

6.12.2. Crystal cryoprotection and freezing

For cryoprotection, the crystals were soaked for three minutes in reservoir solution supplemented with 10% glycerol and then transferred to reservoir solution with 15% glycerol and allowed to soak to another three minutes. Crystals were then flash frozen in loops under liquid nitrogen.

6.12.3. Data collection and processing

Diffraction data were collected from a single crystal at the X6A beam-line of the National Synchrotron Light Source. The crystal temperature was maintained at 100K during data collection. Data were collected at a single X-ray wavelength of 0.979 Å. Diffraction data were collected using the rotation/oscillation method (Dauter, 1999, Minor et al., 2000), with a rotation angle of 1° per frame. Two hundred frames of diffraction data were collected with an X-ray exposure time of 90 seconds. Two hundred additional frames were collected with an increased crystal to detector distance using an exposure time of 10 seconds in order to accurately record low resolution reflections. The diffraction spots were indexed, integrated and scaled using the

DENZO/SCALEPACK software package (Otwinowski, 1997). Statistics for data collection and data processing are shown in Table 3.1.

6.12.4. Structure determination and refinement

The structure of the A138T-d(TCGCGAATTCGCG) complex was solved by molecular replacement, using the high resolution structure of the *Eco*RI-d(TCGCGAATTCGCG) complex (Grigorescu, 2003) from which the solvent molecules were removed was used as the initial atomic model. The Crystallography and NMR System (CNS) (Brünger et al., 1998) was utilized for all stages of refinement. 10% of the reflections were randomly chosen to be used in the cross validation set for calculating R_{free} (Brünger, 1992) during refinement. After initial rigid body refinement, several rounds of simulated annealing torsion angle refinement and Cartesian refinement with maximum likelihood target were carried out in CNS (Brünger, 1987, Rice et al., 1994, Adams et al., 1997). Engh and Huber bond and angle parameters were used for the protein (Engh, 1991), and parameters from Parkinson et al. were used for the DNA (Parkinson, 1996). Using the phases from the resultant model, $2F_o$-F_c and F_o-F_c difference maps were calculated in CNS with σ_a weights for refitting the model. Simulated annealing omit maps and a composite omit map were calculated in CNS as tools to reduce model bias during refinement. The resultant maps and real space R-factors (Jones et al., 1991) were used as guides for making adjustments to the atomic model. Manual adjustments to the main chain torsion angles were not necessary, and the adjustments to side-chain conformations were made using the program O (Jones et al., 1991).

I used the 'water_pick' utility within CNS to build solvent molecules into residual electron density that was not assigned to the macromolecules. The bound solvent was treated as water, although the electron density might be attributable to other species present in the crystallization buffer (NH_4^+, H_3O^+, OH^-). Putative solvent peaks must exist above the 3σ contour level in difference maps, and have good hydrogen bonding geometry with other solvent molecules, the protein, or the DNA. Waters were deleted if their corresponding electron could be interpreted as an alternative position of a side-chain, if the peak was not visible at the 1σ level in subsequent $2F_o$-F_c maps, or if the B-factor for a given solvent atom rose above 75 $Å^2$ during subsequent refinements. 136 bound water molecules which satisfied the above criteria were added to the atomic model.

The 'optimize_wa.inp' utility within CNS was used in order to adjust the weights given to the diffraction data and stereochemical restraints during refinement. A weight was chosen that yielded the lowest R_{free} while maintaining reasonable values for the bond lengths and angles of the model. An additional round of simulated annealing Cartesian refinement with maximum likelihood target was then carried out in CNS. Finally, an anisotropic B-factor correction was applied, and individual restrained B-factor refinement was carried out in CNS. The restraints on the bonded atoms were derived from the weight optimization protocol implemented in CNS.

6.12.5. Quality checks on the final model

Refinement statistics for the final model are given in table 3.1. The software package, SFCHECK (Vaguine et al., 1999) was used in order to obtain estimates of the coordinate error in the model, as well as to receive detailed feedback on the agreement between the diffraction pattern and the final atomic model. The output from the SFCHECK analysis is given in Appendix D. The stereochemistry of the final model was analyzed by PROCHECK (Laskowski, 1993) and WHATIF (Hooft et al., 1996). Both of the PROCHECK and WHATIF analyses were carried out by a single web server at: http://biotech.ebi.ac.uk:8400/. The outputs from these checks are given in Appendix E.

6.12.6. Structural analysis

The protein secondary structure was assigned using the program STRIDE (Heinig and Frishman, 2004). The local base-pair and base-step conformational parameters of the DNA in the model were obtained using the program 3DNA (Lu and Olson, 2003). Output from this analysis and a complete comparison of DNA conformations from the wild-type and A138T models can be found in Appendix F. Intra and inter-molecular hydrogen bonds within the structures were assigned using the program HBPLUS (McDonald and Thornton, 1994). Superposition of the various structures, and rmsd calculations were carried out using O (Jones et al., 1991), Midas Plus (Ferrin, 1988), VMD (Humphrey et al., 1996), or SPLAT (J. M. Rosenberg, unpublished). The estimation of expected coordinate error (σ) between two

structures as calculated by the SPLAT program is based on a relation proposed by Stroud and co-workers (Perry et al., 1990), which takes into account resolution of the data sets and atomic B-factors of the models being compared:

$$\sigma = 0.75 * R * r * (0.0015B^2 - 0.203B + 0.359) \qquad (6.10)$$

where *R* is the (conventional) crystallographic R-factor, *r* is the resolution of the data-set in Å, and B is the Debye-Waller temperature factor.

6.13. Determination of *Eco*RI* site abundance in the Escherichia coli genome

The DNA sequence of the *Escherichia coli* K12 genome (strain MG1655, GenBank accession number U00096) was scanned for *Eco*RI* sites using 'DNA Master' software (freeware written by Dr. Jeffrey G. Lawrence, Department of Biological Sciences, University of Pittsburgh; software available at http://cobamide2.bio.pitt.edu/). The exact number of each *Eco*RI* site was then used to calculate concentration of that miscognate DNA in the *E. coli* cell based on an estimate of 1.6 E^{-15} L for the volume of the *E. coli* cytoplasm.

APPENDIX A

DNA SUBSTRATES USED IN THIS WORK

Sequence [a]	Length	Employed for techniques
5'-GGGCGGGCGCGAATTCGCGGGCGC CCCGCCGCGGCTTAAGCGCCCGCG-5'	24 bp	Equilibrium binding, pH dependence
5'-GGGCGGGGGTGAATTCACCGGCGC CCCGCCGCCACTTAAGTGGCCGCG-5	24 bp	Equilibrium Binding
5'-GGGCGGGGATGAATTCATCGGCGC CCCGCCGCTACTTAAGTAGCCGCG-5'	24 bp	Equilibrium Binding
5'-GGGCGGGCCCGAATTCGGGGGCGC CCCGCCCGGGCTTAAGCCCCCGCG-5'	24 bp	Equilibrium Binding
5'-GGGCGGGCCCGAATTCGGGGGCGC CCCGCCCGGGCTTAAGCCCCCGCG-5'	24 bp	Equilibrium Binding
5'-GGGCGGGGGCGAATTCGCCGGCGC CCCGCCCCCGCTTAAGCGGCCGCG-5'	24 bp	Equilibrium Binding
5'-GGGCGGGGCAGAATTCTGCGGCGC CCCGCCCCGTCTTAAGACGCCGCG-5'	24 bp	Equilibrium Binding
5'-GGGCGGGAAAGAATTCTTTGGCGC CCCGCCCTTTCTTAAGAAACCGCG-5'	24 bp	Equilibrium Binding
5'-GATGGGTGCAGAATTCTGCAGGTA CTACCCACGTCTTAAGACGTCCAT-5'	24 bp	pH dependence
5'-TCGCGAATTCGCG GCGCTTAAGCGCT-5'	12 bp	X-ray crystallography
5'-GGGCGGGCGCAAATTCGCGGGCGC CCCGCCCGCGTTTAAGCGCCCGCG-5'	24 bp	Equilibrium Binding, Cleavage
5'-GGGCGGGCGCTAATTCGCGGGCGC CCCGCCCGCGATTAAGCGCCCGCG-5'	24 bp	Equilibrium Binding, Cleavage

5’-GGGCGGGCGCGACTTCGCGGGCGC CCCGCCCGCGCTGAAGCGCCCGCG-5’	24 bp	Equilibrium binding, Cleavage
5’-GGGCGGGCGCGACTTCGCGGGCGC CCCGCCCGCGCTGAAGCGCCCGCG-5’	24 bp	Equilibrium binding, Cleavage
5’-GGGCGGGCGCCTTAAGGCGGGCGC CCCGCCCGCGGAATTCCGCCCGCG-5’	24 bp	Equilibrium binding, Cleavage
5’-GGGCGGGCGCAAATTTGCGGGCGC CCCGCCCGCGTTTAAACGCCCGCG-5’	24 bp	Cleavage
5’-GGGCGGGCGCTAATTAGCGGGCGC CCCGCCCGCGATTAATCGCCCGCG-5’	24 bp	Cleavage
5’-GGGCGGGCGCGACGTCGCGGGCGC CCCGCCCGCGCTGCATCGCCCGCG-5’	24 bp	Cleavage
5’-GGGCGGGGGCAAATTCGCCGGCGC CCCGCCCCCGTTTAAGCGGCCGCG-5’	24 bp	Equilibrium binding, Cleavage
5’-GGGCGGGGGTAAATTCACCGGCGC CCCGCCCCCATTTAAGTGGCCGCG-5’	24 bp	Equilibrium binding, Cleavage
5’-GGGCGGGCGTAAATTCACGGGCGC CCCGCCCGCATTTAAGTGCCCGCG-5’	24 bp	Cleavage
5’-GGGCGGGGCAAAATTCTGCGGCGC CCCGCCCCGTTTTAAGACGCCGCG-5’	24 bp	Equilibrium binding, Cleavage
5’-GGGCGGGTGTAAATTCACAGGCGC CCCGCCCACATTTAAGTGTCCGCG-5’	24 bp	Cleavage
5’-GGGCGGGGATAAATTCATCGGCGC CCCGCCCCTATTTAAGTAGCCGCG-5’	24 bp	Equilibrium binding, Cleavage
5’-GGGCGGGCCCAAATTCGGGGGCGC CCCGCCCGGGTTTAAGCCCCCGCG-5’	24 bp	Cleavage
5’-GGGCGGAAAAAAATTCTTTTGCGC CCCGCCTTTTTTTAAGAAAACGCG-5’	24 bp	Equilibrium binding, Cleavage
5’-CTGATGACGCAAATTCGCGTAG GACTACTGCGTTTAAGCGCATC-5’	22 bp	Ethylation interference
5’-CTGATGAGGTAAATTCACCTAG GACTACTCCATTTAAGTGGATC-5’	22 bp	Ethylation interference

5’-GGGCGGGC**7cG**CGAATTCG CGGGCGC CCCGCCCG CGCTTAAGC**7cG**CCCGCG	24 bp	Equilibrium binding
5’-GGGC**7cG**GGCGCGAATTCGCGGG CGC CCCG CCCGCGCTTAAGCGCCC**7cG**CG	24 bp	Equilibrium binding
5’-GGGCTGGGGTGAATTCACCGGAGC CCCGACCCCACTTAAGTGGCCACG-5’	24 bp	Equilibrium binding
5’-GGGCTGGGG**U**GAATTCACCGGAGC CCCGACCCCACTTAAG**U**GGCCACG-5’	24 bp	Equilibrium binding
5’-GGGC**U**GGGGTGAATTCACCGGAGC CCCGACCCCACTTAAGTGGCC**U**CG-5’	24 bp	Equilibrium binding

[a] Specific sites are colored red, miscognate sites blue, and non-specific sites green. We recognize that each of these oligonucleotides contain non-specific binding sites, however DNA substrates with a regions highlighted in green lack miscognate and specific sites.

APPENDIX B

BUFFER RECIPES

Buffer A: Buffer containing counter-ion for reversed phase purification of DNA over NENsorb, or Alltech C-18 columns.

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

Binding Buffer:

- 10 mM BTP
- 1 mM EDTA
- 0.01% w/v NaN_3
- 100 mg/ml bovine serum albumin
- 100 μM dithiothreitol (DTT)

pH and salt concentration as listed in text.

Filter Buffer:

- 10 mM BTP
- 1 mM EDTA
- 0.01% w/v NaN_3

pH and salt concentration as listed in text.

1X TBE: Tris-Borate-EDTA

- 100 mM Tris
- 89 mM Boric Acid
- 2.5 mM EDTA

pH 8.3 at 25°C without titration

1X TAE: Tris-Acetate-EDTA

- 40 mM Tris
- 20 mM Acetic acid
- 2.5 mM EDTA

1X TA + Ca Cl_2: Tris-Acetate plus $CaCl_2$

- 40 mM Tris
- 20 mM Acetic acid
- 2 mM $CaCl_2$

Cleavage Stop Buffer:

- 9 M Urea
- 50 mM EDTA

Enzyme Storage Buffer:

- 20 mM Na_2HPO_4 pH 7.0, 21°C
- 0.6 M NaCl
- 1 mM EDTA
- 0.01% w/v NaN_3
- 100 μM DTT
- 10% v/v Glycerol

APPENDIX C

ERROR ANALYSIS

The following lists the equations used to determine the means and standard deviations for the parameters used in this thesis.

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

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

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

2) $\Delta G°_{bind}$ and its standard deviation $\sigma_{\Delta G°bind}$ are calculated by the following equations:

$$\Delta G°_{bind} = -RT \ln \overline{K}_A \qquad \text{(C.3)}$$

$$\sigma_{\Delta G°_{bind}} = -RT \left(\frac{\sigma_{K_A}}{\overline{K}_A} \right) \qquad \text{(C.4)}$$

Calculation of $\Delta G°^{‡}$ and its standard deviation simply involves replacing $\overline{K}_A$ with $\overline{k}_{cleave}$ and σ_{K_A} with $\sigma_{k_{cleave}}$ in equations X.3 and X.4 respectively.

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

$$\Delta\Delta G^\circ_{bind} = \Delta G^\circ_{bind\,2} - \Delta G^\circ_{bind\,1} = -RT \ln \frac{\overline{K}_{A2}}{\overline{K}_{A1}} \qquad \text{(C.5)}$$

$$\sigma_{\Delta\Delta G^\circ_{bind}} = RT \sqrt{\left(\frac{\sigma_{\overline{K}_{A1}}}{\overline{K}_{A1}}\right)^2 + \left(\frac{\sigma_{\overline{K}_{A2}}}{\overline{K}_{A2}}\right)^2} \qquad \text{(C.6)}$$

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

$$\Delta\Delta G^{\circ\ddagger}_{I} = \Delta\Delta G^\circ_{bind} + \Delta\Delta G^{\circ\ddagger} \qquad \text{(C.7)}$$

$$\sigma_{\Delta\Delta G^{\circ\ddagger}_{I}} = \sqrt{\left(\sigma_{\Delta\Delta G^\circ_{bind}}\right)^2 - \left(\sigma_{\Delta\Delta G^{\circ\ddagger}}\right)^2} \qquad \text{(C.8)}$$

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:

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

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

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

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

where S^2 in equations X.11 and X.12 represent the covariance given by:

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

6) Propagation of errors for $\Delta H°_{obs}$ and $\Delta S°_{obs}$ parameters obtained from van't Hoff analysis is as follows:

$$\Delta H°_{obs} = \Delta C°_{p,obs}\left(T - T_H\right) \qquad \text{(C.14)}$$

$$\Delta S°_{obs} = \Delta C°_{p,obs} \ln\left(\frac{T}{T_S}\right) \qquad \text{(C.15)}$$

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

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

APPENDIX D

SFCHECK OF A138T-d(TCGCGAATTCGCG) CRYSTALLOGRAPHIC MODEL

Analysis of agreement between the observed and calculated structure factors for A138T-d(TCGCGAATTCGCG) complex atomic model using SFCHECK (Vaguine et al., 1999).

Title: MOL_ID: 1;
MOLECULE: DNA (5'-D(*TP*CP*GP*CP*GP*AP*AP*TP*TP*CP*GP*CP*G)-3');
CHAIN: C;
ENGINEERED: YES;
MOL_ID: 2;
MOLECULE:;
CHAIN: A;
ENGINEERED: YES
Date: XX-XXX-XX
PDB code: UNNA

Crystal

Cell parameters:
a: 116.47 Å b: 116.47 Å c: 48.48 Å
α: 90.00° β: 90.00° γ: 120.00°
Space group: P 3 2 1

Model

2451 atoms (136 water molecules)

Number of chains:	3
Volume not occupied by model:	49.0 %
<B> (for atomic model):	40.9 Å ²
σ(B):	13.27 Å ²
Matthews coefficient:	2.93
Corresponding solvent % :	57.67

Refinement

Program:	CNS
Nominal resolution range:	12.0 – 1.95 Å
Reported R-factor:	0.203
Reported Rfree:	0.24
Sigma cut-off:	N.A.

Structure Factors

Input

Nominal resolution range:		25.0 – 1.95 Å
Reflections in file:		27856
Unique reflections	above 0:	26211
	above 1σ:	25636
	above 3σ:	24212
Reflections ≤ 0:		1645

SFCHECK

Nominal resolution range: (max. from input data, min. from author)	25.0 – 1.95 Å
Used reflections:	26179
Reflections out of resolution:	32
Completeness:	94.1 %
R_stand(F) = <σ(F)>/<F> :	0.028
Anisotropic distribution of Structure Factors ratio of eigen values:	1.0000 1.0000 0.7300
B_overall (by Patterson):	34.0Å ²
Optical resolution:	1.64 Å
Expected opt. resol. for complete data set:	1.64 Å
Estimated minimal error:	0.028 Å

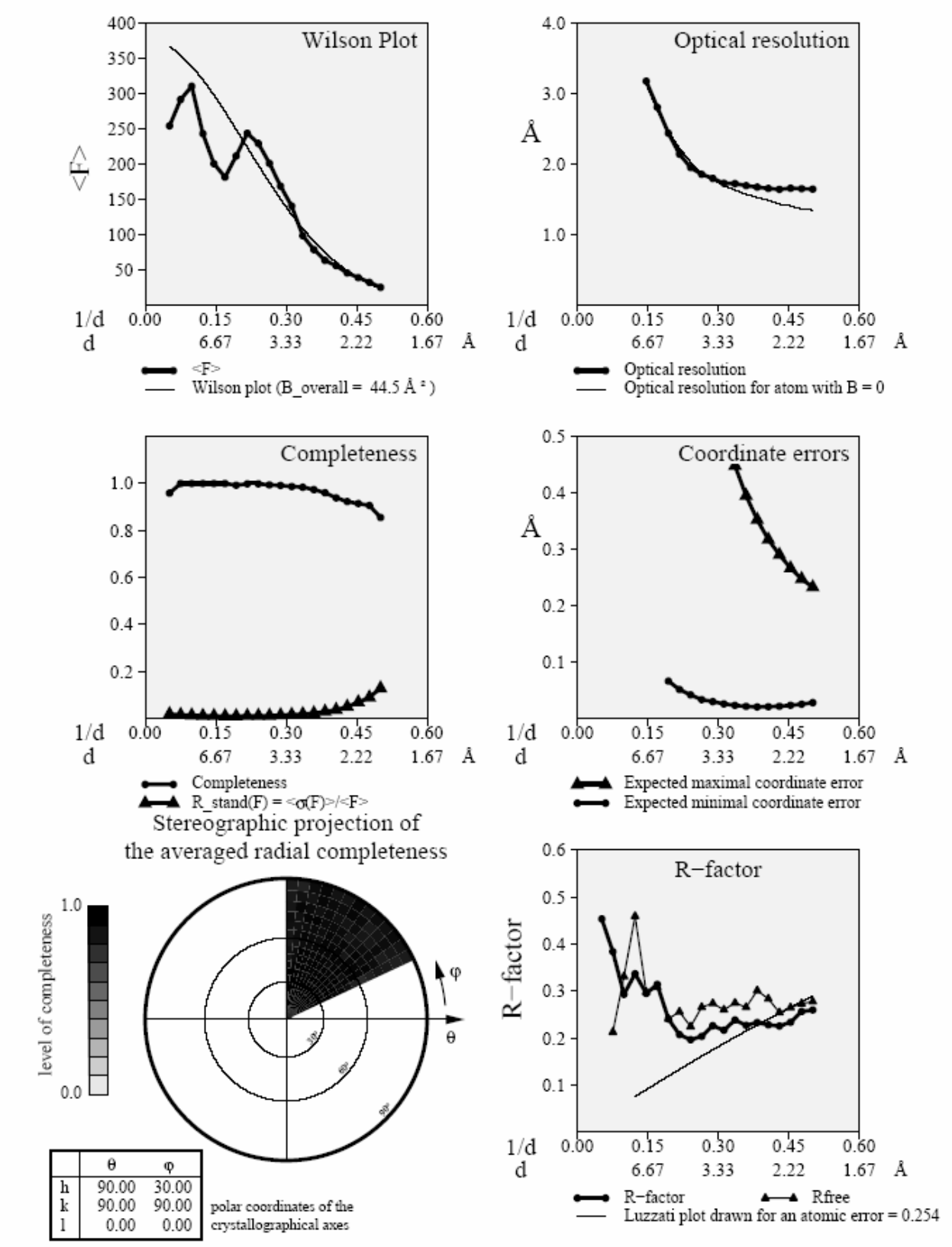
Model vs. Structure Factors

R-factor for all reflections:	0.232
Correlation factor:	0.936
R-factor:	0.231
for F > 2.0σ	
nom. resolution range:	12.00 – 1.95Å
reflections used:	25483
Rfree:	0.271
Nfree:	2543
R-factor without free-refl.:	0.227
Non free-reflections:	22940
<u> (error in coords by Luzzati plot):	0.254 Å
Estimated maximal error:	0.233 Å
DPI:	0.148 Å

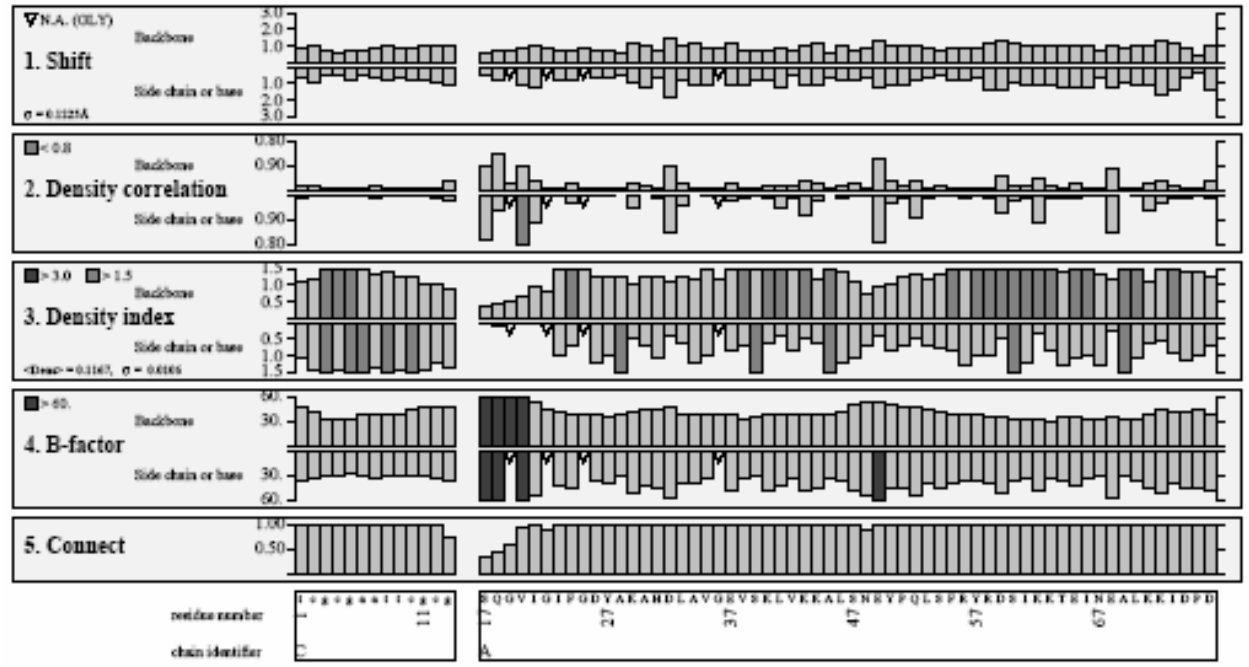
Scaling

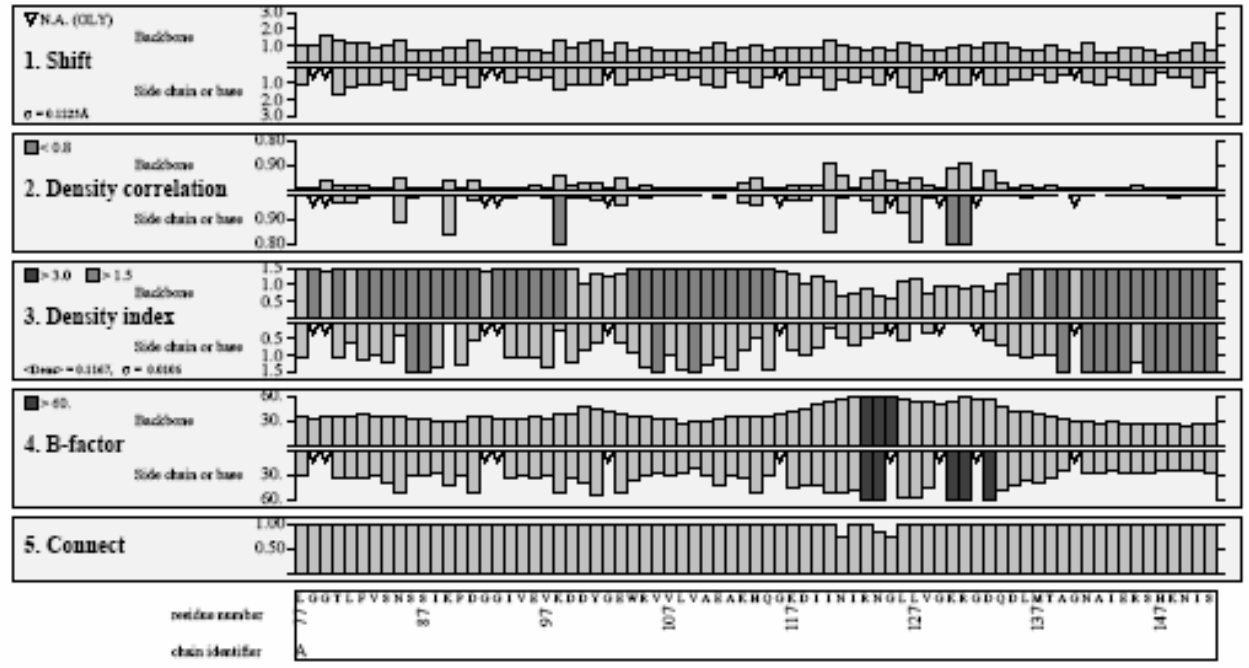
Scale:	0.440
Bdiff:	-4.20

Anisothermal Scaling (Beta):
1.9448 1.9448 -0.9313 0.9724 0.0000 0.0000
Solvent correction - Ks,Bs: 0.811 250.017

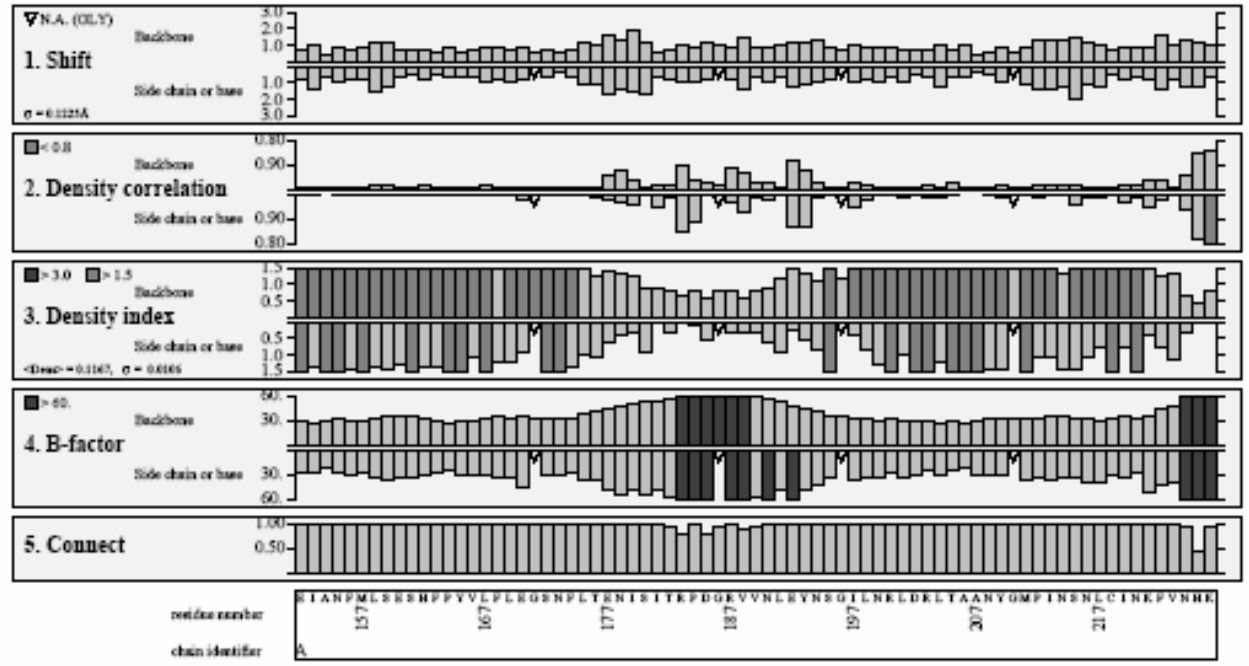

Wilson Plot
<F>
1/d
d
Å
<F>
Wilson plot (B_overall = 44.5 Å²)
Optical resolution
Optical resolution
Optical resolution for atom with B = 0
Completeness
Completeness
R_stand(F) = <σ(F)>/<F>
Coordinate errors
Expected maximal coordinate error
Expected minimal coordinate error
Stereographic projection of
the averaged radial completeness
level of completeness
θ
φ
h 90.00 30.00
k 90.00 90.00
l 0.00 0.00
polar coordinates of the
crystallographical axes
R−factor
R−factor
Rfree
Luzzati plot drawn for an atomic error = 0.254

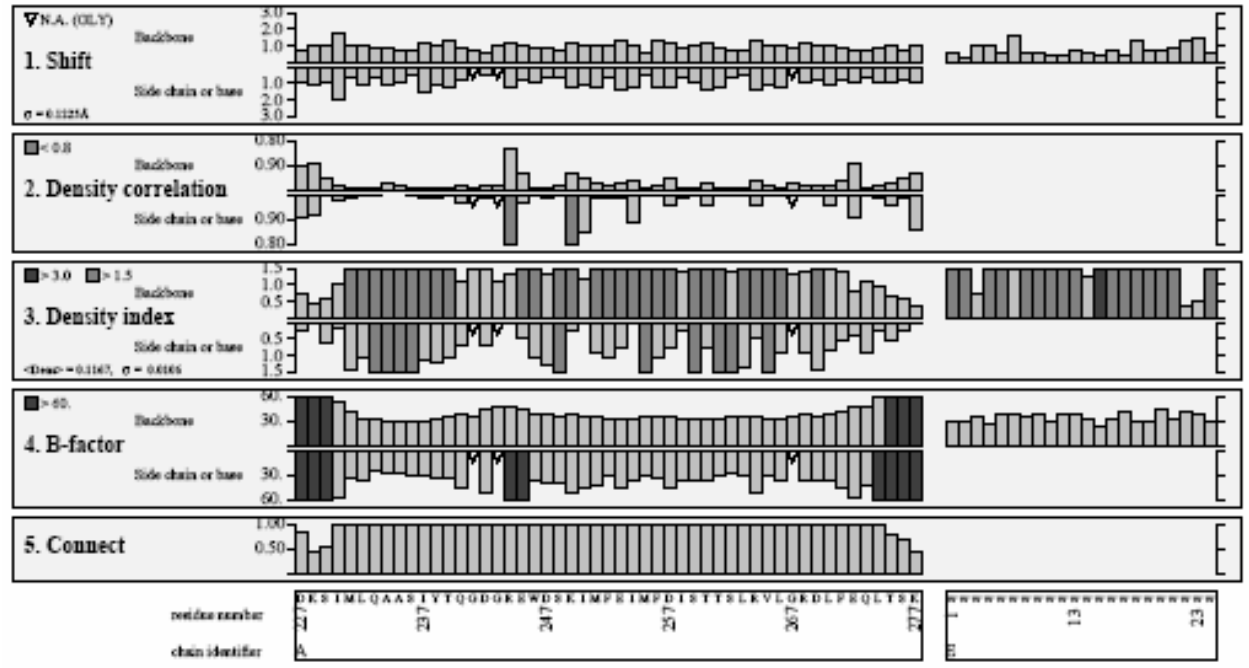
Local estimation



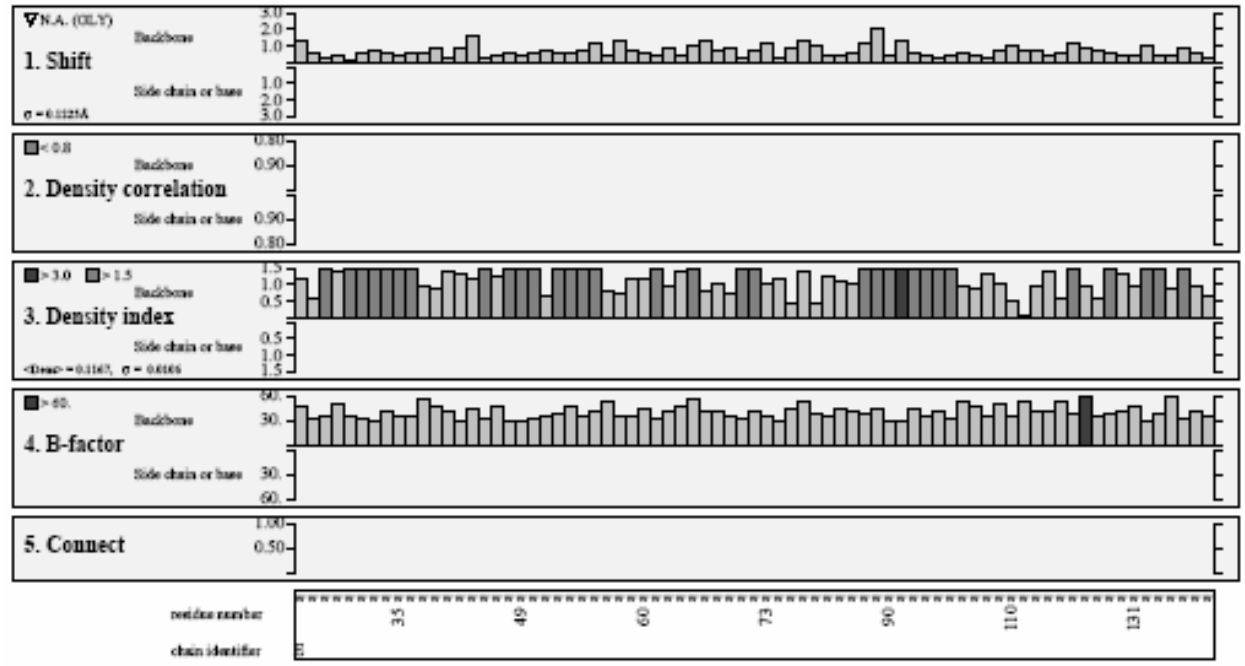


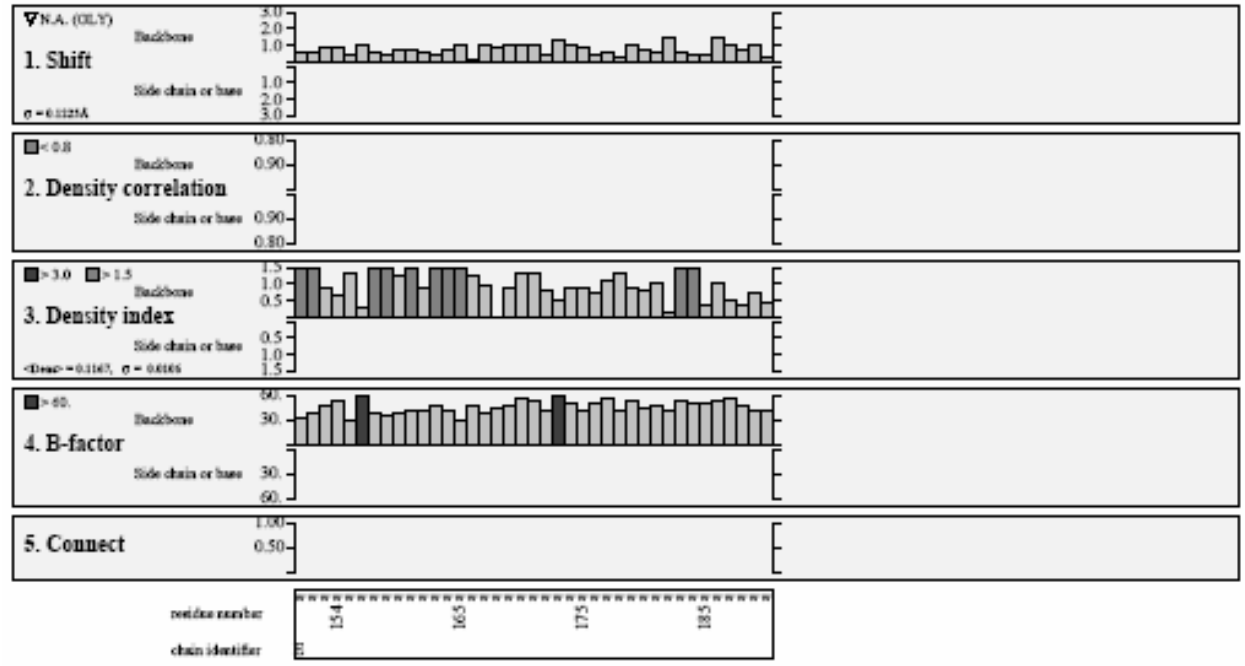
Local estimation (2)





Local estimation (3)

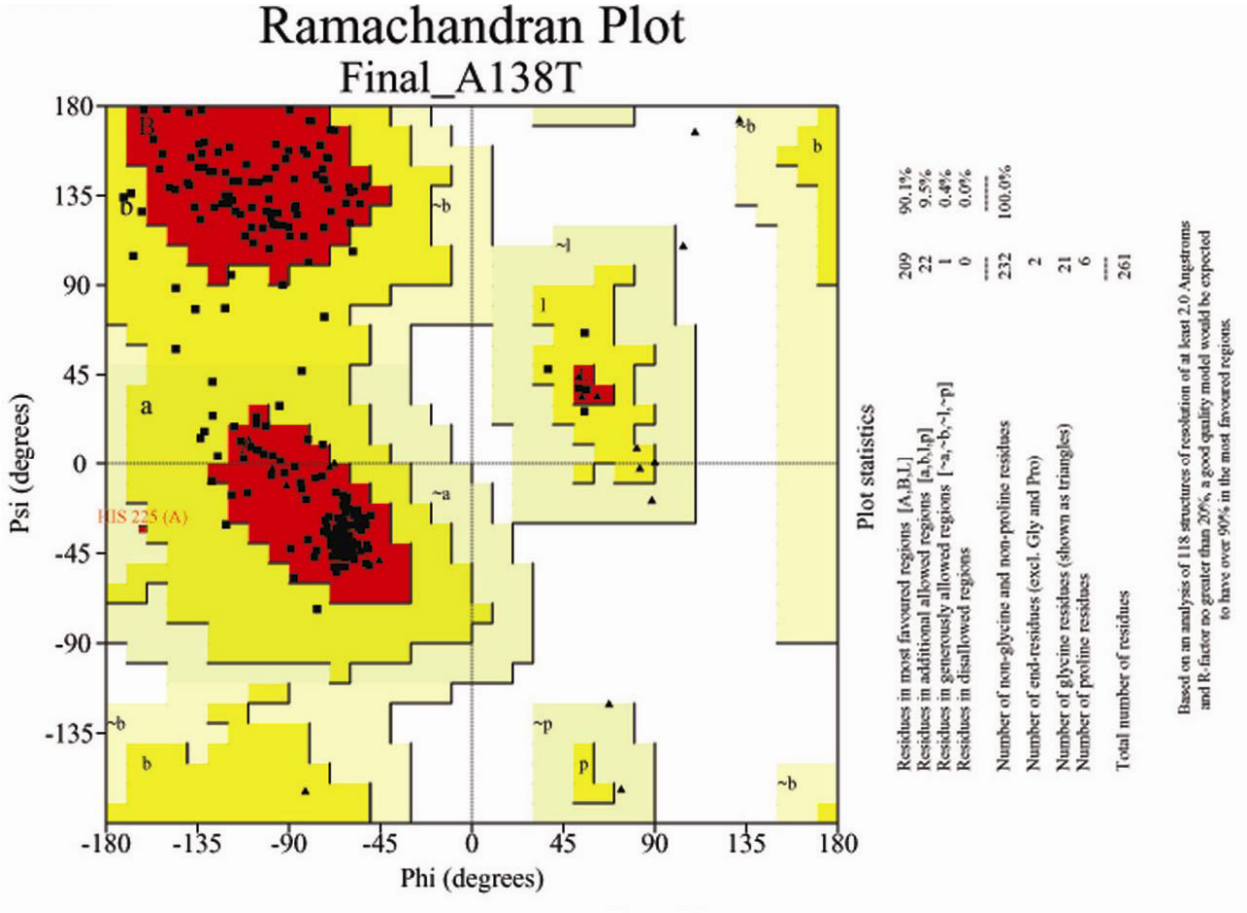




APPENDIX E

PROCHECK OUTPUT FOR MODEL OF A138T *Eco*RI-d(TCGCGAATTCGCG) COMPLEX

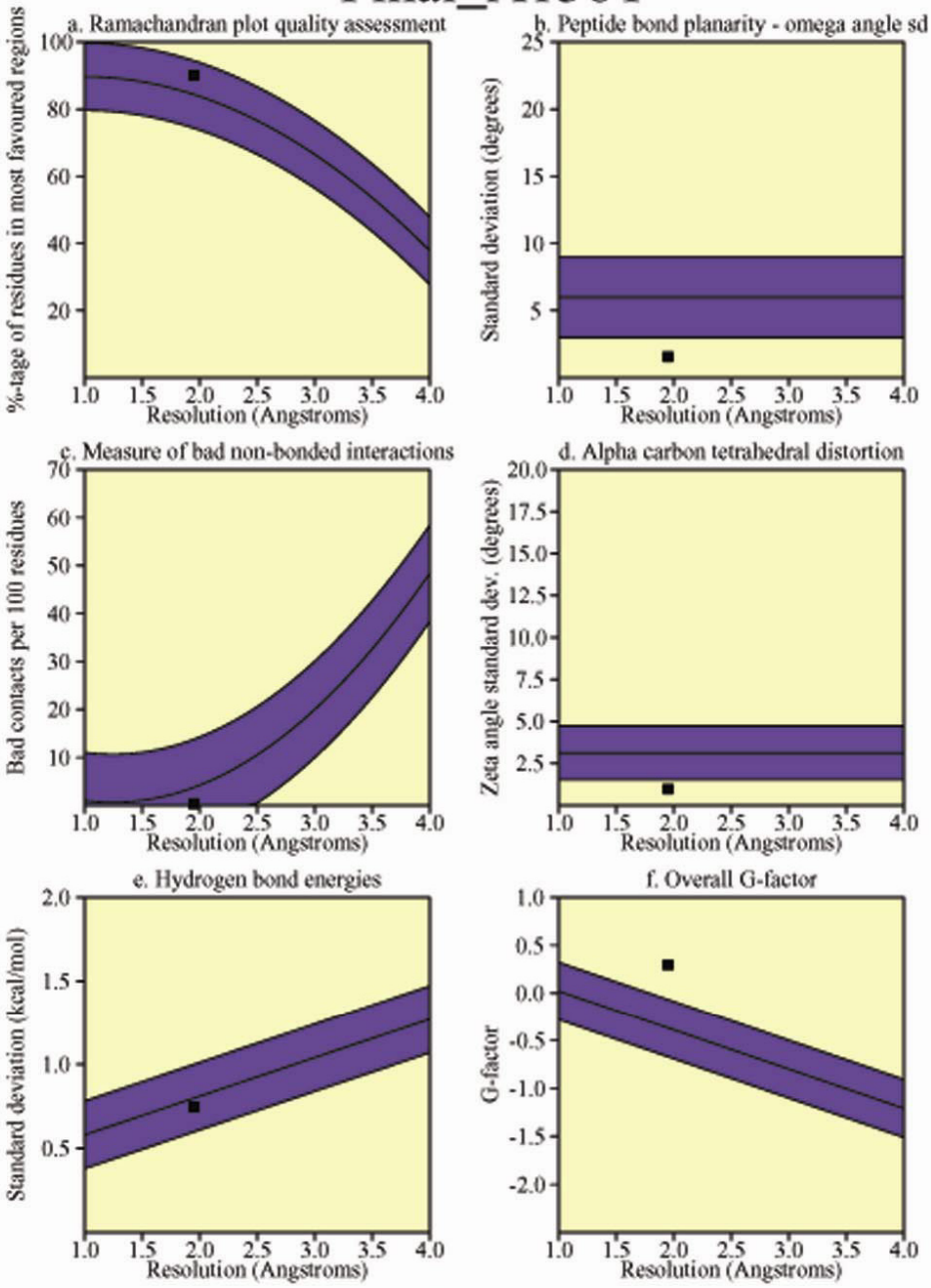
Checks on the stereochemistry of the refined model of the A138T-d(CGCGAATTCGCG) complex were performed using PROCHECK (Laskowski, 1993); the following reports were generated by the http://biotech.ebi.ac.uk:8400/ web server.



PROCHECK

Main-chain parameters

Final_A138T



Plot statistics

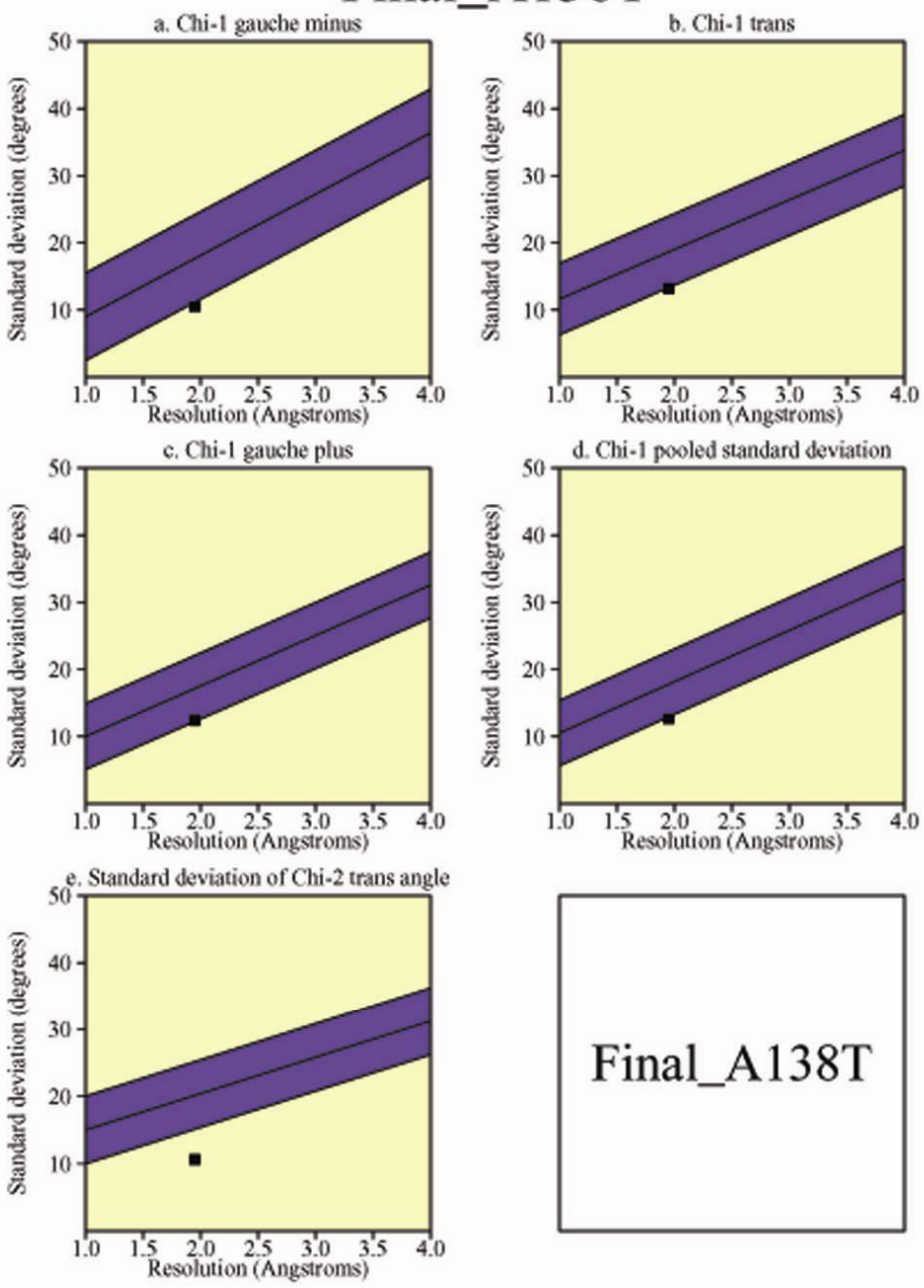
Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. %-tage residues in A, B, L	232	90.1	84.4	10.0	0.6	Inside
b. Omega angle st dev	260	1.6	6.0	3.0	-1.5	BETTER
c. Bad contacts / 100 residues	1	0.4	3.7	10.0	-0.3	Inside
d. Zeta angle st dev	240	1.0	3.1	1.6	-1.3	BETTER
e. H-bond energy st dev	169	0.7	0.8	0.2	-0.3	Inside
f. Overall G-factor	261	0.3	-0.4	0.3	2.2	BETTER

Final_A138T_04.ps

PROCHECK

Side-chain parameters

Final_A138T



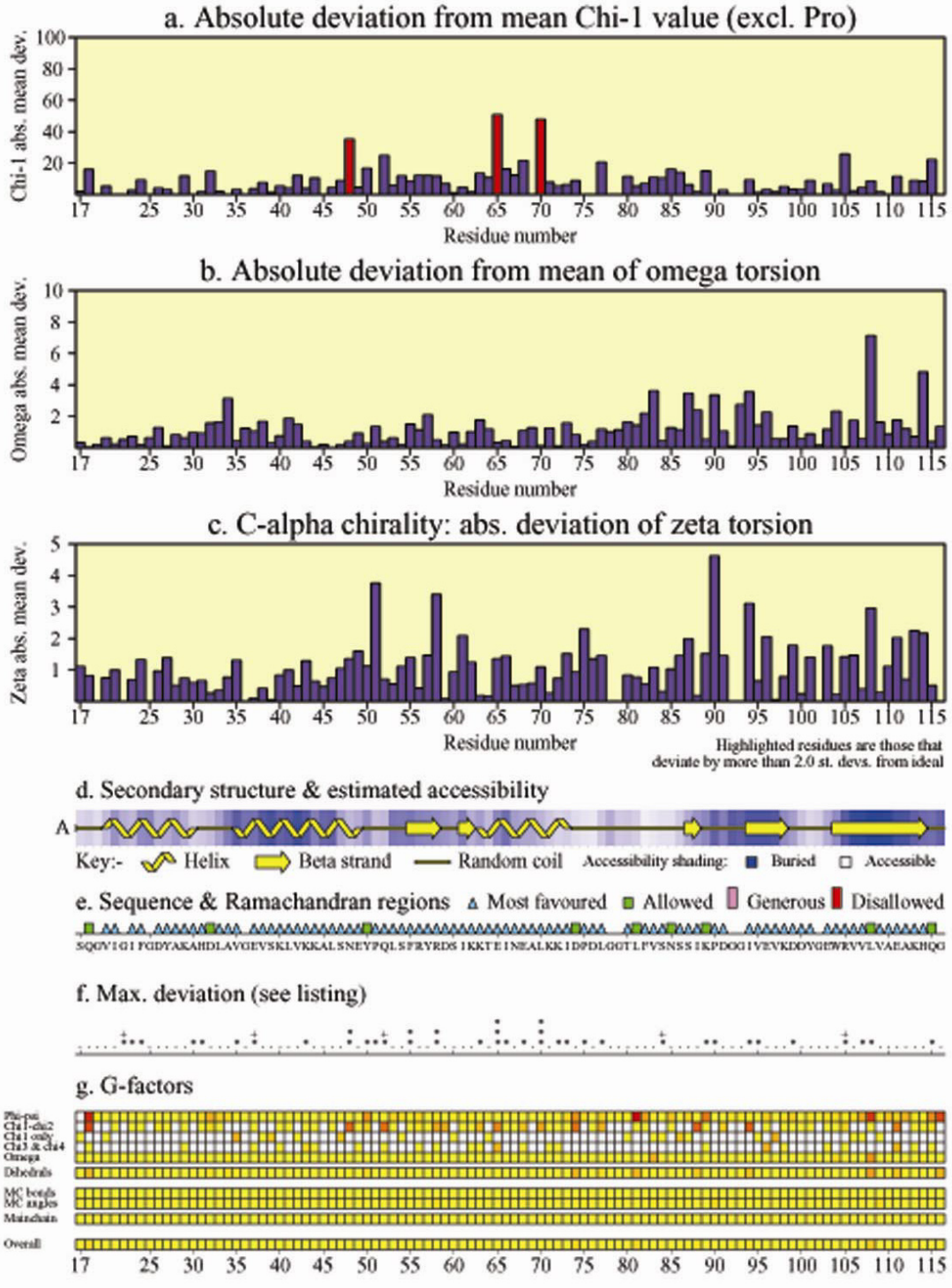
Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. Chi-1 gauche minus st dev	32	10.5	17.7	6.5	-1.1	BETTER
b. Chi-1 trans st dev	58	13.2	18.6	5.3	-1.0	BETTER
c. Chi-1 gauche plus st dev	130	12.4	17.2	4.9	-1.0	Inside
d. Chi-1 pooled st dev	220	12.6	17.8	4.8	-1.1	BETTER
e. Chi-2 trans st dev	56	10.6	20.1	5.0	-1.9	BETTER

Final_A138T_05.ps

Residue properties

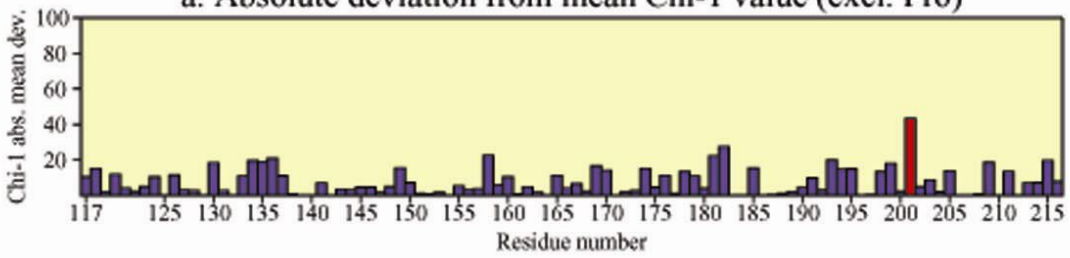
Final_A138T



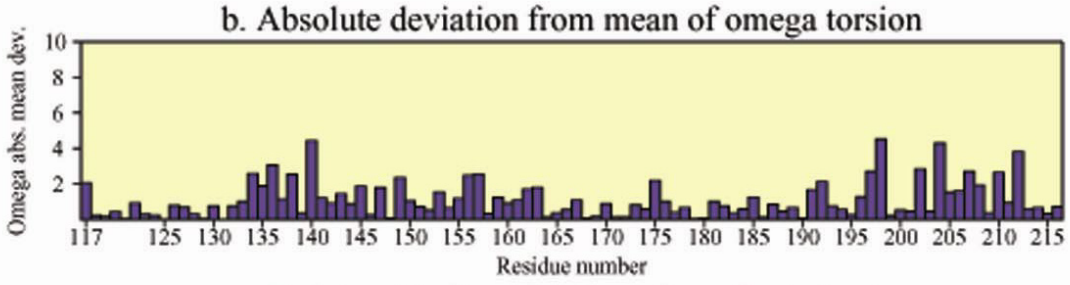
Final_A138T_06.ps

Residue properties
Final_A138T

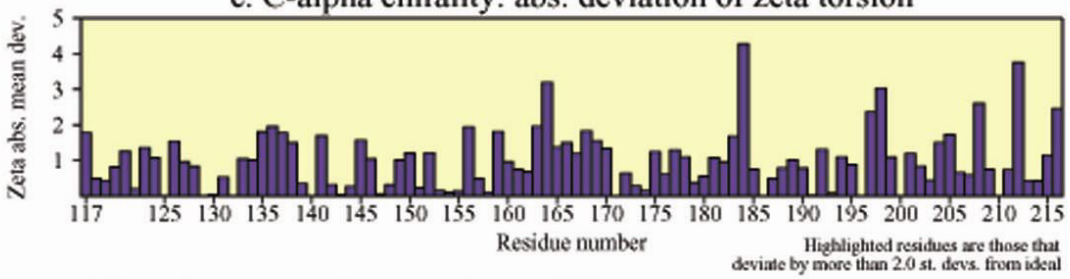
a. Absolute deviation from mean Chi-1 value (excl. Pro)



b. Absolute deviation from mean of omega torsion



c. C-alpha chirality: abs. deviation of zeta torsion



Highlighted residues are those that deviate by more than 2.0 st. devs. from ideal

d. Secondary structure & estimated accessibility

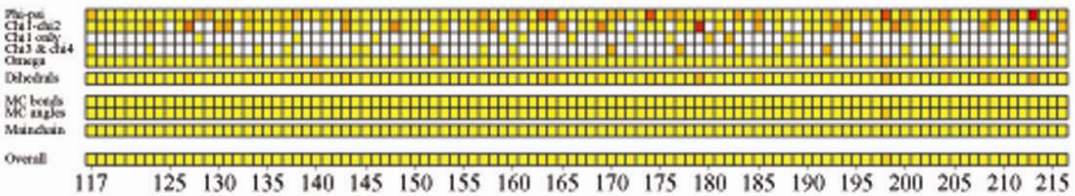
Key:- Helix Beta strand Random coil Accessibility shading: Buried Accessible

e. Sequence & Ramachandran regions Most favoured Allowed Generous Disallowed

KD I IN I RNGI L VGKRGDQDDLMTAGNA I ERSHKN I SE I ANFML SESHF PYVL FLEGSNFL T EN I S I TRPDGRVVNL EYNSG I L NRLDRL T AANYGMP I NSN

f. Max. deviation (see listing)

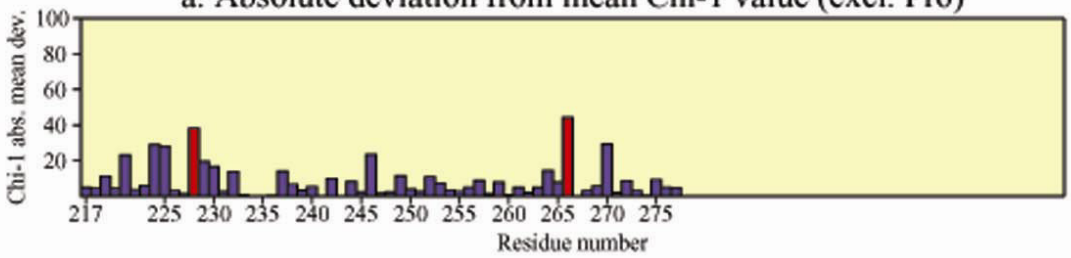
g. G-factors



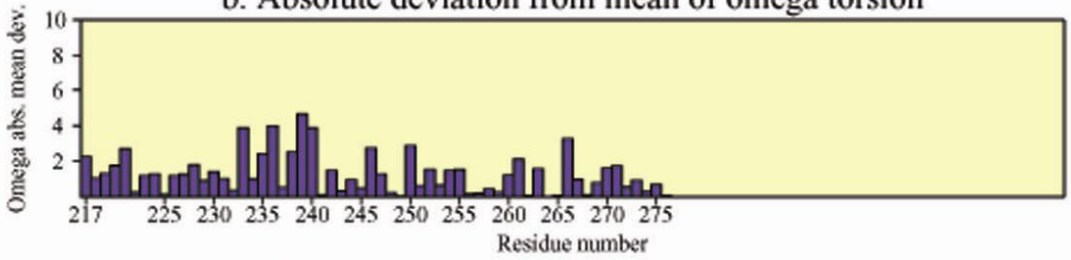
Final_A138T_06.ps

Residue properties
Final_A138T

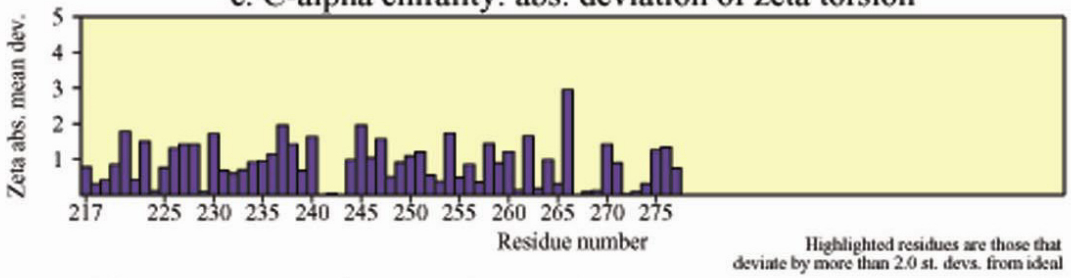
a. Absolute deviation from mean Chi-1 value (excl. Pro)



b. Absolute deviation from mean of omega torsion

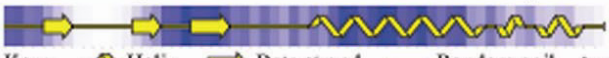


c. C-alpha chirality: abs. deviation of zeta torsion



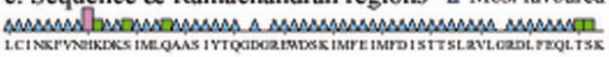
Highlighted residues are those that deviate by more than 2.0 st. devs. from ideal

d. Secondary structure & estimated accessibility



Key:- Helix Beta strand Random coil Accessibility shading: Buried Accessible

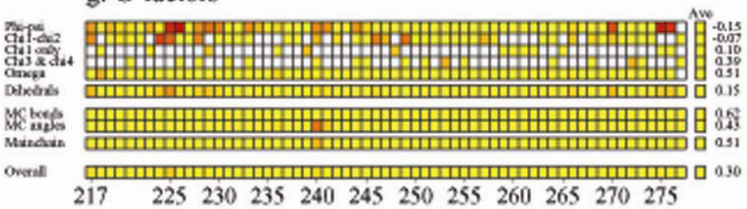
e. Sequence & Ramachandran regions Most favoured Allowed Generous Disallowed



LC I NKPVNHKDKS IMLQAAS I YTQCDGRIWDSK IMPE IMPD I STTSLRVLGRDLFEQLTSK

f. Max. deviation (see listing)

g. G-factors



Final_A138T_06.ps

APPENDIX F

LOCAL DNA BASE-PAIR AND BASE-STEP PARAMETERS OBTAINED FROM THE CRYSTAL STRUCTURES OF THE WILD-TYPE AND A138T *Eco*RI-d(TCGCGAATTCGCG) COMPLEXES

The program 3DNA (Lu and Olson, 2003) was used to extract the local base-pair and base-step parameters from the DNAs observed in the wild-type (Grigorescu, 2003) and A138T-d(TCGCGAATTCGCG) crystal structures. 3DNA uses the base-centered reference frame that was proposed (Olson et al., 2001) as a standard for the description of nucleic acid structures. The definitions and nomenclature for DNA structural parameters resultant from the 1988 EMBO Workshop on DNA Curvature and Bending (Dickerson, 1989) are shown in figure F.1. These parameters for the DNA base-pairs and base-steps in the wild-type and A138T structures are given in Figures F.2, F.3, and F.4. In each of these plots, parameters extracted from the bound DNAs are compared with those for averaged B-DNA (B-DNA crystal structures of 2.0 Å or better resolution without chemical modification, mismatches, drugs or proteins (Olson et al., 2001) from the Nucleic Acid Database (Berman et al., 2002)). Backbone parameters (all obtained using 3DNA) for the bound DNA molecules are given in table F.1.

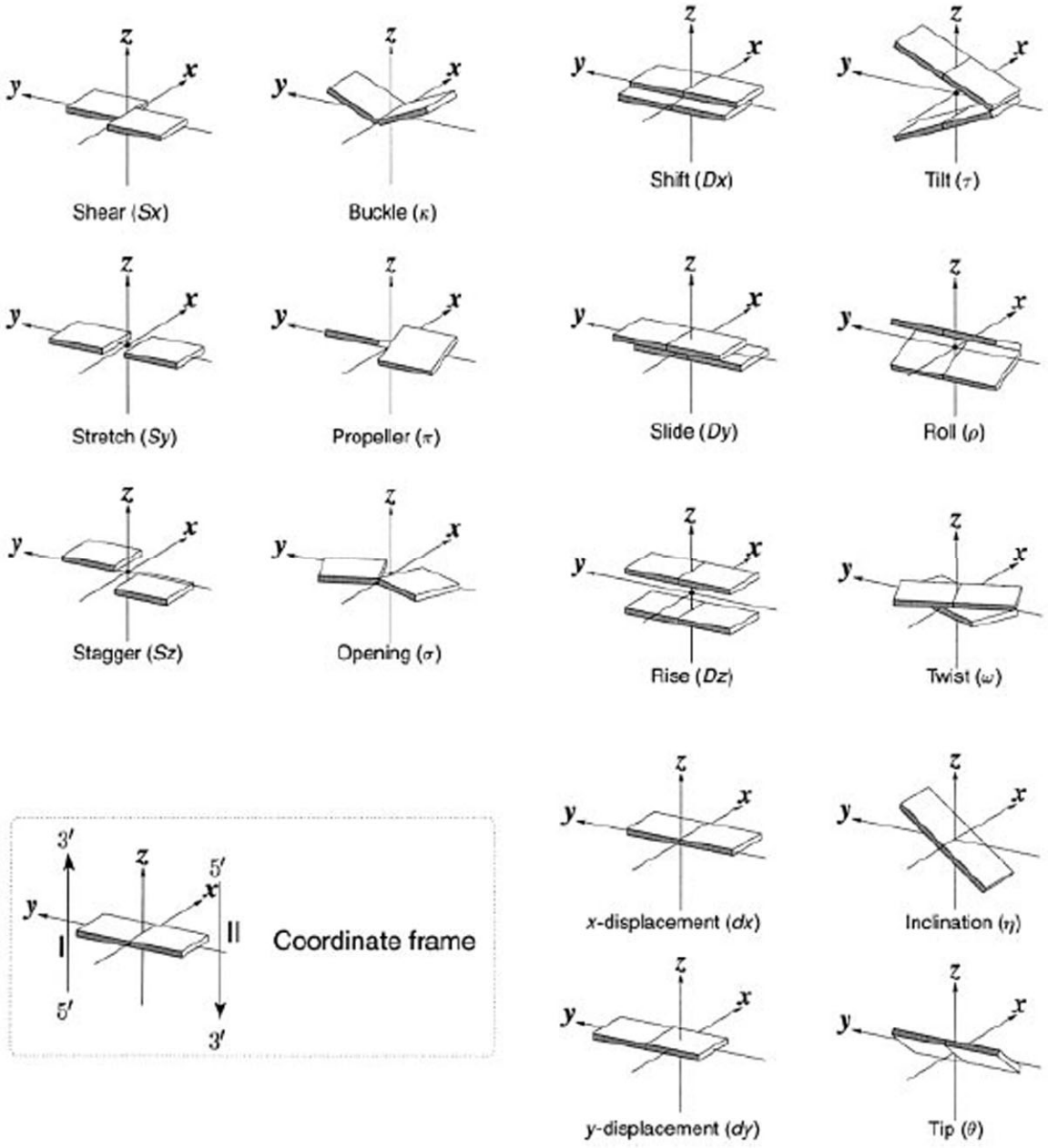


Figure F.1 Definitions and nomenclature for local DNA structural parameters. The base pair reference frame (lower left) is constructed such that the x-axis points away from the (shaded) minor groove edge of a base or base pair and the y-axis points toward strand (I).

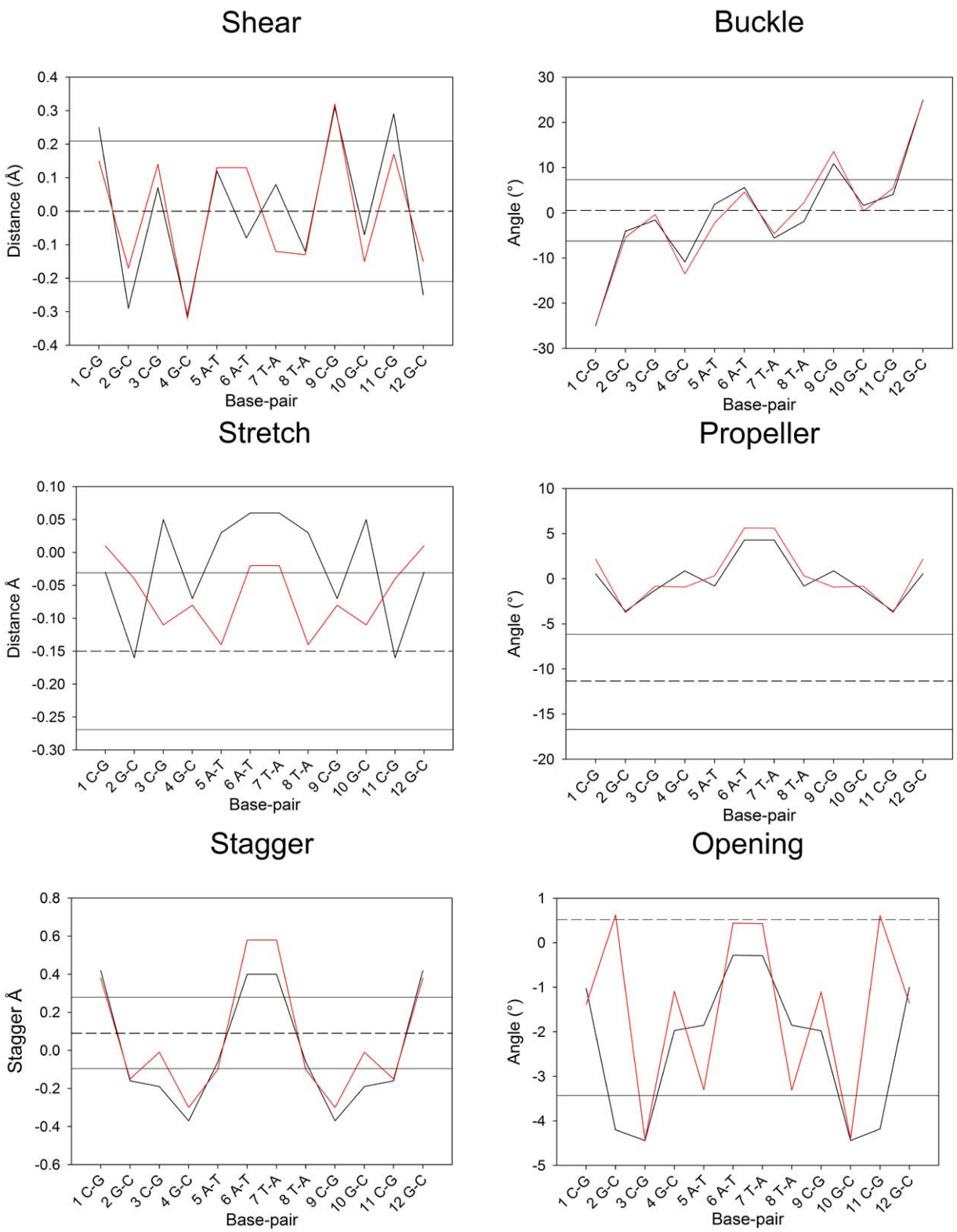


Figure F.2 Local DNA base-pair parameters for d(TCGCGAATTCGCG) bound in structures of the wild-type and A138T *Eco*RI specific complexes. Values were determined using the computer program, 3DNA. Data for the DNA bound to the wild-type are black, and A138T data are red. Mean and ± standard deviation values for base-pairs in standard B DNA (Olson et al., 2001) are shown as solid and dotted lines respectively.

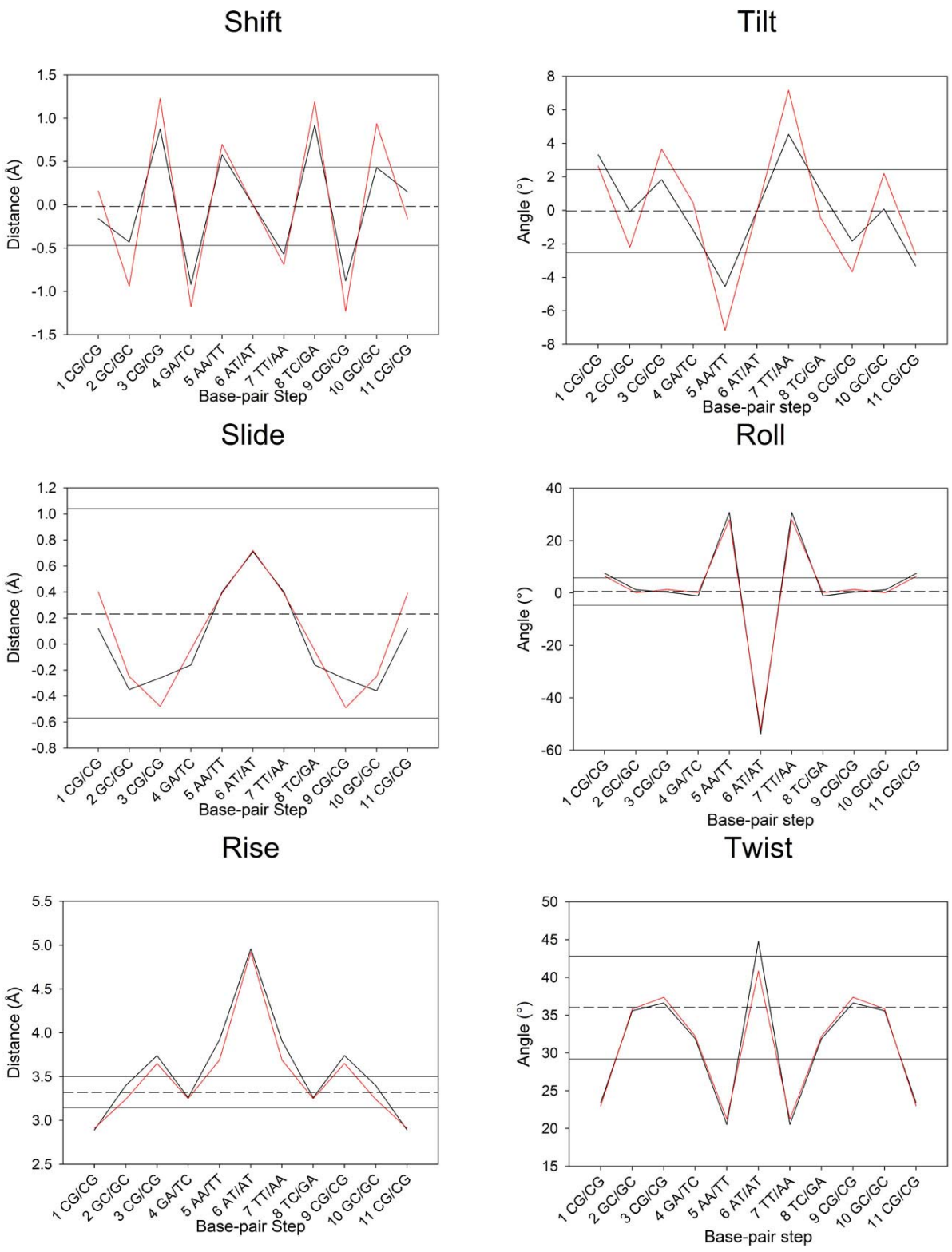


Figure F.3 Local DNA base-step parameters for d(TCGCGAATTCGCG) bound in structures of the wild-type and A138T *Eco*RI specific complexes. Values were determined using the computer program, 3DNA. Data for the DNA bound to the wild-type are black, and A138T data are red. Mean and ± standard deviation values for base-pairs in standard B DNA (Olson et al., 2001) are shown as solid and dotted lines respectively.

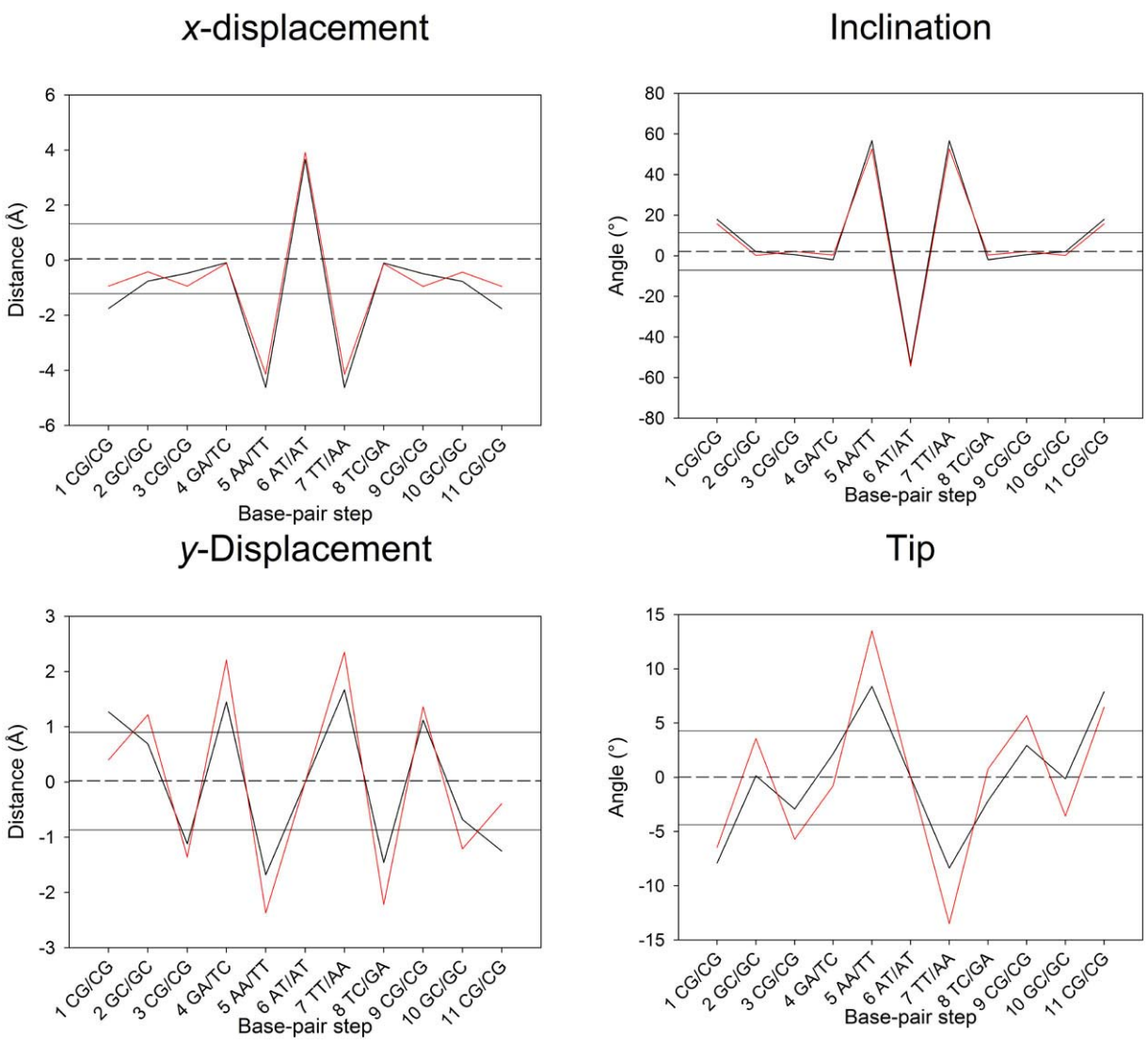


Figure F.4 Local DNA base-step parameters for d(TCGCGAATTCGCG) bound in structures of the wild-type and A138T *Eco*RI specific complexes continued. Values were determined using the computer program, 3DNA. Data for the DNA bound to the wild-type are black, and A138T data are red. Mean and ± standard deviation values for base-pairs in standard B DNA (Olson et al., 2001) are shown as solid and dotted lines respectively.

Table F.1 Torsion and phase angles for DNA backbone of nucleic acids in specific complexes of the wild-type and A138T *Eco*RI endonucleases

Torsions (°)	χ C1'-N	γ C5'-C4'	δ C4'-C3'	ε C3'-O3'	ζ O3'-P	α P-O5'	β O5'-C5'	Phase	Pucker
C	-105.7	51.7	139.3	-162.8	-102.6	-	-178.4	163.4	C2'-endo
	-106.3	57.2	141.9	-160.4	-96.8	-	177.6	168.4	C2'-endo
G	-112.4	53.3	119.0	168.1	-112.4	-66.8	165.8	132.4	C1'-exo
	-99.4	49.5	119.3	171.2	-121.3	-64.5	164.8	133.2	C1'-exo
C	-111.8	51.7	138.4	175.3	-96.2	-56.7	-171.2	161.2	C2'-endo
	-110.8	49.7	139.9	174.0	-92.7	-59.3	-171.5	167.3	C2'-endo
G	-122.1	50.6	132.1	-174.6	-113.3	-59.3	-177.1	150.4	C2'-endo
	-128.9	51.2	130.9	-169.3	-115.6	-58.1	176.2	142.5	C1'-exo
A	-107.6	45.4	140.8	-150.1	-74.2	-55.0	-175.8	164.6	C2'-endo
	-102.9	44.7	144.7	-153.9	-77.2	-58.2	-170.1	172.0	C2'-endo
A	-91.9	56.8	130.4	-79.0	128.2	-90.5	171.4	140.0	C1'-exo
	-90.4	44.9	131.8	-80.3	124.7	-79.8	170.7	139.8	C1-exo
T	-131.4	33.3	103.3	-155.0	-100.9	-73.4	145.3	78.2	O4'-exo
	-133.4	31.1	102.9	-155.8	-103.5	-72.4	143.0	74.4	O4'-endo
T	-105.4	51.5	139.6	-173.5	-95.6	-52.8	179.8	165.5	C2'-endo
	-101.8	48.2	140.9	-175.1	-93.8	-47.5	178.2	172.6	C2'-endo
C	-106.7	49.1	127.2	-157.3	-142.9	-49.1	168.4	144.0	C2'-endo
	-100.2	50.2	144.7	-125.1	178.2	-56.5	165.2	151.8	C2'-endo
G	-115.7	43.8	142.0	-172.9	-97.3	-50.9	168.2	170.9	C2'-endo
	-120.5	36.1	140.1	-172.1	-86.5	-53.2	154.2	170.0	C2'-endo
C	-108.3	35.7	128.1	-158.2	-104.7	-48.7	176.3	142.3	C1'-exo
	-107.0	32.4	121.3	-159.8	-108.1	-53.1	178.5	131.5	C1'-exo
G	-85.3	45.6	146	-	-	-67.1	171.6	154.3	C2'-eno
	-82.6	43.4	146.2	-	-	-58.8	169.0	156.2	C2'-endo

BIBLIOGRAPHY

1. Adams, P.D., Pannu, N.S., Read, R.J. & Brunger, A.T. (1997). Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc Natl Acad Sci U S A* **94**, pp. 5018-5023.

2. Albery, W.J. & Knowles, J.R. (1976). Evolution of enzyme function and the development of catalytic efficiency. *Biochemistry* **15**, pp. 5631-5640.

3. Albright, R.A. & Matthews, B.W. (1998). How Cro and lambda-repressor distinguish between operators: the structural basis underlying a genetic switch. *Proc Natl Acad Sci U S A* **95**, pp. 3431-3436.

4. Albright, R.A., Mossing, M.C. & Matthews, B.W. (1998). Crystal structure of an engineered Cro monomer bound nonspecifically to DNA: possible implications for nonspecific binding by the wild-type protein. *Protein Sci* **7**, pp. 1485-1494.

5. Altenbach, C., Oh, K.J., Trabanino, R.J., Hideg, K. & Hubbell, W.L. (2001). Estimation of inter-residue distances in spin labeled proteins at physiological temperatures: experimental strategies and practical limitations. *Biochemistry* **40**, pp. 15471-15482.

6. Alves, J., Ruter, T., Geiger, R., Fliess, A., Maass, G. & Pingoud, A. (1989). Changing the hydrogen-bonding potential in the DNA binding site of EcoRI by site-directed mutagenesis drastically reduces the enzymatic activity, not, however, the preference of this restriction endonuclease for cleavage within the site-GAATTC. *Biochemistry* **28**, pp. 2678-2684.

7. Arber, W. (1965). Host Specificity of DNA Produced by Escherichia Coli V . The Role of Methionine in the Production of Host Specificity. *J Mol Biol* **11**, pp. 247-256.

8. Arber, W. (1979). Promotion and limitation of genetic exchange. *Science* **205**, pp. 361-365.

9. Berg, O.G., Winter, R.B. & von Hippel, P.H. (1981). Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, pp. 6929-6948.

10. Bergqvist, S., Williams, M.A., O'Brien, R. & Ladbury, J.E. (2004). Heat capacity effects of water molecules and ions at a protein-DNA interface. *J Mol Biol* **336**, pp. 829-842.

11. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. & Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Res* **28**, pp. 235-242.

12. Berman, H.M., Westbrook, J., Feng, Z., Iype, L., Schneider, B. & Zardecki, C. (2002). The Nucleic Acid Database. *Acta Crystallogr D Biol Crystallogr* **58**, pp. 889-898.

13. Bertani, G. & Weigle, J.J. (1953). Host controlled variation in bacterial viruses. *J Bacteriol* **65**, pp. 113-121.

14. Betlach, M., Hershfield, V., Chow, L., Brown, W., Goodman, H. & Boyer, H.W. (1976). A restriction endonuclease analysis of the bacterial plasmid controlling the EcoRI restriction and modification of DNA. *Fed Proc* **35**, pp. 2037-2043.

15. Birge, E.A. (1994). *Bacterial and bacteriophage genetics*. Springer-Verlag: New York.

16. Brünger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, pp. 472-475.

17. Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. & Warren, G.L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54 (Pt 5)**, pp. 905-921.

18. Brünger, A.T., Kuriyan, J., Karplus, M. (1987). Crystallographic R factor Refinement by Molecular Dynamics. *Science* **235**, pp. 458-460.

19. Bujnicki, J.M. (2003). Crystallographic and bioinformatic studies on restriction endonucleases: inference of evolutionary relationships in the "midnight zone" of homology. *Curr Protein Pept Sci* **4**, pp. 327-337.

20. Campbell, A. (1996). Bacteriophages. In *Escherichia coli and Salmonella : cellular and molecular biology* (Neidhardt, F.C. & Curtiss, R., eds.). pp. 2325-2338, ASM Press: Washington, D.C.

21. Cao, D.F. (2001). *Co-solute effects as probes of the role of water release in DNA binding and catalysis by three restriction endonucleases* (Ph.D. Thesis) Biological Sciences, University of Pittsburgh: Pittsburgh, PA.

22. Cheatham, T.E., 3rd & Young, M.A. (2000). Molecular dynamics simulation of nucleic acids: successes, limitations, and promise. *Biopolymers* **56**, pp. 232-256.

23. Cheng, S.C., Kim, R., King, K., Kim, S.H. & Modrich, P. (1984). Isolation of gram quantities of EcoRI restriction and modification enzymes from an overproducing strain. *J Biol Chem* **259**, pp. 11571-11575.

24. Christensen, J.J., Hansen, L.D. & Izatt, R.M. (1976). *Handbook of proton ionization heats and related thermodynamic quantities*. Wiley: New York.

25. Clore, G.M., Gronenborn, A.M. & Davies, R.W. (1982). Theoretical aspects of specific and non-specific equilibrium binding of proteins to DNA as studied by the nitrocellulose filter binding assay. Co-operative and non-co-operative binding to a one-dimensional lattice. *J Mol Biol* **155**, pp. 447-466.

26. Columbus, L. & Hubbell, W.L. (2002). A new spin on protein dynamics. *Trends Biochem Sci* **27**, pp. 288-295.

27. Connelly, P.R. (1997). The cost of releasing site-specific, bound water molecules from proteins: Toward a quantitative guide for structure-based drug design. In *Structure Based Drug Design: Thermodynamics, Modeling and Strategy* (Ladbury, J.E., and Connelly, P.R., ed.), vol. Chapter 5. pp. 143-159, Springer-Verlag: New York.

28. Cooper, A. (2000). Heat capacity of hydrogen-bonded networks: an alternative view of protein folding thermodynamics. *Biophys Chem* **85**, pp. 25-39.

29. Dauter, Z. (1999). Data-collection strategies. *Acta Crystallogr D Biol Crystallogr* **55 (Pt 10)**, pp. 1703-1717.

30. Deibert, M., Grazulis, S., Janulaitis, A., Siksnys, V. & Huber, R. (1999). Crystal structure of MunI restriction endonuclease in complex with cognate DNA at 1.7 A resolution. *Embo J* **18**, pp. 5805-5816.

31. Deng, W.P. & Nickoloff, J.A. (1992). Site-directed mutagenesis of virtually any plasmid by eliminating a unique site. *Anal Biochem* **200**, pp. 81-88.

32. DePristo, M.A., de Bakker, P.I. & Blundell, T.L. (2004). Heterogeneity and inaccuracy in protein structures solved by x-ray crystallography. *Structure (Camb)* **12**, pp. 831-838.

33. Deva, T. & Krishnaswamy, S. (2001). Structure-based sequence alignment of type-II restriction endonucleases. *Biochim Biophys Acta* **1544**, pp. 217-228.

34. Dickerson, R.E. (1989). Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res* **17**, pp. 1797-1803.

35. Dickerson, R.E. & Chiu, T.K. (1997). Helix bending as a factor in protein/DNA recognition. *Biopolymers* **44**, pp. 361-403.

36. Doig, A.J. & Sternberg, M.J. (1995). Side-chain conformational entropy in protein folding. *Protein Sci* **4**, pp. 2247-2251.

37. Dorner, L.F., Bitinaite, J., Whitaker, R.D. & Schildkraut, I. (1999). Genetic analysis of the base-specific contacts of BamHI restriction endonuclease. *J Mol Biol* **285**, pp. 1515-1523.

38. Duan, Y., Wilkosz, P. & Rosenberg, J.M. (1996). Dynamic contributions to the DNA binding entropy of the EcoRI and EcoRV restriction endonucleases. *J Mol Biol* **264**, pp. 546-555.

39. Dubendorff, J.W., deHaseth, P.L., Rosendahl, M.S. & Caruthers, M.H. (1987). DNA functional groups required for formation of open complexes between Escherichia coli RNA polymerase and the lambda PR promoter. Identification via base analog substitutions. *J Biol Chem* **262**, pp. 892-898.

40. Dunitz, J.D. (1994). The Entropic Cost of Bound Water in Crystals and Biomolecules. *Science* **264**, pp. 670.

41. Dunitz, J.D. (1995). Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions. *Chem Biol* **2**, pp. 709-712.

42. Eftink, M.R., Anusiem, A.C. & Biltonen, R.L. (1983). Enthalpy-entropy compensation and heat capacity changes for protein-ligand interactions: general thermodynamic models and data for the binding of nucleotides to ribonuclease A. *Biochemistry* **22**, pp. 3884-3896.

43. Ehbrecht, H.J., Pingoud, A., Urbanke, C., Maass, G. & Gualerzi, C. (1985). Linear diffusion of restriction endonucleases on DNA. *J Biol Chem* **260**, pp. 6160-6166.

44. Engh, R.A., Huber, R. (1991). Accurate bond and angle parameters for X-ray structure refinement. *Acta Cryst. A* **47**, pp. 392-400.

45. Engler, L.E. (1998). *Specificity determinants in the BamHI-DNA interaction* (Ph.D. Thesis) Biological Sciences, University of Pittsburgh: Pittsburgh, PA.

46. Engler, L.E., Sapienza, P., Dorner, L.F., Kucera, R., Schildkraut, I. & Jen-Jacobson, L. (2001). The energetics of the interaction of BamHI endonuclease with its recognition site GGATCC. *J Mol Biol* **307**, pp. 619-636.

47. Engler, L.E., Welch, K.K. & Jen-Jacobson, L. (1997). Specific binding by EcoRV endonuclease to its DNA recognition site GATATC. *J Mol Biol* **269**, pp. 82-101.

48. Etzkorn, C. & Horton, N.C. (2004). Ca2+ binding in the active site of HincII: implications for the catalytic mechanism. *Biochemistry* **43**, pp. 13256-13270.

49. Ferrin, T.E., Huang, C. C. Jarvis, L. E., & and Langridge, R. (1988). The MIDAS display system'. *J. Mol. Graphics* **6**, pp. 13-27.

50. Fersht, A. (1985). Enzyme structure and mechanism. pp. Ch. 2, 3, 12, W.H. Freeman: New York.

51. Flores, H., Osuna, J., Heitman, J. & Soberon, X. (1995). Saturation mutagenesis of His114 of EcoRI reveals relaxed-specificity mutants. *Gene* **157**, pp. 295-301.

52. Frank, D.E., Saecker, R.M., Bond, J.P., Capp, M.W., Tsodikov, O.V., Melcher, S.E., Levandoski, M.M. & Record, M.T., Jr. (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. *J Mol Biol* **267**, pp. 1186-1206.

53. Frauenfelder, H., Sligar, S.G. & Wolynes, P.G. (1991). The energy landscapes and motions of proteins. *Science* **254**, pp. 1598-1603.

54. Frederick, C.A., Grable, J., Melia, M., Samudzi, C., Jen-Jacobson, L., Wang, B.C., Greene, P., Boyer, H.W. & Rosenberg, J.M. (1984). Kinked DNA in crystalline complex with EcoRI endonuclease. *Nature* **309**, pp. 327-331.

55. Freire, E. (1998). Statistical thermodynamic linkage between conformational and binding equilibria. *Adv Protein Chem* **51**, pp. 255-279.

56. Fritz, A., Kuster, W. & Alves, J. (1998). Asn141 is essential for DNA recognition by EcoRI restriction endonuclease. *FEBS Lett* **438**, pp. 66-70.

57. Giudice, E. & Lavery, R. (2003). Nucleic acid base pair dynamics: the impact of sequence and structure using free-energy calculations. *J Am Chem Soc* **125**, pp. 4998-4999.

58. Goeddel, D.V., Yansura, D.G. & Caruthers, M.H. (1977). Studies on gene control regions. VI. The 5- methyl of thymine, a lac repressor recognition site. *Nucleic Acids Res* **4**, pp. 3039-3054.

59. Grabowski, G., Jeltsch, A., Wolfes, H., Maass, G. & Alves, J. (1995). Site-directed mutagenesis in the catalytic center of the restriction endonuclease EcoRI. *Gene* **157**, pp. 113-118.

60. Grigorescu, A. (2003). *Structural and Energetic Determinants of the Binding Specificity of EcoRI Endonuclease* (Ph.D. Thesis) Molecular Biophysics, University of Pittsburgh: Pittsburgh, PA.

61. Grigorescu, A., Horvath, M., Wilkosz, P.A., Chandrasekhar, K., Rosenberg, J.M. (2003). The integration of recognition and cleavage: X-ray structures of pre-transition state complex, post reactive complex and the DNA-free endonuclease. In *Restriction Endonucleases* (Pingoud, A.M., ed.). pp. 137-173, Springer: Berlin.

62. Gubler, M., Braguglia, D., Meyer, J., Piekarowicz, A. & Bickle, T.A. (1992). Recombination of constant and variable modules alters DNA sequence recognition by type IC restriction-modification enzymes. *Embo J* **11**, pp. 233-240.

63. Ha, J.H., Spolar, R.S. & Record, M.T., Jr. (1989). Role of the hydrophobic effect in stability of site-specific protein-DNA complexes. *J Mol Biol* **209**, pp. 801-816.

64. Hager, P.W., Reich, N.O., Day, J.P., Coche, T.G., Boyer, H.W., Rosenberg, J.M. & Greene, P.J. (1990). Probing the role of glutamic acid 144 in the EcoRI endonuclease using aspartic acid and glutamine replacements. *J Biol Chem* **265**, pp. 21520-21526.

65. Hagerman, P.J. (1990). Pyrimidine 5-methyl groups influence the magnitude of DNA curvature. *Biochemistry* **29**, pp. 1980-1983.

66. Halford, S.E. & Johnson, N.P. (1983). Single turnovers of the EcoRI restriction endonuclease. *Biochem J* **211**, pp. 405-415.

67. Handa, N., Ichige, A., Kusano, K. & Kobayashi, I. (2000). Cellular responses to postsegregational killing by restriction-modification genes. *J Bacteriol* **182**, pp. 2218-2229.

68. Handa, N. & Kobayashi, I. (1999). Post-segregational killing by restriction modification gene complexes: observations of individual cell deaths. *Biochimie* **81**, pp. 931-938.

69. Harrison, B. & Zimmerman, S.B. (1986). T4 polynucleotide kinase: macromolecular crowding increases the efficiency of reaction at DNA termini. *Anal Biochem* **158**, pp. 307-315.

70. Harrison, S.C. & Aggarwal, A.K. (1990). DNA recognition by proteins with the helix-turn-helix motif. *Annu Rev Biochem* **59**, pp. 933-969.

71. Hassur, S.M. & Whitlock, H.W., Jr. (1974). UV shadowing--a new and convenient method for the location of ultraviolet-absorbing species in polyacrylamide gels. *Anal Biochem* **59**, pp. 162-164.

72. Heinig, M. & Frishman, D. (2004). STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res* **32**, pp. W500-502.

73. Heitman, J. (1993). On the origins, structures and functions of restriction-modification enzymes. *Genet Eng (N Y)* **15**, pp. 57-108.

74. Heitman, J., Ivanenko, T. & Kiss, A. (1999). DNA nicks inflicted by restriction endonucleases are repaired by a RecA- and RecB-dependent pathway in Escherichia coli. *Mol Microbiol* **33**, pp. 1141-1151.

75. Heitman, J. & Model, P. (1990). Mutants of the EcoRI endonuclease with promiscuous substrate specificity implicate residues involved in substrate recognition. *Embo J* **9**, pp. 3369-3378.

76. Heitman, J. & Model, P. (1987). Site-specific methylases induce the SOS DNA repair response in Escherichia coli. *J Bacteriol* **169**, pp. 3243-3250.

77. Heitman, J. & Model, P. (1990). Substrate recognition by the EcoRI endonuclease. *Proteins* **7**, pp. 185-197.

78. Heitman, J., Zinder, N.D. & Model, P. (1989). Repair of the Escherichia coli chromosome after in vivo scission by the EcoRI endonuclease. *Proc Natl Acad Sci U S A* **86**, pp. 2281-2285.

79. Holbrook, J.A., Tsodikov, O.V., Saecker, R.M. & Record, M.T., Jr. (2001). Specific and non-specific interactions of integration host factor with DNA: thermodynamic evidence for disruption of multiple IHF surface salt-bridges coupled to DNA binding. *J Mol Biol* **310**, pp. 379-401.

80. Holmbeck, S.M., Dyson, H.J. & Wright, P.E. (1998). DNA-induced conformational changes are the basis for cooperative dimerization by the DNA binding domain of the retinoid X receptor. *J Mol Biol* **284**, pp. 533-539.

81. Hooft, R.W., Vriend, G., Sander, C. & Abola, E.E. (1996). Errors in protein structures. *Nature* **381**, pp. 272.

82. Horn, J.R., Brandts, J.F. & Murphy, K.P. (2002). van't Hoff and calorimetric enthalpies II: effects of linked equilibria. *Biochemistry* **41**, pp. 7501-7507.

83. Horton, J.R. & Cheng, X. (2000). PvuII endonuclease contains two calcium ions in active sites. *J Mol Biol* **300**, pp. 1049-1056.

84. Horton, N.C., Newberry, K.J. & Perona, J.J. (1998). Metal ion-mediated substrate-assisted catalysis in type II restriction endonucleases. *Proc Natl Acad Sci U S A* **95**, pp. 13489-13494.

85. Horton, N.C. & Perona, J.J. (1998). Recognition of flanking DNA sequences by EcoRV endonuclease involves alternative patterns of water-mediated contacts. *J Biol Chem* **273**, pp. 21721-21729.

86. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: visual molecular dynamics. *J Mol Graph* **14**, pp. 33-38, 27-38.

87. Ivanenko, T., Heitman, J. & Kiss, A. (1998). Mutational analysis of the function of Met137 and Ile197, two amino acids implicated in sequence-specific DNA recognition by the EcoRI endonuclease. *Biol Chem* **379**, pp. 459-465.

88. Jack, W.E., Terry, B.J. & Modrich, P. (1982). Involvement of outside DNA sequences in the major kinetic path by which EcoRI endonuclease locates and leaves its recognition sequence. *Proc Natl Acad Sci U S A* **79**, pp. 4010-4014.

89. Janin, J. (1999). Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition. *Structure Fold Des* **7**, pp. R277-279.

90. Jeltsch, A., Alves, J., Wolfes, H., Maass, G. & Pingoud, A. (1994). Pausing of the restriction endonuclease EcoRI during linear diffusion on DNA. *Biochemistry* **33**, pp. 10215-10219.

91. Jeltsch, A. & Pingoud, A. (1996). Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. *J Mol Evol* **42**, pp. 91-96.

92. Jen-Jacobson, L. (1997). Protein-DNA recognition complexes: conservation of structure and binding energy in the transition state. *Biopolymers* **44**, pp. 153-180.

93. Jen-Jacobson, L. (1995). Structural-perturbation approaches to thermodynamics of site-specific protein-DNA interactions. *Methods Enzymol* **259**, pp. 305-344.

94. Jen-Jacobson, L., Engler, L.E., Lesser, D.R., Kurpiewski, M.R., Yee, C. & McVerry, B. (1996). Structural adaptations in the interaction of EcoRI endonuclease with methylated GAATTC sites. *Embo J* **15**, pp. 2870-2882.

95. Jen-Jacobson, L., Kurpiewski, M., Lesser, D., Grable, J., Boyer, H.W., Rosenberg, J.M. & Greene, P.J. (1983). Coordinate ion pair formation between EcoRI endonuclease and DNA. *J Biol Chem* **258**, pp. 14638-14646.

96. Jen-Jacobson, L., L.E. Engler, J.T. Ames, M.R. Kurpiewski, and A. Grigorescu (2000). Thermodynamic parameters of specific and nonspecific protein-DNA binding. *Supramolecular Chemistry* **12**, pp. 143-160.

97. Jen-Jacobson, L., Lesser, D. & Kurpiewski, M. (1986). The enfolding arms of EcoRI endonuclease: role in DNA binding and cleavage. *Cell* **45**, pp. 619-629.

98. Johnson, M.S., Sutcliffe, M.J. & Blundell, T.L. (1990). Molecular anatomy: phyletic relationships derived from three-dimensional structures of proteins. *J Mol Evol* **30**, pp. 43-59.

99. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* **47 (Pt 2)**, pp. 110-119.

100. Kalodimos, C.G., Biris, N., Bonvin, A.M., Levandoski, M.M., Guennuegues, M., Boelens, R. & Kaptein, R. (2004). Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science* **305**, pp. 386-389.

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

102. Kim, Y., Coi, J., Grable, J.C., Greene, P., Hager, P., and Rosenberg, J.M. (1994). Studies on the canonical DNA-EcoRI endonuclease complex and the EcoRI kink. In *Structural biology : the state of the art : proceedings of the eighth Conversation, SUNY, Albany* (Sarma, R., and Sarma, M., ed.). pp. 225-246, Adenine Press: New York.

103. Kim, Y., Geiger, J.H., Hahn, S. & Sigler, P.B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, pp. 512-520.

104. Kim, Y., Grable, J.C., Greene, P.J. and Rosenberg, J.M. (1990). Refine of EcoRI endonuclease crystal structure: a revised chain tracing. *Science* **249**, pp. 1307-1309.

105. Kim, Y.C., Grable, J.C., Love, R., Greene, P.J. & Rosenberg, J.M. (1990). Refinement of Eco RI endonuclease crystal structure: a revised protein chain tracing. *Science* **249**, pp. 1307-1309.

106. King, K., Benkovic, S.J. & Modrich, P. (1989). Glu-111 is required for activation of the DNA cleavage center of EcoRI endonuclease. *J Biol Chem* **264**, pp. 11807-11815.

107. Kleywegt, G.J. & Jones, T.A. (1998). Databases in protein crystallography. *Acta Crystallogr D Biol Crystallogr* **54**, pp. 1119-1131.

108. Klingen, A.R. & Ullmann, G.M. (2004). Negatively charged residues and hydrogen bonds tune the ligand histidine pKa values of Rieske iron-sulfur proteins. *Biochemistry* **43**, pp. 12383-12389.

109. Kozlov, A.G. & Lohman, T.M. (1998). Calorimetric studies of E. coli SSB protein-single-stranded DNA interactions. Effects of monovalent salts on binding enthalpy. *J Mol Biol* **278**, pp. 999-1014.

110. Kozlov, A.G. & Lohman, T.M. (2000). Large contributions of coupled protonation equilibria to the observed enthalpy and heat capacity changes for ssDNA binding to Escherichia coli SSB protein. *Proteins* **Suppl 4**, pp. 8-22.

111. Kruger, D.H. & Bickle, T.A. (1983). Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol Rev* **47**, pp. 345-360.

112. Kuhn, A. & Kellenberger, E. (1985). Productive phage infection in Escherichia coli with reduced internal levels of the major cations. *J Bacteriol* **163**, pp. 906-912.

113. Kulakauskas, S., Lubys, A. & Ehrlich, S.D. (1995). DNA restriction-modification systems mediate plasmid maintenance. *J Bacteriol* **177**, pp. 3451-3454.

114. Kuriyan, J., Osapay, K., Burley, S.K., Brunger, A.T., Hendrickson, W.A. & Karplus, M. (1991). Exploration of disorder in protein structures by X-ray restrained molecular dynamics. *Proteins* **10**, pp. 340-358.

115. Kuriyan, J. & Weis, W.I. (1991). Rigid protein motion as a model for crystallographic temperature factors. *Proc Natl Acad Sci U S A* **88**, pp. 2773-2777.

116. Kurpiewski, M.R., Engler, L.E., Wozniak, L.A., Kobylanska, A., Koziolkiewicz, M., Stec, W.J. & Jen-Jacobson, L. (2004). Mechanisms of coupling between DNA recognition specificity and catalysis in EcoRI endonuclease. *Structure (Camb)* **12**, pp. 1775-1788.

117. Kurpiewski, M.R., Koziolkiewicz, M., Wilk, A., Stec, W.J. & Jen-Jacobson, L. (1996). Chiral phosphorothioates as probes of protein interactions with individual DNA phosphoryl oxygens: essential interactions of EcoRI endonuclease with the phosphate at pGAATTC. *Biochemistry* **35**, pp. 8846-8854.

118. Kusano, K., Naito, T., Handa, N. & Kobayashi, I. (1995). Restriction-modification systems as genomic parasites in competition for specific sequences. *Proc Natl Acad Sci U S A* **92**, pp. 11095-11099.

119. Ladbury, J.E., Wright, J.G., Sturtevant, J.M. & Sigler, P.B. (1994). A thermodynamic study of the trp repressor-operator interaction. *J Mol Biol* **238**, pp. 669-681.

120. Lankas, F., Sponer, J., Langowski, J. & Cheatham, T.E., 3rd (2003). DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys J* **85**, pp. 2872-2883.

121. Laskowski, R.A., MacArthur, M. W., Moss, D. S. & Thornton J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, pp. 283-291.

122. Lesser, D.R., Grajkowski, A., Kurpiewski, M.R., Koziolkiewicz, M., Stec, W.J. & Jen-Jacobson, L. (1992). Stereoselective interaction with chiral phosphorothioates at the central DNA kink of the EcoRI endonuclease-GAATTC complex. *J Biol Chem* **267**, pp. 24810-24818.

123. Lesser, D.R., Kurpiewski, M.R. & Jen-Jacobson, L. (1990). The energetic basis of specificity in the Eco RI endonuclease--DNA interaction. *Science* **250**, pp. 776-786.

124. Lesser, D.R., Kurpiewski, M.R., Waters, T., Connolly, B.A. & Jen-Jacobson, L. (1993). Facilitated distortion of the DNA site enhances EcoRI endonuclease-DNA recognition. *Proc Natl Acad Sci U S A* **90**, pp. 7548-7552.

125. Levin, B.R. (1993). The accessory genetic elements of bacteria: existence conditions and (co)evolution. *Curr Opin Genet Dev* **3**, pp. 849-854.

126. Lide, D.R. ed. (2004). *CRC Handbook of Chemistry and Physics*. CRC Press: Boca Raton, FL.

127. Lin, S.Y. & Riggs, A.D. (1972). Lac repressor binding to non-operator DNA: detailed studies and a comparison of eequilibrium and rate competition methods. *J Mol Biol* **72**, pp. 671-690.

128. Liu, W., Chen, Y., Watrob, H., Bartlett, S.G., Jen-Jacobson, L. & Barkley, M.D. (1998). N-termini of EcoRI restriction endonuclease dimer are in close proximity on the protein surface. *Biochemistry* **37**, pp. 15457-15465.

129. Lohman, T.M., DeHaseth, P.L. & Record, M.T., Jr. (1978). Analysis of ion concentration effects of the kinetics of protein-nucleic acid interactions. Application to lac repressor-operator interactions. *Biophys Chem* **8**, pp. 281-294.

130. Love, J.J., Li, X., Chung, J., Dyson, H.J. & Wright, P.E. (2004). The LEF-1 high-mobility group domain undergoes a disorder-to-order transition upon formation of a complex with cognate DNA. *Biochemistry* **43**, pp. 8725-8734.

131. Lu, X.J. & Olson, W.K. (2003). 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* **31**, pp. 5108-5121.

132. Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R. & Sigler, P.B. (1991). Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* **352**, pp. 497-505.

133. Lundback, T., van Den Berg, S. & Hard, T. (2000). Sequence-specific DNA binding by the glucocorticoid receptor DNA-binding domain is linked to a salt-dependent histidine protonation. *Biochemistry* **39**, pp. 8909-8916.

134. Luria, S.E. & Human, M.L. (1952). A nonhereditary, host-induced variation of bacterial viruses. *J Bacteriol* **64**, pp. 557-569.

135. 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. *Biophys Chem* **7**, pp. 95-102.

136. Manning, G.S. (1978). The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q Rev Biophys* **11**, pp. 179-246.

137. Martin, A.M., Sam, M.D., Reich, N.O. & Perona, J.J. (1999). Structural and energetic origins of indirect readout in site-specific DNA cleavage by a restriction endonuclease. *Nat Struct Biol* **6**, pp. 269-277.

138. McClarin, J.A., Frederick, C.A., Wang, B.C., Greene, P., Boyer, H.W., Grable, J. & Rosenberg, J.M. (1986). Structure of the DNA-Eco RI endonuclease recognition complex at 3 A resolution. *Science* **234**, pp. 1526-1541.

139. McDonald, I.K. & Thornton, J.M. (1994). Satisfying hydrogen bonding potential in proteins. *J Mol Biol* **238**, pp. 777-793.

140. Merabet, E. & Ackers, G.K. (1995). Calorimetric analysis of lambda cI repressor binding to DNA operator sites. *Biochemistry* **34**, pp. 8554-8563.

141. Minor, W., Tomchick, D. & Otwinowski, Z. (2000). Strategies for macromolecular synchrotron crystallography. *Structure Fold Des* **8**, pp. R105-110.

142. Misra, V.K., Hecht, J.L., Sharp, K.A., Friedman, R.A. & Honig, B. (1994). Salt effects on protein-DNA interactions. The lambda cI repressor and EcoRI endonuclease. *J Mol Biol* **238**, pp. 264-280.

143. Modrich, P. & Zabel, D. (1976). EcoRI endonuclease. Physical and catalytic properties of the homogenous enzyme. *J Biol Chem* **251**, pp. 5866-5874.

144. Muir, R.S., Flores, H., Zinder, N.D., Model, P., Soberon, X. & Heitman, J. (1997). Temperature-sensitive mutants of the EcoRI endonuclease. *J Mol Biol* **274**, pp. 722-737.

145. Murray, N.E. (2000). Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol Mol Biol Rev* **64**, pp. 412-434.

146. Murzin, A.G. (1998). How far divergent evolution goes in proteins. *Curr Opin Struct Biol* **8**, pp. 380-387.

147. Nagaraja, V., Shepherd, J.C. & Bickle, T.A. (1985). A hybrid recognition sequence in a recombinant restriction enzyme and the evolution of DNA sequence specificity. *Nature* **316**, pp. 371-372.

148. Naito, T., Kusano, K. & Kobayashi, I. (1995). Selfish behavior of restriction-modification systems. *Science* **267**, pp. 897-899.

149. Needels, M.C., Fried, S.R., Love, R., Rosenberg, J.M., Boyer, H.W. & Greene, P.J. (1989). Determinants of EcoRI endonuclease sequence discrimination. *Proc Natl Acad Sci U S A* **86**, pp. 3579-3583.

150. Nelson, H.C. & Sauer, R.T. (1985). Lambda repressor mutations that increase the affinity and specificity of operator binding. *Cell* **42**, pp. 549-558.

151. Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I. & Aggarwal, A.K. (1994). Structure of restriction endonuclease BamHI phased at 1.95 A resolution by MAD analysis. *Structure* **2**, pp. 439-452.

152. Nicholls, A., Sharp, K.A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, pp. 281-296.

153. Novoa, J.J., Whangbo, M.-H. & Williams, J.M. (1991). Interactions energies associated with short intermolecular contacts of C--H bonds. II. Ab initio computational study of the C--H--H--C interactions in methane dimer. *J. Chem. Phys* **94**, pp. 4835-4841.

154. Oelgeschlager, T., Geiger, R., Ruter, T., Alves, J., Fliess, A. & Pingoud, A. (1990). Probing the function of individual amino acid residues in the DNA binding site of the EcoRI restriction endonuclease by analysing the toxicity of genetically engineered mutants. *Gene* **89**, pp. 19-27.

155. Oertel-Buchheit, P., Porte, D., Schnarr, M. & Granger-Schnarr, M. (1992). Isolation and characterization of LexA mutant repressors with enhanced DNA binding affinity. *J Mol Biol* **225**, pp. 609-620.

156. Olson, W.K., Bansal, M., Burley, S.K., Dickerson, R.E., Gerstein, M., Harvey, S.C., Heinemann, U., Lu, X.J., Neidle, S., Shakked, Z., Sklenar, H., Suzuki, M., Tung, C.S., Westhof, E., Wolberger, C. & Berman, H.M. (2001). A standard reference frame for the description of nucleic acid base-pair geometry. *J Mol Biol* **313**, pp. 229-237.

157. Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M. & Zhurkin, V.B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc Natl Acad Sci U S A* **95**, pp. 11163-11168.

158. O'Neill, M., Chen, A. & Murray, N.E. (1997). The restriction-modification genes of Escherichia coli K-12 may not be selfish: they do not resist loss and are readily replaced by alleles conferring different specificities. *Proc Natl Acad Sci U S A* **94**, pp. 14596-14601.

159. Osuna, J., Flores, H. & Soberon, X. (1991). Combinatorial mutagenesis of three major groove-contacting residues of EcoRI: single and double amino acid replacements retaining methyltransferase-sensitive activities. *Gene* **106**, pp. 7-12.

160. Otwinowski, Z., Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. In *Methods in Enzymology* (C.W. Carter, J., R.M. Sweet, ed.), vol. 276. pp. 307-326, Academic Press.

161. Otwinowski, Z., Schevitz, R.W., Zhang, R.G., Lawson, C.L., Joachimiak, A., Marmorstein, R.Q., Luisi, B.F. & Sigler, P.B. (1988). Crystal structure of trp repressor/operator complex at atomic resolution. *Nature* **335**, pp. 321-329.

162. Pace, C.N. & Scholtz, J.M. (1998). A helix propensity scale based on experimental studies of peptides and proteins. *Biophys J* **75**, pp. 422-427.

163. Paillard, G. & Lavery, R. (2004). Analyzing protein-DNA recognition mechanisms. *Structure (Camb)* **12**, pp. 113-122.

164. Parkinson, G. (1996). New parameters for the refinement of nucleic acid-containing structures. *Acta Crystallogr D Biol Crystallogr* **52**, pp. 57-64.

165. Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y.W., Ebright, R.E. & Berman, H.M. (1996). Structure of the CAP-DNA complex at 2.5 angstroms resolution: a complete picture of the protein-DNA interface. *J Mol Biol* **260**, pp. 395-408.

166. Perona, J.J. & Martin, A.M. (1997). Conformational transitions and structural deformability of EcoRV endonuclease revealed by crystallographic analysis. *J Mol Biol* **273**, pp. 207-225.

167. Perry, K.M., Fauman, E.B., Finer-Moore, J.S., Montfort, W.R., Maley, G.F., Maley, F. & Stroud, R.M. (1990). Plastic adaptation toward mutations in proteins: structural comparison of thymidylate synthases. *Proteins* **8**, pp. 315-333.

168. Plaxco, K.W. & Goddard, W.A., 3rd (1994). Contributions of the thymine methyl group to the specific recognition of poly- and mononucleotides: an analysis of the relative free energies of solvation of thymine and uracil. *Biochemistry* **33**, pp. 3050-3054.

169. Poon, G.M. & Macgregor, R.B., Jr. (2004). A thermodynamic basis of DNA sequence selectivity by the ETS domain of murine PU.1. *J Mol Biol* **335**, pp. 113-127.

170. Rambach, A. & Tiollais, P. (1974). Bacteriophage lambda having EcoRI endonuclease sites only in the nonessential region of the genome. *Proc Natl Acad Sci U S A* **71**, pp. 3927-3930.

171. Record, M.T., Jr., Courtenay, E.S., Cayley, S. & Guttman, H.J. (1998). Biophysical compensation mechanisms buffering E. coli protein-nucleic acid interactions against changing environments. *Trends Biochem Sci* **23**, pp. 190-194.

172. Record, M.T., Jr., Ha, J.H. & Fisher, M.A. (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 Enzymol* **208**, pp. 291-343.

173. Rice, L.M. & Brunger, A.T. (1994). Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins* **19**, pp. 277-290.

174. Richey, B., Cayley, D.S., Mossing, M.C., Kolka, C., Anderson, C.F., Farrar, T.C. & Record, M.T., Jr. (1987). Variability of the intracellular ionic environment of Escherichia coli. Differences between in vitro and in vivo effects of ion concentrations on protein-DNA interactions and gene expression. *J Biol Chem* **262**, pp. 7157-7164.

175. Riggs, A.D., Bourgeois, S. & Cohn, M. (1970). The lac repressor-operator interaction. 3. Kinetic studies. *J Mol Biol* **53**, pp. 401-417.

176. Riggs, A.D., Suzuki, H. & Bourgeoss, S. (1970). Lac repressor-operator interaction. I. Equilibrium studies. *J Mol Biol* **48**, pp. 67-83.

177. Rimseliene, R., Maneliene, Z., Lubys, A. & Janulaitis, A. (2003). Engineering of restriction endonucleases: using methylation activity of the bifunctional endonuclease Eco57I to select the mutant with a novel sequence specificity. *J Mol Biol* **327**, pp. 383-391.

178. Ringe, D. & Petsko, G.A. (2003). The 'glass transition' in protein dynamics: what it is, why it occurs, and how to exploit it. *Biophys Chem* **105**, pp. 667-680.

179. Ringe, D. & Petsko, G.A. (1986). Study of protein dynamics by X-ray diffraction. *Methods Enzymol* **131**, pp. 389-433.

180. Roberts, R.J. (1976). Restriction endonucleases. *CRC Crit Rev Biochem* **4**, pp. 123-164.

181. Roberts, R.J., Vincze, T., Posfai, J. & Macelis, D. (2003). REBASE: restriction enzymes and methyltransferases. *Nucleic Acids Res* **31**, pp. 418-420.

182. Rocha, E.P., Danchin, A. & Viari, A. (2001). Evolutionary role of restriction/modification systems as revealed by comparative genome analysis. *Genome Res* **11**, pp. 946-958.

183. Saito, M. & Sarai, A. (2003). Free energy calculations for the relative binding affinity between DNA and lambda-repressor. *Proteins* **52**, pp. 129-136.

184. Schowen, R. (1978). *Transition States of Biological Processes*.

185. Schultz, S.C., Shields, G.C. & Steitz, T.A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science* **253**, pp. 1001-1007.

186. Seeman, N.C., Rosenberg, J.M. & Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proc Natl Acad Sci U S A* **73**, pp. 804-808.

187. Senior, M., Jones, R.A. & Breslauer, K.J. (1988). Influence of dangling thymidine residues on the stability and structure of two DNA duplexes. *Biochemistry* **27**, pp. 3879-3885.

188. Sharp, K.A., Friedman, R.A., Misra, V., Hecht, J. & Honig, B. (1995). Salt effects on polyelectrolyte-ligand binding: comparison of Poisson-Boltzmann, and limiting law/counterion binding models. *Biopolymers* **36**, pp. 245-262.

189. Shui, X., McFail-Isom, L., Hu, G.G. & Williams, L.D. (1998). The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry* **37**, pp. 8341-8355.

190. Siebenlist, U. & Gilbert, W. (1980). Contacts between Escherichia coli RNA polymerase and an early promoter of phage T7. *Proc Natl Acad Sci U S A* **77**, pp. 122-126.

191. Sinha, N. & Nussinov, R. (2001). Point mutations and sequence variability in proteins: redistributions of preexisting populations. *Proc Natl Acad Sci U S A* **98**, pp. 3139-3144.

192. Spitzner, N., Lohr, F., Pfeiffer, S., Koumanov, A., Karshikoff, A. & Ruterjans, H. (2001). Ionization properties of titratable groups in ribonuclease T1. I. pKa values in the native state determined by two-dimensional heteronuclear NMR spectroscopy. *Eur Biophys J* **30**, pp. 186-197.

193. Spolar, R.S. & Record, M.T., Jr. (1994). Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263**, pp. 777-784.

194. Srinivasan, A.R., Torres, R., Clark, W. & Olson, W.K. (1987). Base sequence effects in double helical DNA. I. Potential energy estimates of local base morphology. *J Biomol Struct Dyn* **5**, pp. 459-496.

195. Sturtevant, J.M. (1977). Heat capacity and entropy changes in processes involving proteins. *Proc Natl Acad Sci U S A* **74**, pp. 2236-2240.

196. Sun, J., Viadiu, H., Aggarwal, A.K. & Weinstein, H. (2003). Energetic and structural considerations for the mechanism of protein sliding along DNA in the nonspecific BamHI-DNA complex. *Biophys J* **84**, pp. 3317-3325.

197. Takeda, Y., Ross, P.D. & Mudd, C.P. (1992). Thermodynamics of Cro protein-DNA interactions. *Proc Natl Acad Sci U S A* **89**, pp. 8180-8184.

198. Takeda, Y., Sarai, A. & Rivera, V.M. (1989). Analysis of the sequence-specific interactions between Cro repressor and operator DNA by systematic base substitution experiments. *Proc Natl Acad Sci U S A* **86**, pp. 439-443.

199. Ten Eyck, L.F. (1995). Gro-ing pains. *Nat Struct Biol* **2**, pp. 1038-1042.

200. Terry, B.J., Jack, W.E. & Modrich, P. (1985). Facilitated diffusion during catalysis by EcoRI endonuclease. Nonspecific interactions in EcoRI catalysis. *J Biol Chem* **260**, pp. 13130-13137.

201. Uversky, V.N. (2002). Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* **11**, pp. 739-756.

202. Vaguine, A.A., Richelle, J. & Wodak, S.J. (1999). SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr D Biol Crystallogr* **55 (Pt 1)**, pp. 191-205.

203. Viadiu, H. & Aggarwal, A.K. (1998). The role of metals in catalysis by the restriction endonuclease BamHI. *Nat Struct Biol* **5**, pp. 910-916.

204. Viadiu, H. & Aggarwal, A.K. (2000). Structure of BamHI bound to nonspecific DNA: a model for DNA sliding. *Mol Cell* **5**, pp. 889-895.

205. Vipond, I.B. & Halford, S.E. (1995). Specific DNA recognition by EcoRV restriction endonuclease induced by calcium ions. *Biochemistry* **34**, pp. 1113-1119.

206. von Hippel, P.H., Revzin, A., Gross, C.A. & Wang, A.C. (1974). Non-specific DNA binding of genome regulating proteins as a biological control mechanism: I. The lac operon: equilibrium aspects. *Proc Natl Acad Sci U S A* **71**, pp. 4808-4812.

207. Warmlander, S., Sponer, J.E., Sponer, J. & Leijon, M. (2002). The influence of the thymine C5 methyl group on spontaneous base pair breathing in DNA. *J Biol Chem* **277**, pp. 28491-28497.

208. Warshaw, M.M. & Cantor, C.R. (1970). Oligonucleotide interactions. IV. Conformational differences between deoxy- and ribodinucleoside phosphates. *Biopolymers* **9**, pp. 1079-1103.

209. Watrob, H., Liu, W., Chen, Y., Bartlett, S.G., Jen-Jacobson, L. & Barkley, M.D. (2001). Solution conformation of EcoRI restriction endonuclease changes upon binding of cognate DNA and Mg2+ cofactor. *Biochemistry* **40**, pp. 683-692.

210. Weiss, M.A., Ellenberger, T., Wobbe, C.R., Lee, J.P., Harrison, S.C. & Struhl, K. (1990). Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature* **347**, pp. 575-578.

211. Wilkosz, P.A. (1993). *Preliminary Structure of an EcoRI* Endonuclease-DNA Complex* (Ph.D. Thesis) Biological Sciences, University of Pittsburgh: Pittsburgh, PA.

212. Wilson, G.G. (1991). Organization of restriction-modification systems. *Nucleic Acids Res* **19**, pp. 2539-2566.

213. Winkler, F.K., Banner, D.W., Oefner, C., Tsernoglou, D., Brown, R.S., Heathman, S.P., Bryan, R.K., Martin, P.D., Petratos, K. & Wilson, K.S. (1993). The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *Embo J* **12**, pp. 1781-1795.

214. Winter, R.B., Berg, O.G. & von Hippel, P.H. (1981). Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor--operator interaction: kinetic measurements and conclusions. *Biochemistry* **20**, pp. 6961-6977.

215. Wolfes, H., Alves, J., Fliess, A., Geiger, R. & Pingoud, A. (1986). Site directed mutagenesis experiments suggest that Glu 111, Glu 144 and Arg 145 are essential for endonucleolytic activity of EcoRI. *Nucleic Acids Res* **14**, pp. 9063-9080.

216. Xu, S.Y. & Schildkraut, I. (1991). Isolation of BamHI variants with reduced cleavage activities. *J Biol Chem* **266**, pp. 4425-4429.

217. Yanofsky, S.D., Love, R., McClarin, J.A., Rosenberg, J.M., Boyer, H.W. & Greene, P.J. (1987). Clustering of null mutations in the EcoRI endonuclease. *Proteins* **2**, pp. 273-282.

218. Yoshimori, R., Roulland-Dussoix, D. & Boyer, H.W. (1972). R factor-controlled restriction and modification of deoxyribonucleic acid: restriction mutants. *J Bacteriol* **112**, pp. 1275-1279.

219. Zhu, Z., Zhou, J., Friedman, A.M. & Xu, S.Y. (2003). Isolation of BsoBI restriction endonuclease variants with altered substrate specificity. *J Mol Biol* **330**, pp. 359-372.

220. Zhurkin, V.B., Ulyanov, N.B., Gorin, A.A. & Jernigan, R.L. (1991). Static and statistical bending of DNA evaluated by Monte Carlo simulations. *Proc Natl Acad Sci U S A* **88**, pp. 7046-7050.