

# **Lanthanides to Probe Metal-induced Structural and Dynamic Changes in EcoRV-DNA Complex**

by

Sahil S. Sangani

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at

Carnegie Mellon University

Department of Biological Sciences

Pittsburgh, Pennsylvania

Advised by Dr. Gordon S. Rule

August 19, 2016

I would like to dedicate this thesis to my loving parents, my brother, and my best friend  
and fiancée ...

## Acknowledgements

I would like to express my sincere gratitude to my advisor, Dr. Gordon S. Rule, for his continuous guidance, patience and scientific and professional support. His knowledge and passion for science has helped me immensely throughout my graduate career and I would like to thank him for helping me give shape to my dissertation and my scientific career.

I would like to thank our collaborator, Dr. Linda Jen-Jacobson, for her valuable insights and advice throughout this project. I would also like to thank my committee members, Dr. David D. Hackney and Dr. Bruce A. Armitage for their time and their valuable inputs.

I have been fortunate to meet extremely supportive colleagues who made the laboratory environment a productive and entertaining place to be. I would like to express my genuine thanks to Dr. Kaustubh Sinha for his help and guidance with all NMR experiments. I would also like to acknowledge the past and current members of the Rule laboratory, Dr. Andy Kehr, Dr. John Pettersson, Dr. Shantanu Bhattacharyya, Paul Barton and Ian Fucci for all their guidance and assistance, Virgil Simplaceanu for his help and training of the NMR spectrometers and Doowon Lee at University of Pittsburgh X-ray crystallography training.

During the course of my studies, I have been lucky to find friends like Dr. Madhumitha Ramesh and Dr. Jigar Desai who have made my journey here at CMU memorable. Their personal support and friendship through all the ups and downs of during this time has been invaluable and I will always cherish our scientific and non-scientific discussions, laughs and good times that we shared at Mellon Institute. I would also like to thank all my other classmates and friends, Dr. Salini Konnikat, Dr. Karen Adria, Dr. Shanna Bowersox-

Bowman, Dr. Gino Gallo, Dr. Olyvya Molinar and Ram Shankar Siva Kumar for making my time at CMU a very enjoyable experience.

One of the things I enjoyed the most about studying in Pittsburgh was getting to play cricket on the weekends. The Kent and CMU cricket team have become like a family to me and I would like to specially acknowledge Sameen Khan, Mudassar Zahoor, Kashif Sheikh, Saurabh Shintre, Pritish Kamath, Prajna Pati, Vishal Naik, Anoop Shashidhran, Yaseen Karche, Kushal Kapoor, Agniv Bannerjee, Chetan Rajput, Shiva Ramalingam, Sagar Shah, Sahil Jolly, Gautam Nagaraju, Sandeep Khatri, Arun Srivatsan, Ravi Vishwanathan, Karan Kalsaria and Dhruv Shah for the good times we shared as a team.

My dissertation and the entire journey leading up to it would not be possible without the unfaltering support and encouragement of my family. My father Sailesh Sangani, my mother Manisha Sangani and my brother Yash Sangani have been a constant source of inspiration and comfort. Words are not enough to describe their love and care and I will always be indebted to them for giving me the opportunity to become the person I am today.

Finally, I would like to thank my fiancée Sneha Potdar. She has been my best friend and unwavering confidante for the past 12 years. Her perpetual motivation and heartwarming support kept me going in the good times and the bad. Thank you for being by my side, now and forever.

## Abstract

Protein-DNA interactions regulate key biological processes such as gene expression, genetic recombination, DNA synthesis and repair. DNA binding proteins that are enzymes often depend on metal ions (typically  $Mg^{2+}$ ) for catalysis. Metal ions neutralize the active site electrostatic repulsion generated by a constellation of negatively charged active site residues. They therefore enhance site-specific protein-DNA binding by making the interaction more energetically favorable. Although, specific protein-DNA-metal complexes have been studied extensively by various biophysical and biochemical methods, little is known about the dynamics of these complexes. Many crystal structures of protein-DNA-metal complexes are available but they are very similar to the metal-free structures and show little or no change due to metal-ion binding. Therefore, to understand the altered energetics of protein-DNA recognition complexes due to metal-ion binding, it is essential to study them in solution. In our study, we have used EcoRV-DNA complex, with  $Lu^{3+}$  as the metal ion, to detect metal-induced structural and/or dynamic changes in solution using paramagnetic relaxation enhancement (PRE).  $Lu^{3+}$  inhibits  $Mg^{2+}$  dependent DNA hydrolysis in EcoRV, stimulates EcoRV-DNA affinity, binds in the same site and with the same stoichiometry as  $Mg^{2+}$  and is therefore an excellent substitute for  $Mg^{2+}$ . In addition, to verify that  $Lu^{3+}$  binding is confined to the active sites, we have solved crystal structures of the EcoRV-DNA- $Lu^{3+}$  complex with uncleaved and cleaved DNA. Also, for a comprehensive residue-specific analysis, we have acquired both backbone NH and methyl side-chain (ILV) assignments for EcoRV-DNA complex with and without  $Lu^{3+}$ . In order to detect metal-induced changes in

the EcoRV-DNA complex, backbone and methyl side-chain PRE rates were measured with spin-labels at S2C, K197C and S234C of EcoRV. The measured PRE rates suggest that the crystal structures of EcoRV-DNA complexes better represent the complex with metal-ion in solution whereas the nometal complex in solution shows differences in structure and/or dynamics when compared to the crystal structure. The C-terminal region of EcoRV is observed to be more flexible in solution in the nometal complex and its conformational flexibility is reduced due to metal-ion binding.

# Contents

<b>Contents</b>	<b>vi</b>
<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 EcoRV-DNA complex as model system . . . . .	5
1.1.1 Catalytic mechanism of EcoRV . . . . .	5
1.1.2 DNA-induced conformational changes in EcoRV . . . . .	8
1.1.3 Comparison of EcoRV-DNA complex crystal structures $\pm$ metal ion	10
1.2 Metal ions . . . . .	12
1.3 Paramagnetic relaxation enhancement (PRE) . . . . .	13
<b>2 Lanthanide ions as substitutes for Magnesium</b>	<b>15</b>
2.1 Ln <sup>3+</sup> inhibit Mg <sup>2+</sup> dependent DNA hydrolysis in EcoRV . . . . .	16
2.2 DNA affinity and binding stoichiometry . . . . .	20
2.3 Ln <sup>3+</sup> bind in the active sites . . . . .	23
2.4 EcoRV-DNA-Lu <sup>3+</sup> complex crystal structures . . . . .	27
2.4.1 Local environment of Lu <sup>3+</sup> in the crystal structures . . . . .	32
2.4.2 Comparison with EcoRV-DNA-Mg <sup>2+</sup> structures . . . . .	34

---

2.5	Conclusion . . . . .	38
<b>3</b>	<b>Backbone amide and methyl side-chain assignments of EcoRV</b>	<b>39</b>
3.1	Backbone NH assignments . . . . .	40
3.2	Methyl assignments . . . . .	45
3.2.1	Correlation spectroscopy . . . . .	49
3.2.2	Site-directed mutagenesis . . . . .	49
3.2.3	Paramagnetic relaxation enhancement (PRE) . . . . .	51
3.2.4	4D methyl-methyl NOE . . . . .	53
3.2.5	Stereospecific assignments of Leu and Val . . . . .	56
3.3	Conclusion . . . . .	58
<b>4</b>	<b>Paramagnetic relaxation enhancement (PRE)</b>	<b>59</b>
4.1	Detecting domain motions by PRE . . . . .	59
4.2	Spin-labeling . . . . .	62
4.3	Cysteine-free EcoRV and nitroxide labeling sites . . . . .	64
4.4	PRE measurements on EcoRV-DNA complex . . . . .	69
4.4.1	Back-calculating PREs from crystal structures . . . . .	72
4.4.2	Spin-label at S2C . . . . .	73
4.4.3	Spin-label at K197C and S234C . . . . .	76
4.5	Discussion . . . . .	79
<b>5</b>	<b>Summary, Conclusions and Future Directions</b>	<b>80</b>
5.1	Summary and Conclusions . . . . .	80
5.2	Future Directions . . . . .	81
5.2.1	Residual dipolar coupling (RDC) experiments . . . . .	81
5.2.2	Inter-subunit distances and PRE with d(His) Cu <sup>2+</sup> - IDA . . . . .	82

---

5.2.3 Spin-labeling the DNA, PRE with paramagnetic lanthanides and DNA dynamics . . . . .	83
<b>6 Materials and Methods</b>	<b>85</b>
6.1 Expression of wild type and mutant EcoRV endonucleases . . . . .	85
6.2 Protein purification . . . . .	86
6.3 Oligodeoxynucleotide substrates . . . . .	86
6.4 NMR sample preparation . . . . .	87
6.4.1 Paramagnetic relaxation enhancement (PRE) samples . . . . .	87
6.5 NMR spectroscopy . . . . .	89
6.5.1 Methyl assignments . . . . .	89
6.5.2 PRE experiments . . . . .	90
6.6 Crystal structures of EcoRV-DNA-Lu <sup>3+</sup> complexes . . . . .	91
6.6.1 Crystallization and data collection . . . . .	91
6.6.2 Structure determination, model building and refinement . . . . .	92
<b>References</b>	<b>93</b>
<b>I Appendix</b>	<b>104</b>
<b>II Appendix</b>	<b>124</b>

# List of Figures

1.1	Dynamics by NMR spectroscopy. . . . .	4
1.2	Catalytic mechanism of EcoRV. . . . .	7
1.3	DNA-induced conformational changes in EcoRV. . . . .	9
1.4	Superimpositions of EcoRV-DNA complex structures with and without metal ion. . . . .	11
2.1	Competitive inhibition of DNA cleavage in EcoRV by lanthanides. . . . .	17
2.2	Stimulation of EcoRV-DNA affinity by lanthanides. . . . .	19
2.3	Stoichiometries of lanthanide ion binding. . . . .	22
2.4	Lu <sup>3+</sup> and Mg <sup>2+</sup> ions occupy the same sites. . . . .	25
2.5	Ln <sup>3+</sup> ions bind in the active site. . . . .	26
2.6	Crystal structure of the EcoRV-DNA-Lu <sup>3+</sup> complex with substrate DNA. . . . .	29
2.7	Crystal structure of the EcoRV-DNA-Lu <sup>3+</sup> complex with product DNA. . . . .	30
2.8	Local environment of Lu <sup>3+</sup> in uncleaved and cleaved EcoRV-DNA-Lu <sup>3+</sup> complex crystal structures. . . . .	33
2.9	Superimposition of the substrate EcoRV-DNA-Lu <sup>3+</sup> complex crystal structure with the EcoRV structures of substrate and product DNA with Mg <sup>2+</sup> . . . . .	36
2.10	Superimposition of the product EcoRV-DNA-Lu <sup>3+</sup> complex crystal structure with the EcoRV structures of substrate and product DNA with Mg <sup>2+</sup> . . . . .	37

---

3.1	Schematic of triple resonance experiments used for backbone NH assignments. . . . .	41
3.2	Backbone assignment of EcoRV-DNA complex without metal. . . . .	42
3.3	Backbone assignment of EcoRV-DNA-Lu <sup>3+</sup> complex. . . . .	43
3.4	Coverage map of assigned backbone NH. . . . .	44
3.5	Assignment of Ile- $\delta$ 1-CH <sub>3</sub> . . . . .	46
3.6	Assignment of Val- $\gamma$ 2 and Leu- $\delta$ 2 methyls. . . . .	47
3.7	Coverage map of assigned ILV-methyls. . . . .	48
3.8	Assignment of ILV methyls using site-directed mutagenesis. . . . .	50
3.9	PRE to verify ILV methyl assignments. . . . .	52
3.10	Biosynthetic precursors used to make 4D methyl-methyl NOE samples. . .	54
3.11	Example of assignment by 4D methyl-methyl NOE. . . . .	55
3.12	Stereospecific methyl assignment of Leu and Val residues. . . . .	57
4.1	Schematic representation of detecting domain motions by PRE. . . . .	61
4.2	Structure and conjugation of MTSL and MTS spin-labels. . . . .	63
4.3	Cysteine-free EcoRV to prevent undesired spin-labeling. . . . .	66
4.4	Sites mutated to cysteine for spin-labeling. . . . .	67
4.5	Spin-label does not affect the protein structure. . . . .	68
4.6	<sup>1</sup> H transverse relaxation decay curves. . . . .	71
4.7	PRE correlation plots for spin-label at S2C. . . . .	75
4.8	PRE correlation plots for spin-label at K197C. . . . .	77
4.9	PRE correlation plots for spin-label at S234C. . . . .	78

# List of Tables

2.1	Binding and cleavage inhibition constants of lanthanides in EcoRV-DNA complex. . . . .	18
2.2	Data collection and refinement statistics for EcoRV-DNA-Lu <sup>3+</sup> complexes.	28
2.3	DNA conformation analysis in EcoRV-DNA complex crystal structures. . .	31

# Chapter 1

## Introduction

Protein-DNA interactions play a crucial role in biology as they regulate key processes such as gene expression, DNA repair, replication and recombination. Thus, understanding the molecular mechanisms that govern these interactions is of great importance to engineer novel therapeutic interventions. Many nucleic acid-binding proteins are enzymes that depend on metal ions (typically  $Mg^{2+}$ ) for nucleic acid synthesis and/or hydrolysis. The 3'-exonuclease of DNA polymerase I, T7 DNA and RNA polymerases, Dicer, RNase P, RNase H, endonucleases PvuII and EcoRV, and group I and group II ribozymes are a few examples of such metalloenzymes. In addition to their role in catalysis, metal ions enhance site-specific protein-DNA binding by several orders of magnitude. The active-sites of metal-ion dependent DNA binding proteins are composed of Asp and/or Glu sidechains as well as the DNA phosphate. Metal ions make DNA binding energetically more favorable by neutralizing the electrostatic repulsion among these negatively charged active-site residues.

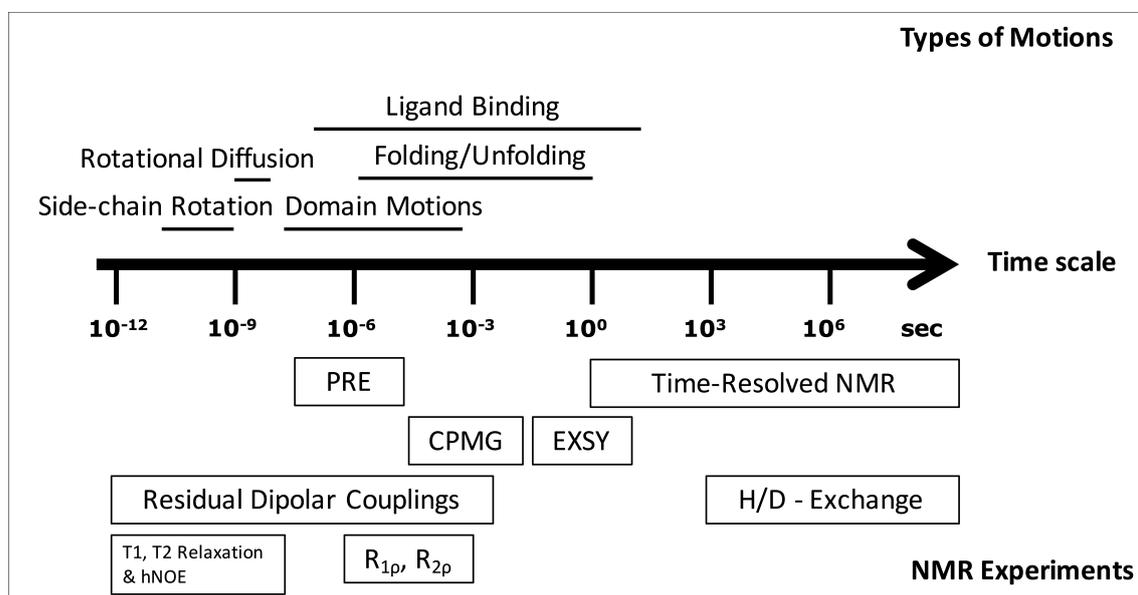
Site-specific protein-DNA interactions are characterized by a decrease in heat capacity ( $\Delta C_p^\circ$ ) as reported for endonucleases EcoRI and BamHI (Jen-Jacobson et al., 2000). In this study, metal-ions have been observed to further decrease the heat capacity ( $\Delta C_p^\circ$ ) in the specific protein-DNA complexes.  $\Delta C_p^\circ$  can be correlated to dynamic fluctuations (picosecond - nanosecond) and thus decrease in heat capacity upon metal-ion binding may

---

suggest changes in conformational flexibility of these complexes. Although, such specific protein-DNA-metal complexes have been studied extensively by X-ray crystallography, crystal structures of these complexes show little or no change due to metal-ion binding and do not reveal any information about the dynamics of these complexes  $\pm$  metal ions. Thus, the long-term goal of our laboratory is to detect metal-induced changes in the structure and dynamics of protein-DNA complexes and draw correlations with the observed thermodynamic parameters ( $\Delta C_p^\circ$ ). Structural changes refer to changes in mean positions of atomic coordinates whereas dynamics refer to the time-dependent fluctuations in structure.

Over the last couple decades, conformational selection theory or the existence of multiple conformational states in solution has become popular (Vega et al., 2016), underscoring the importance of dynamic studies in biomolecular recognition. Therefore, for a comprehensive analysis of metal-induced structural and/or dynamic changes in protein-DNA complexes, it is essential to study them in solution to avoid constraints from the crystal lattice. In recent studies with PvuII endonuclease-DNA complex, conformational changes were detected due to  $\text{Ca}^{2+}$  binding in solution. However, specific structural and/or dynamic changes due to  $\text{Ca}^{2+}$  binding were not clearly identified (Dupureur, 2005). Recent advances in solution NMR spectroscopy, ranging from availability of more powerful and sensitive spectrometers to the development of various isotope labeling strategies, have permitted the study of dynamics over a broad range of time-scales (Figure 1.1 on page 4). Spin relaxation experiments such as T1, T2 and hNOE measure fast reorientational motions on picosecond - nanosecond time scales. Relaxation dispersion experiments ( $R_{1\rho}, R_{2\rho}$ ) measure slower collective motions characterized by microsecond - millisecond time scales and are sensitive to chemical exchange processes. H/D - exchange studies report on much slower motions or conformational transitions (folding/unfolding) due to ligand binding. Paramagnetic relaxation enhancement (PRE) and residual dipolar coupling (RDC) experiments can be used to probe both structural and/or dynamic changes. PRE allows the detection of larger domain

motions characterized by microsecond - millisecond time scales. Some of these larger domain motions may not be observable by other techniques such as chemical shift perturbations (CSPs) that are sensitive to changes in environment or relaxation dispersion experiments which require substantial chemical shift differences. RDC experiments can be used to extract information on picosecond - millisecond dynamics. Therefore, we have used solution NMR spectroscopy for a comprehensive analysis of metal-induced changes in a site-specific protein-DNA complex. In this chapter, first an introduction of EcoRV as our model system is given. Then, the metal ions used in the study are discussed followed by a background of the NMR technique used to study structural and/or dynamic changes. In the next three chapters, results of this study are presented and discussed.



**Figure 1.1: Dynamics by NMR spectroscopy.** The central arrow represents the time-scale of protein motions. Types of motions are indicated above the arrow and the corresponding NMR experiments are shown at the bottom.

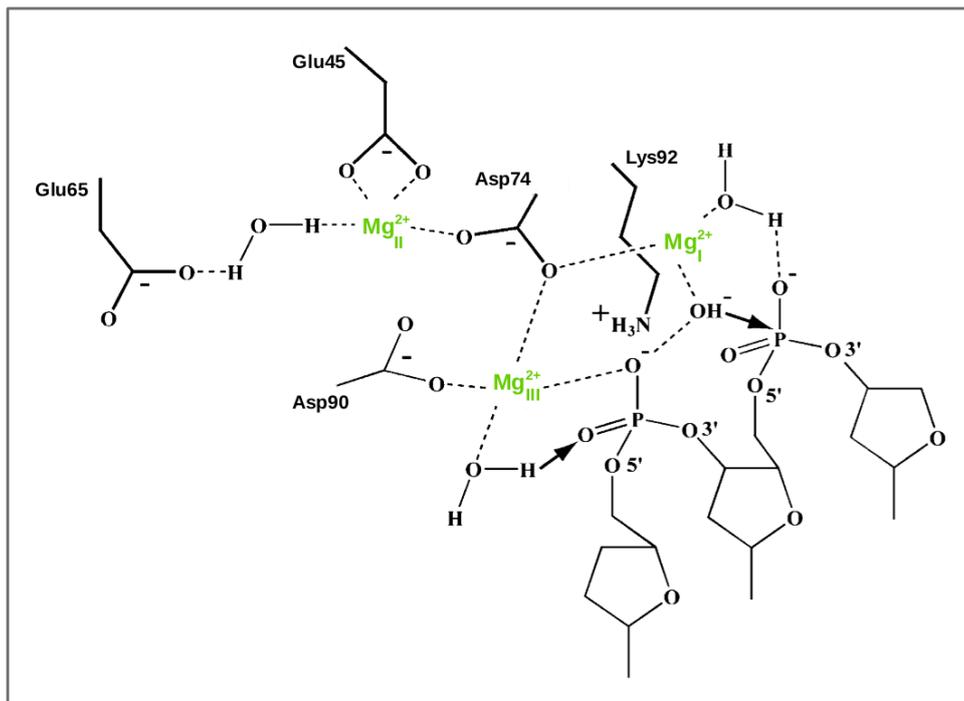
## 1.1 EcoRV-DNA complex as model system

In order to study the metal-induced dynamic changes in a specific protein-DNA complex, we chose EcoRV endonuclease as our model system. EcoRV is a homodimer of 29 kDa subunits and specifically binds the DNA sequence GATATC. EcoRV is stable even at high temperatures (40°C), unusual for a mesophile, making it suitable for NMR studies. There is a large body of thermodynamic and kinetic information available for EcoRV-DNA complexes (Engler et al., 1997; Hancock et al., 2011; Hiller et al., 2003; Martin et al., 1999). There are over 30 crystal structures of EcoRV in the PDB, including the apo-enzyme (Perona and Martin, 1997; Winkler et al., 1993), DNA complexes with cognate (Kostrewa and Winkler, 1995; Winkler et al., 1993) and non-specific sites (Winkler et al., 1993) as well as numerous forms with bound metal ions (Perona and Martin, 1997; Thomas et al., 1999). Thus, the specific recognition interactions between EcoRV and various DNA sequences are very well characterized. In addition, the exceptionally high yields of purified EcoRV (~100 mg/L) make it feasible to obtain NMR samples with expensive precursors.

### 1.1.1 Catalytic mechanism of EcoRV

The active site of EcoRV has a cluster of negatively charged residues (Glu 45, Asp 74 and Asp 90) surrounding the scissile DNA phosphate. EcoRV, like other Type II restriction endonucleases, catalyzes the blunt-end double-strand DNA hydrolysis by binding two  $Mg^{2+}$  ions in each of the two symmetrical active sites. These Type II nucleases follow an  $S_N2$ -type mechanism, in which first an attacking nucleophile (hydroxide ion) is activated by deprotonation of a water molecule. The activated hydroxide ion attacks the phosphorous of the scissile phosphate, forming a pentavalent transition state which is then followed by protonation of the 3' hydroxyl leaving group. For EcoRV, a catalytic mechanism with two metal ions in three different sites has been proposed (Figure 1.2 on page 7). Metal ion in site I is involved in generating an activated nucleophilic hydroxyl by deprotonation of a water

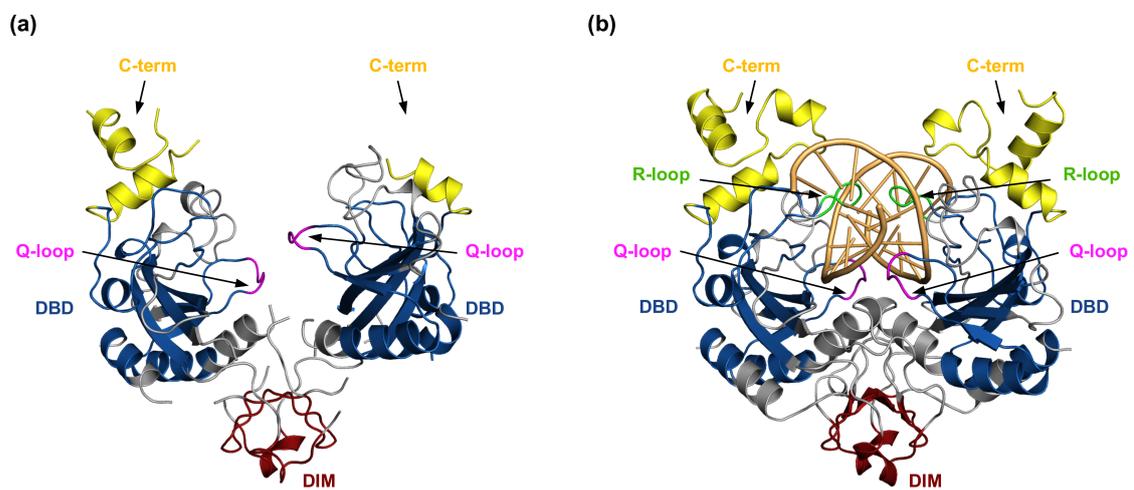
molecule. The second metal ion initially occupies site III, where it stabilizes the transition state formed from nucleophilic attack by the hydroxide ion. Later, the site III metal ion shifts to site II where it is implicated in maintaining the active site geometry and protonation of the 3'-oxyanion (Horton and Perona, 2004; Pingoud et al., 2005).



**Figure 1.2: Catalytic mechanism of EcoRV.** Two metal ions occupy three positions during the reaction. Site I Mg<sup>2+</sup> activates the nucleophilic water, site III Mg<sup>2+</sup> stabilizes the transition state while site II Mg<sup>2+</sup> has a structural role. Figure adapted from Pingoud et al., 2005

### 1.1.2 DNA-induced conformational changes in EcoRV

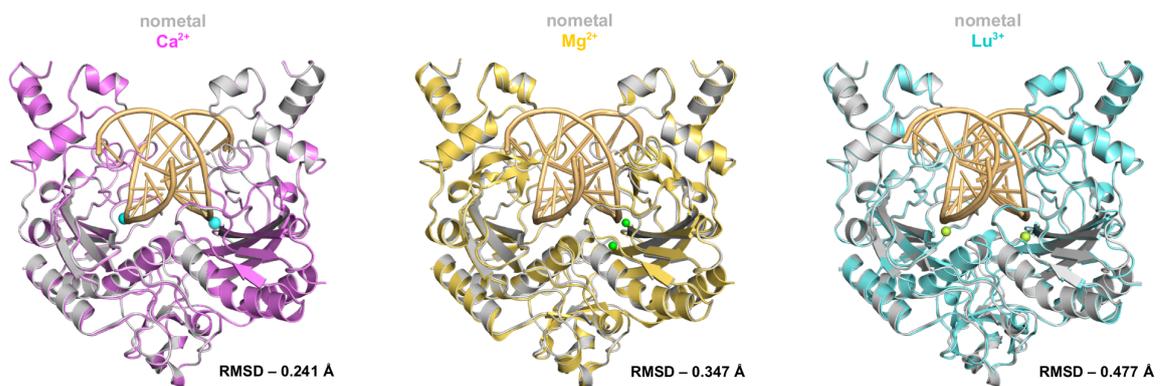
A characteristic structural feature of EcoRV-DNA complex is the sharp DNA bend ( $\sim 50^\circ$ ) when bound to cognate sequence (GATATC). In addition, there are local structural rearrangements of EcoRV domains that occur due to DNA binding. The structures of unliganded and cognate DNA bound EcoRV are shown in Figure 1.3 on page 9 and the structural domains of EcoRV that undergo transitions are highlighted. DNA-induced movements are observed in the DNA binding domain (DBD), the C-terminal domain and the dimerization interface (DIM) of each monomer. The two DNA binding domains (DBD) rotate by  $25^\circ$  with respect to each other. In addition, the R-loops (183-187) and Q-loops (67-70) that bind in the major and minor grooves of DNA become more ordered upon DNA binding. Most notably, the C-terminal region of EcoRV undergoes a disorder-to-order transition upon DNA binding (Horton and Perona, 2000). The C-terminal domain is highly disordered in one of the subunits of the unliganded EcoRV but becomes well-defined with DNA bound (Figure 1.3 on page 9). Tyr 219, Arg 221, Ser 223, Gln 224 and Arg 226 residues of the C-terminal region make contacts with the DNA and stabilize the nearby residues (217-245), promoting their folding. Studies with truncated EcoRV mutant lacking the C-terminal domain show that the C-terminal domain plays a key role in DNA association, bending and cleavage (Hiller and Perona, 2006). Computational studies of DNA-free and DNA-bound EcoRV suggest that the dynamics of the C-terminal region is reduced due to DNA binding (Zahran et al., 2010).



**Figure 1.3: DNA-induced conformational changes in EcoRV.** Crystal structures of (a) DNA-free (PDB ID 1AZ3) and (b) cognate DNA-bound (PDB ID 1B95) EcoRV. The domains that undergo conformational transitions are labeled.

### 1.1.3 Comparison of EcoRV-DNA complex crystal structures $\pm$ metal ion

Unlike the conformational transitions observed in response to DNA binding in the DNA-free and DNA-bound structures, metal-ion binding does not show any major changes in EcoRV-DNA complex crystal structures. Superimpositions of the structures with and without metals ( $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$  and  $\text{Lu}^{3+}$ ) are shown in Figure 1.4 on page 11.  $\text{RMSD}_{C\alpha}$  values for superimpositions of structures with  $\text{Ca}^{2+}$  (PDB ID 1B94),  $\text{Mg}^{2+}$  (PDB ID 1RVB) and  $\text{Lu}^{3+}$  (PDB ID 5F8A) with the no-metal complex (PDB ID 1B95) are 0.241, 0.347 and 0.477 respectively. Barring minor changes in loop positions and residues that interact with the metal ions, the crystal structures with and without metal are identical. The C-terminal region that undergoes disorder-to-order transition in response to DNA binding, appears unaffected due to metal-ion binding in the crystal structures  $\pm$  metal ions. This may be because many C-terminal residues (Q224, L225, N227, S234, N238, Y241, K245) participate in crystal lattice contacts, which may prevent the observation of any metal-induced changes in their positions in the crystal structures.



**Figure 1.4: Superimpositions of EcoRV-DNA complex structures with and without metal ion.** EcoRV-DNA complex crystal structure without metal (gray; PDB ID 1B95) is superimposed with the structure with Ca<sup>2+</sup> (magenta; PDB 1B94), Mg<sup>2+</sup> (yellow; PDB ID 1RVB) and Lu<sup>3+</sup> (cyan; PDB ID 5F8A). RMSD of superimpositions were calculated using C<sub>α</sub> atoms and are indicated.

## 1.2 Metal ions

We cannot use  $Mg^{2+}$  to study the ternary EcoRV-DNA-metal complex as it would lead to DNA cleavage. Previously  $Ca^{2+}$  was reported to not support catalysis in EcoRV and PvuII and was used as a substitute for  $Mg^{2+}$  to study DNA binding (Bowen and Dupureur, 2003; Vipond et al., 1995). However, in our NMR experiments with  $Ca^{2+}$  we observed DNA hydrolysis over time (>24 hours). We are unsure if this is due to  $Ca^{2+}$  ions supporting slow catalysis with EcoRV or is it because of trace amounts of divalent metal ion contaminants ( $Mg^{2+}$ ,  $Mn^{2+}$  etc) in our calcium salts. Lanthanide ions have been reported to not support catalysis in Dicer ribonuclease (Macrae et al., 2006), Klenow fragment of DNA polymerase I, T4 DNA polymerase (Brautigam et al., 1999; Frey et al., 1996) and PvuII endonuclease (Bowen and Dupureur, 2003). Crystal structure of Dicer shows two  $Er^{3+}$  ions bound similar to two  $Mg^{2+}$  in the active site. On the other hand, crystal structure of Klenow fragment of DNA polymerase I shows the larger  $Eu^{3+}$  ion occupying only one of two  $Mg^{2+}$  sites.

Lanthanides exhibit an important phenomenon called lanthanide contraction in which their ionic radius decreases, from left to right in the periodic table, as their atomic number increases.  $Lu^{3+}$  has an ionic radius (86 pm) similar to  $Mg^{2+}$  (72 pm), whereas ionic radius of  $La^{3+}$  (103 pm) is similar to  $Ca^{2+}$  (100 pm) (Shannon and IUCr, 1976). In addition, most lanthanides are paramagnetic (except  $Lu^{3+}$  and  $La^{3+}$ ), enabling their use in paramagnetic relaxation enhancement (PRE) and pseudocontact shift studies by NMR (Otting, 2010), or for electron paramagnetic resonance studies (Hubbell et al., 2013; Kaminker et al., 2012). Furthermore, lanthanides such as  $Eu^{3+}$  and  $Tb^{3+}$  have been extensively used as luminescence probes to study metal ion binding in proteins (Bowen and Dupureur, 2003; Brautigam et al., 1999; Frey et al., 1996). All these properties make lanthanides very attractive candidates to be tested as potential substitutes for  $Mg^{2+}$  in the EcoRV-DNA complex.

### 1.3 Paramagnetic relaxation enhancement (PRE)

Paramagnetic relaxation enhancement or PRE is an NMR technique that provides long-range distance information (up to 35 Å) useful for solving macromolecular structures or to study their dynamic processes. Our initial chemical shift perturbation (CSP) and amide exchange studies show widespread structural and dynamic changes in EcoRV-DNA complex due to Lu<sup>3+</sup> binding (Sinha et al., manuscript in preparation). Since PRE can be used to extract both structural and dynamics information, it was used to supplement CSP and amide exchange data and survey additional metal-induced changes in EcoRV-DNA complex. Large domain motions that may not have been detected by CSPs, should be observable by PRE. In addition, the transverse PRE rate ( $\Gamma_2$ ), which is typically measured, is relatively insensitive to fast internal dynamics, making it ideal to study changes in the structure and/or dynamics without the confounding effect of small internal motions.

The theoretical framework for PRE was worked out a long time ago (Solomon, 1955) but its use was restricted to systems with intrinsic paramagnetic groups or metalloproteins (Bertini et al., 2001). Recent advances in site-directed spin labeling (SDSL) have permitted the use of PRE beyond these systems, by facilitating the conjugation of macromolecules with an extrinsic paramagnetic group at a specific site of choice (Kosen, 1989). In addition, the computational framework necessary for refinement and back-calculating PREs from structures has been developed making it feasible to analyze the structural aspects of dynamic processes in macromolecules (Clare and Iwahara, 2009). In PRE, the unpaired electron of the paramagnetic center influences the relaxation of nearby nuclear spins within a certain radius (up to 35 Å, depending on the paramagnetic probe) through dipolar interactions and causes them to relax faster. PRE is highly sensitive to distance changes as it has a steep distance dependence ( $r^{-6}$ ), making it a versatile tool for studying macromolecular structure and dynamics. The paramagnetic centers in PRE can be classified into two groups - isotropic and anisotropic. Probes with isotropic g-tensors such as nitroxide spin

labels or EDTA-Mn<sup>2+</sup> are commonly used because they lack pseudo-contact shifts (PCS) and other unfavorable relaxation properties observed in the anisotropic probes, making their PRE analysis relatively simple. Such isotropic paramagnetic probes have been widely used in structural characterization of various soluble proteins, membrane proteins, protein-oligosaccharide complexes, protein-protein complexes and protein-nucleic acid complexes (Clore and Iwahara, 2009; Iwahara et al., 2007, and references therein). In addition, PRE has been used to study large-scale dynamic changes in non-specific protein-DNA interactions (Iwahara et al., 2004a, 2006), inter domain motions (Baker et al., 2006; Tang et al., 2007) and transient protein-protein associations (Hansen et al., 2003; Tang et al., 2008a,b). Moreover, PRE has also been used to detect transient intermediates or lowly populated states in macromolecular binding events that are invisible to other biophysical techniques (Iwahara and Clore, 2006; Tang et al., 2006).

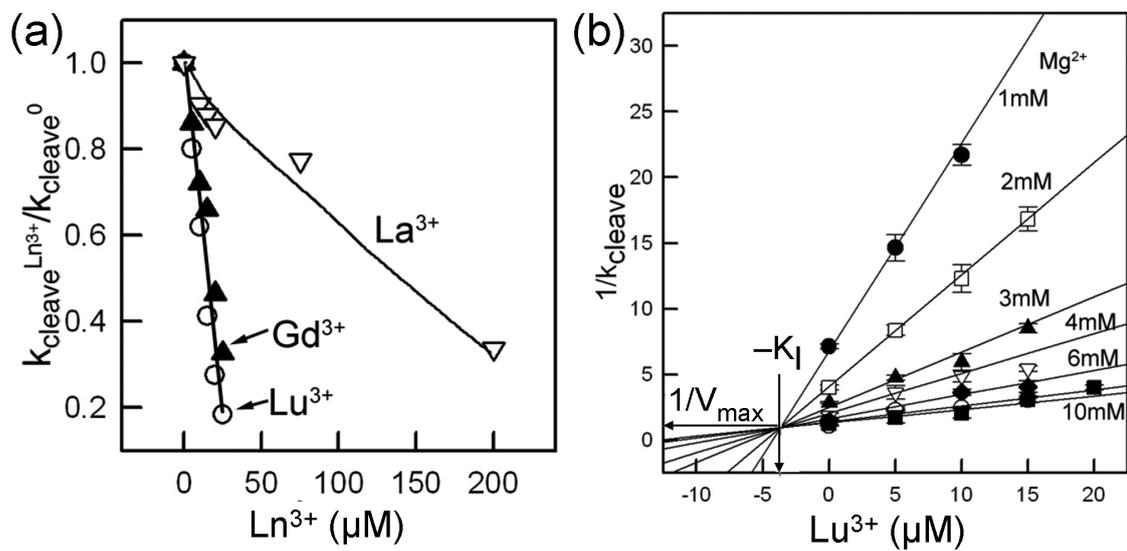
## Chapter 2

# Lanthanide ions as substitutes for Magnesium

EcoRV, like other restriction endonucleases, depends on  $Mg^{2+}$  for its catalytic activity. However, we cannot use  $Mg^{2+}$  to study the ternary EcoRV-DNA-metal complex as it would lead to DNA cleavage. Thus, there was a need to identify metal ions that do not support  $Mg^{2+}$  dependent DNA cleavage in EcoRV. Lanthanide ( $Ln^{3+}$ ) ions were tested as potential inhibitors of  $Mg^{2+}$ -dependent cleavage in EcoRV-DNA complex. In this chapter, it is shown that  $Ln^{3+}$  are excellent substitutes for  $Mg^{2+}$  for studying metal-ion induced changes in EcoRV-DNA complex. The cleavage-inhibition assays and DNA binding studies done previously (Jen-Jacobson lab) show that  $Ln^{3+}$  ions inhibit  $Mg^{2+}$  dependent DNA hydrolysis in EcoRV and enhance the DNA binding affinity of EcoRV. In addition, overlays of NMR spectra of cleaved complexes with  $Lu^{3+}$  and  $Mg^{2+}$  and NMR titration experiments, done previously in our lab (Dr. Kaustubh Sinha), suggest that  $Lu^{3+}$  binds in the same site and with the same stoichiometry (2 per monomer) as  $Mg^{2+}$ . Further, to verify that  $Lu^{3+}$  binds in the active site and not any other location in the complex, we solved crystal structures of EcoRV-DNA- $Lu^{3+}$  complex with cleaved and uncleaved DNA. The two crystal structures show that  $Lu^{3+}$  binds only in the active site and occupies a site similar to  $Mg^{2+}$ .

## 2.1 $\text{Ln}^{3+}$ inhibit $\text{Mg}^{2+}$ dependent DNA hydrolysis in EcoRV

Preliminary characterization of inhibition of EcoRV activity by lanthanides was done by Yi Wang in our lab. Using oligonucleotide-cleavage assays, it was shown that none of the lanthanide ions support catalytic activity of EcoRV (unpublished). Furthermore, cleavage-inhibition assays suggest that  $\text{Ln}^{3+}$  ions competitively inhibit  $\text{Mg}^{2+}$ -dependent DNA cleavage in EcoRV and bind at the active sites (unpublished, Jen-Jacobson lab). The smaller  $\text{Ln}^{3+}$  ( $\text{Lu}^{3+}$  and  $\text{Gd}^{3+}$ ) are better inhibitors than the larger  $\text{La}^{3+}$  (Figure 2.1 on page 17). Their calculated  $K_i$  values range from  $3.7 \mu\text{M}$  for  $\text{Lu}^{3+}$  to  $15.7 \mu\text{M}$  for  $\text{La}^{3+}$  (Table 2.1 on page 18).  $\text{Mg}^{2+}$  has an apparent  $K_D \sim 3.2 \text{ mM}$  whereas  $\text{Ca}^{2+}$  has an apparent  $K_D \sim 4.2 \text{ mM}$ . Thus, lanthanides bind EcoRV-DNA complex with much higher affinity than  $\text{Mg}^{2+}$  and inhibit DNA hydrolysis.



**Figure 2.1: Competitive inhibition of DNA cleavage in EcoRV by lanthanides (unpublished, Jen-Jacobson lab).** (a) The ratio of first-order cleavage rate constant with  $\text{Ln}^{3+}$  ( $k_{\text{cleave, Ln}^{3+}}$ ) to that without metal ( $k_{\text{cleave, 0}}$ ) is plotted as a function of  $\text{Ln}^{3+}$  concentration. (b)  $\text{Lu}^{3+}$  competitively inhibits  $\text{Mg}^{2+}$ -catalyzed DNA cleavage. The intersecting lines in the Dixon plot indicate competitive inhibition with  $K_i = 3.7 \mu\text{M}$  and  $V_{\text{max}} = 1.2 \text{ s}^{-1}$ .

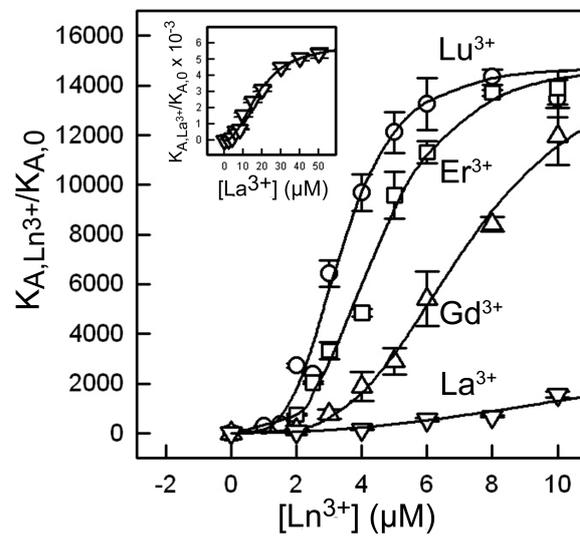
**Table 2.1: Binding and cleavage inhibition constants of lanthanides in EcoRV-DNA complex (unpublished, Jen-Jacobson lab).**

Ln <sup>3+</sup>	Ionic radius (Å) <sup>a</sup>	K <sub>i</sub> (μM) <sup>b</sup>	V <sub>max</sub> (s <sup>-1</sup> ) <sup>b</sup>	K <sub>D</sub> (μM) <sup>c</sup>	Hill coefficient (n) <sup>c</sup>
Lu <sup>3+</sup>	0.86	3.7 ± 0.9	1.2 ± 0.6	3.4 ± 0.2	3.8 ± 0.6
Er <sup>3+</sup>	0.89	6.1 ± 1.6	1.3 ± 0.7	4.5 ± 0.2	3.7 ± 0.7
Gd <sup>3+</sup>	0.94	7.2 ± 0.5	1.1 ± 0.2	7.2 ± 0.3	3.6 ± 0.4
La <sup>3+</sup>	1.03	15.7 ± 0.9	1.2 ± 0.2	18.5 ± 1.3	2.1 ± 0.2

<sup>a</sup> Effective ionic radii (6-coordinate). Mg<sup>2+</sup> ionic radius is 0.72 Å; Ca<sup>2+</sup> is 1.0 Å.

<sup>b</sup> Values derived from cleavage-inhibition assays in Figure 2.1 on page 17.

<sup>c</sup> K<sub>D</sub> is the apparent dissociation constant for Ln<sup>3+</sup> from the EcoRV-DNA-Ln<sup>3+</sup> complex; the nth root of K<sub>D</sub> is equivalent to K<sub>0.5</sub>, the [Ln<sup>3+</sup>] that produces half occupation of the [Ln<sup>3+</sup>] binding sites. Values of K<sub>D</sub> and n were obtained by fits to the binding data in Figure 2.2 on page 19 to the Hill equation. If Ln<sup>3+</sup> binding is completely cooperative, n indicates the number of Ln<sup>3+</sup> binding sites.

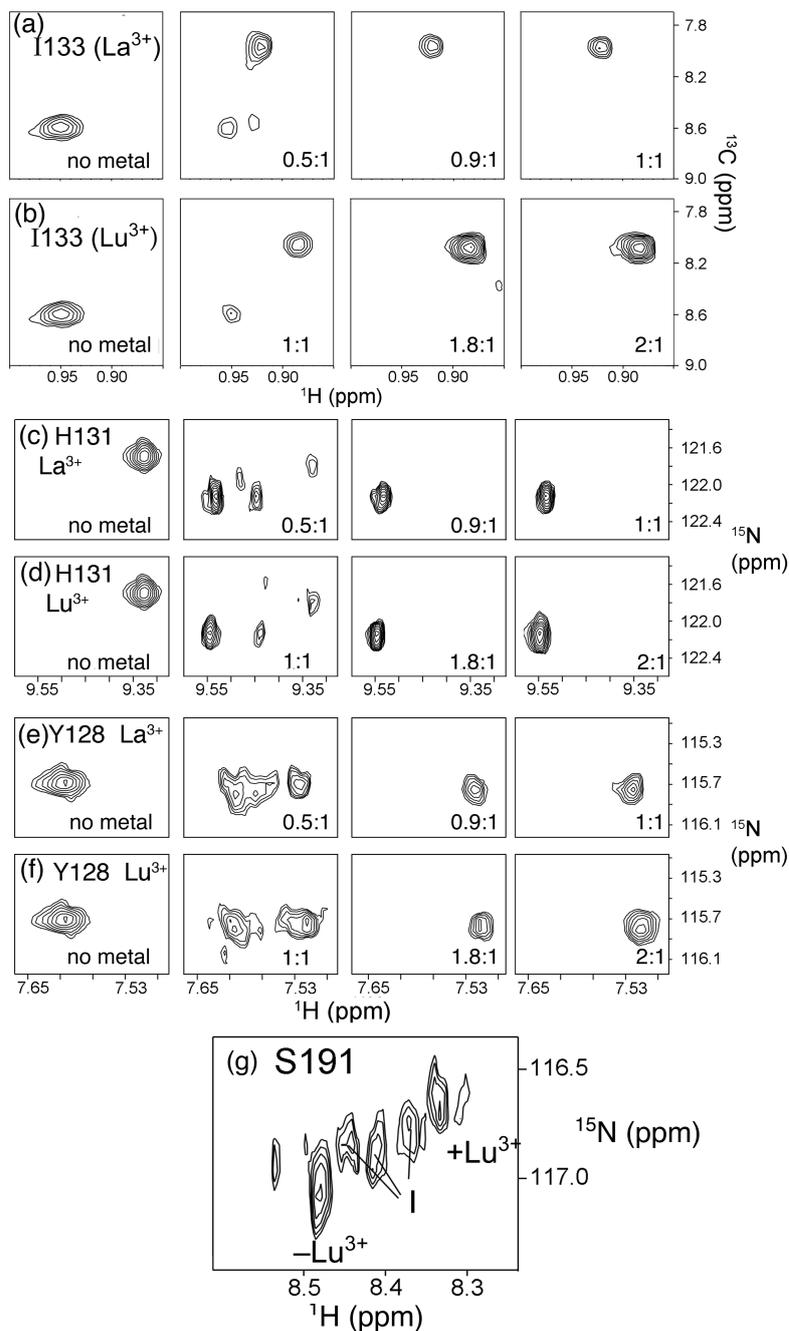


**Figure 2.2: Stimulation of EcoRV-DNA affinity by lanthanides (unpublished, Jen-Jacobson lab).**  $\text{Ln}^{3+}$  strongly enhance EcoRV binding to cognate DNA. The sigmoidal curves (ratio of  $K_A$  in the presence of  $\text{Ln}^{3+}$  to  $K_A$  in the absence of  $\text{Ln}^{3+}$ ) suggest  $\text{Ln}^{3+}$  binding to the EcoRV–DNA complex is cooperative.

## 2.2 DNA affinity and binding stoichiometry

Lanthanide ions bind in the active site, neutralizing the electrostatic repulsion among the negatively charged active-site residues and the scissile phosphate, thus stimulating the DNA binding affinity of EcoRV. This makes it feasible to measure the apparent dissociation constant as well as the binding stoichiometry for each metal ion by measuring the increase in DNA affinity. Binding studies (unpublished, Jen-Jacobson lab) show that saturating concentrations of  $\text{Lu}^{3+}$ ,  $\text{Er}^{3+}$  or  $\text{Gd}^{3+}$  increase the  $K_A$  of EcoRV for its cognate DNA by ~15,000 fold ( $\Delta\Delta G^\circ = -5.7$  kcal/mol), whereas saturating  $\text{La}^{3+}$  has a weaker effect (~6000 fold;  $-5.1$  kcal/mol) as seen in Figure 2.2 on page 19. The sigmoidal dependence of DNA binding on  $\text{Ln}^{3+}$  concentration indicates cooperativity for all  $\text{Ln}^{3+}$ . The apparent  $K_D$  values and Hill coefficients ( $n$ ) for each ion were determined by fitting the binding data (Figure 2.2 on page 19) to the Hill equation. They are in good agreement with the  $K_i$  values calculated from cleavage-inhibition assays (Table 2.1 on page 18). The Hill coefficients ( $n$ ) suggest binding of 4 ions/dimer for  $\text{Lu}^{3+}$  and  $\text{Gd}^{3+}$  whereas 2 ions/dimer for the larger  $\text{La}^{3+}$ . This was verified by NMR titration studies done previously in the lab, where EcoRV-DNA complexes were titrated with  $\text{Lu}^{3+}$  or  $\text{La}^{3+}$  ions and the effect of metal ion concentration on the NMR spectra was observed (unpublished, Dr. Kaustubh Sinha). 1  $\text{La}^{3+}$  per monomer is necessary to complete the metal-induced chemical shift changes whereas 2  $\text{Lu}^{3+}$  ions are required per monomer for saturation (Figure 2.3 on page 22). Because metal ion binding is in the slow exchange regime, distinct crosspeaks are observed at fixed chemical shifts, independent of metal ion concentration. In most cases, for both methyl resonances and amide resonances, only the crosspeaks from the species with no metal and with fully occupied sites were observed. However, for a few amide resonances, additional spectral peaks were observed representing intermediate states that arise when fewer than 4  $\text{Lu}^{3+}$  ions are bound per dimer (Figure 2.3 on page 22), suggesting that these residues experience distinct environments in the partially-saturated complexes. Interestingly, S191 shows three intermediate peaks,

which is consistent with the binding of 4 Lu<sup>3+</sup> ions per dimer. These NMR observations thus fully support Lu<sup>3+</sup> and La<sup>3+</sup> stoichiometries of 4 and 2 ions per dimer respectively as inferred from the Hill coefficients (Table 2.1 on page 18).



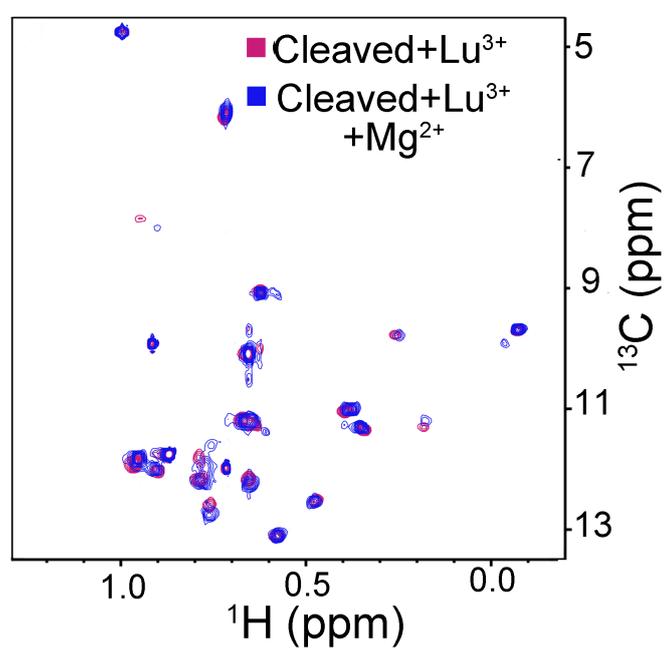
**Figure 2.3: Stoichiometries of lanthanide ion binding (unpublished, Dr. Kaustubh Sinha).** Residues near metal-ion binding site were inspected (a) and (b) Resonance peaks from I133- $\delta$ -CH<sub>3</sub> as the concentration of La<sup>3+</sup> or Lu<sup>3+</sup> is increased in the sample. The ratio Ln<sup>3+</sup>:EcoRV monomer is indicated in each panel. No additional changes are observed after 1 La<sup>3+</sup> or 2 Lu<sup>3+</sup> bind per monomer. (c)-(f) Similar titrations for the <sup>1</sup>H-<sup>15</sup>N amide resonances of H131 and Y128. (g) S191 amide peaks at 1:1 Lu<sup>3+</sup>:EcoRV monomer. In addition to resonances of the metal-free and the (Lu<sup>3+</sup>)<sub>4</sub>-EcoRV-DNA complexes, intermediate species (I) with 1, 2 or 3 ions bound are observed. <sup>13</sup>C chemical shifts were referenced to TMS.

## 2.3 Ln<sup>3+</sup> bind in the active sites

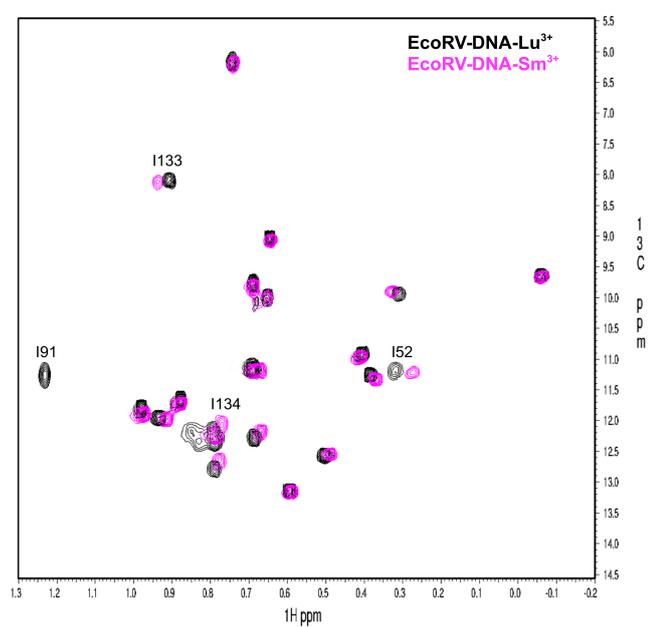
Competitive cleavage-inhibition assays suggest that Lu<sup>3+</sup> ions occupy the same binding sites as Mg<sup>2+</sup> (Figure 2.1 on page 17). This was also verified previously in the lab using NMR spectroscopy (unpublished, Dr. Kaustubh Sinha). Since it is not possible to get NMR spectra of the uncleaved complex with Mg<sup>2+</sup>, complexes of EcoRV with cleaved DNA were generated. The cleaved complexes were either saturated with Lu<sup>3+</sup>, or a mixture of 2 Lu<sup>3+</sup>/dimer to fill half the binding sites and 21 mM Mg<sup>2+</sup> (apparent K<sub>D</sub> of Mg<sup>2+</sup> ~ 3 mM) were added. The 2D-NMR spectra of the cleaved Lu<sup>3+</sup> complex and that of the cleaved Mg<sup>2+</sup>/Lu<sup>3+</sup> complex are overlaid (Figure 2.4 on page 25). Overlay shows that the two spectra are almost perfectly superimposable, which suggests that Lu<sup>3+</sup> binds the active site of EcoRV-DNA complex with the same stoichiometry as Mg<sup>2+</sup> (2 per monomer).

Paramagnetic centers with anisotropic susceptibility tensors ( $\chi$ ) give rise to changes in chemical shifts in NMR spectra called pseudocontact shifts (PCS). Nitroxides, Cu<sup>2+</sup>, Mn<sup>2+</sup> and Gd<sup>3+</sup> have isotropic  $\chi$  tensors whereas Co<sup>2+</sup> and all paramagnetic lanthanides (except Gd<sup>3+</sup>) have anisotropic  $\chi$  tensors. Paramagnetic relaxation enhancement (PRE) and pseudocontact shift (PCS) studies with paramagnetic lanthanides (Sm<sup>3+</sup>, Yb<sup>3+</sup> and Eu<sup>3+</sup>) were carried out to further verify active-site binding of Ln<sup>3+</sup> ions. Figure 2.5 on page 26 shows the overlay of methyl spectra of the Ile region of EcoRV-DNA complex with Lu<sup>3+</sup> (diamagnetic) and Sm<sup>3+</sup> (paramagnetic). Sm<sup>3+</sup> is the weakest paramagnetic ion among all lanthanide ions and shows PRE and PCS effects up to a smaller distance (10 Å). Consequently, only residues within 10 Å of Sm<sup>3+</sup> are affected, allowing us to determine its binding site. PCS effects are observable at greater distances than PRE because of their less steep distance dependence ( $r^{-3}$ ) compared to PRE ( $r^{-6}$ ), where  $r$  is the distance of the nuclear spin from the paramagnetic center. In Figure 2.5 on page 26, I91 which is ~5 Å from the active site, disappears in the spectra with Sm<sup>3+</sup> due to PRE. On the other hand, I52, I133 and I134 residues that are ~9 Å from the active site, move due to PCS effects. Other Ile residues show little or

no effect, confirming that Sm<sup>3+</sup> binds in the active site.



**Figure 2.4:  $\text{Lu}^{3+}$  and  $\text{Mg}^{2+}$  ions occupy the same sites (unpublished, Dr. Kaustubh Sinha).**  $^1\text{H}$ - $^{13}\text{C}$  HMQC spectrum of the Ile- $\delta$ - $\text{CH}_3$  region of the cleaved complex in the presence of 1:1  $\text{Lu}^{3+}:\text{Mg}^{2+}$ , overlaid on the spectrum of the  $\text{Lu}^{3+}$  saturated complex. Cross-peaks in the two spectra overlap well indicating that replacing  $\text{Mg}^{2+}$  with  $\text{Lu}^{3+}$  does not cause changes in local environments or the overall structure.  $^{13}\text{C}$  chemical shifts were referenced to TMS.



**Figure 2.5:  $\text{Ln}^{3+}$  ions bind in the active site.** Overlay of  $^1\text{H}$ - $^{13}\text{C}$  HMQC spectra of the Ile- $\delta$ - $\text{CH}_3$  region of EcoRV-DNA complex with diamagnetic  $\text{Lu}^{3+}$  (black) and paramagnetic  $\text{Sm}^{3+}$  (magenta). I91 disappears due to PRE whereas I52, I133 and I134 show pseudo contact shifts (PCS).  $^{13}\text{C}$  chemical shifts were referenced to TMS.

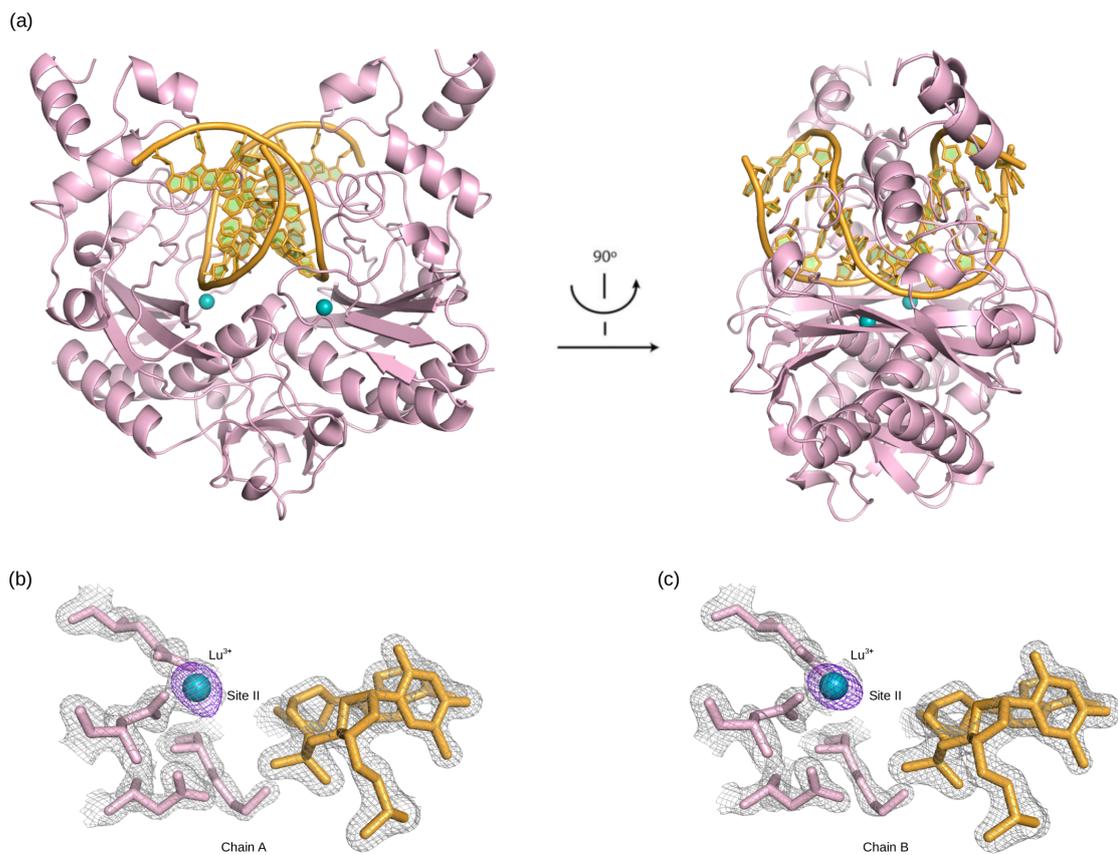
## 2.4 EcoRV-DNA-Lu<sup>3+</sup> complex crystal structures

In order to ascertain that Lu<sup>3+</sup> binds exclusively in the active sites, we solved crystal structures of the EcoRV-DNA-Lu<sup>3+</sup> complexes with substrate (uncleaved) and product (cleaved) DNA to 1.8 Å and 2.0 Å resolution respectively. Dr. Andy Kehr, during his graduate studies in our lab, helped with the refinement of the structure of the product complex. The data collection and refinement statistics are summarized in Table 2.2 on page 28. In both the structures, we did not observe any Lu<sup>3+</sup> binding outside the active sites. Anomalous difference maps (contoured at 10  $\sigma$  - uncleaved and 5  $\sigma$  - cleaved) confirm Lu<sup>3+</sup> binding in the active sites (Figure 2.6 on page 29 and Figure 2.7 on page 30). The crystal structure of the complex with substrate DNA has 1 Lu<sup>3+</sup> ion bound in each of the two active sites (Figure 2.6 on page 29). The structure of EcoRV-DNA-Lu<sup>3+</sup> complex with product DNA has 1 Lu<sup>3+</sup> ion bound in the chain A active site whereas the chain B active site has 3 Lu<sup>3+</sup> ions with different occupancies (Figure 2.7 on page 30). It is not possible for all three Lu<sup>3+</sup> sites to be occupied at the same time, indicating that there is variation in the number of bound Lu<sup>3+</sup> at this site throughout the lattice. Coordinates of the two structures have been deposited in the Protein Data Bank (PDB ID 5F8A - substrate complex and PDB ID 5HLK - product complex). In the structure with substrate DNA, the central TA step of the recognition site GATATC bends into the major groove by approximately 50°, a characteristic feature of the specific contact between EcoRV and its cognate sequence (Table 2.3 on page 31).

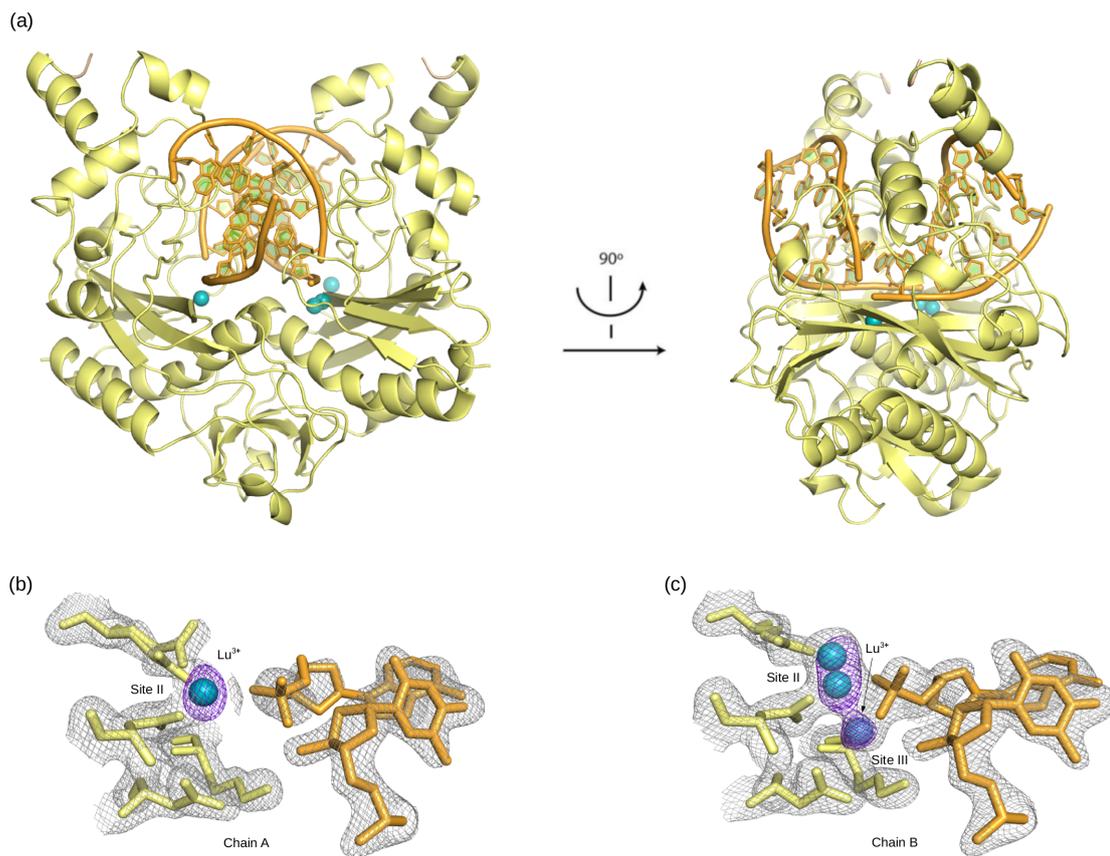
**Table 2.2: Data collection and refinement statistics for EcoRV-DNA-Lu<sup>3+</sup> complexes.**

	5F8A Substrate DNA	5HLK Product DNA
<b>Data collection</b>		
Space group	P1	P1
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	46.43, 52.86, 65.31	47.62, 48.33, 63.53
$\alpha$ , $\beta$ , $\gamma$ (°)	70.61, 73.24, 81.81	96.94, 108.78, 106.70
Resolution range (Å)	24.90-1.76 (1.85-1.76)	38.11-2.00 (2.05-2.00)
Wavelength (Å)	1.54178	1.54178
<i>R</i> <sub>sym</sub> or <i>R</i> <sub>merge</sub>	0.051 (0.196)	0.051 (0.224)
<i>I</i> / $\sigma$ <i>I</i>	14.7 (4.4)	20.3 (3.2)
Completeness (%)	93.0 (88.9)	84.3 (27.8)
Redundancy	3.9 (3.2)	4.3 (1.7)
<b>Refinement</b>		
Resolution (Å)	1.76	2.00
No. reflections	46648	26617
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.159 / 0.202	0.165 / 0.228
<i>R</i> <sub>free</sub> test set	10.71%	5.13%
<b><i>B</i>-factors</b>		
Wilson plot (Å <sup>2</sup> )	21.69	18.60
Overall (Å <sup>2</sup> )	27.61	22.44
<b>Ramachandran plot</b>		
Favored	464 (97.3%)	463 (96.3%)
Outliers	4 (0.84%)	3 (0.62%)
<b>R.m.s. deviations</b>		
Bond lengths (Å)	0.017	0.016
Bond angles (°)	1.876	1.780

Values in parentheses are for highest-resolution shell.



**Figure 2.6: Crystal structure of the EcoRV-DNA-Lu<sup>3+</sup> complex with substrate DNA (PDB ID 5F8A).** (a) Overall structure of the complex where protein dimer is bound to the cognate DNA in the presence of Lu<sup>3+</sup>. (b) and (c) Chain A and chain B active sites respectively with the electron density map ( $2F_o - F_c$ ) contoured at  $2\sigma$  (grey). Anomalous difference map is contoured at  $10\sigma$  (purple) and shows one Lu<sup>3+</sup> bound in site II in each active site.



**Figure 2.7: Crystal structure of the EcoRV-DNA-Lu<sup>3+</sup> complex with product DNA (PDB ID 5HLK).** (a) Overall structure of the cleaved complex in the presence of Lu<sup>3+</sup>. (b) and (c) Chain A and chain B active sites respectively with the electron density map (2F<sub>o</sub>-F<sub>c</sub>) contoured at 1 σ (grey). Anomalous difference map is contoured at 5 σ (purple) and shows the positions of Lu<sup>3+</sup> ions. Chain B active site has 2 Lu<sup>3+</sup> bound in site II (~2 Å apart) with 50 % occupancy each and 1 Lu<sup>3+</sup> in site III with 20% occupancy.

**Table 2.3: DNA conformation analysis in EcoRV-DNA complex crystal structures.**

PDB ID	Metal ion	Space group	DNA sequence <sup>a,b</sup>	Bend overall(°) <sup>c</sup>	Roll angle at TA (°)	Minor groove width at TA, Å
5F8A	Lu <sup>3+</sup>	P1	<u>AAAGATATCTTT</u>	55.5	55.5	10.6
1B95	None	P1	<u>AAAGATATCTT</u>	42.9	53.6	10.5
1B94	Ca <sup>2+</sup>	P1	<u>AAAGATATCTT</u>	43.4	57.1	10.4
1RVB	Mg <sup>2+</sup>	P1	<u>AAAGATATCTT</u>	43.1	54.4	10.2
1RVC	Mg <sup>2+</sup>	P1	<u>AAAGAT</u>  ATCTT	44.1	35.7	7.7
5HLK	Lu <sup>3+</sup>	P1	<u>AAAGAT</u>  ATCTTT	38.3	54	8.0

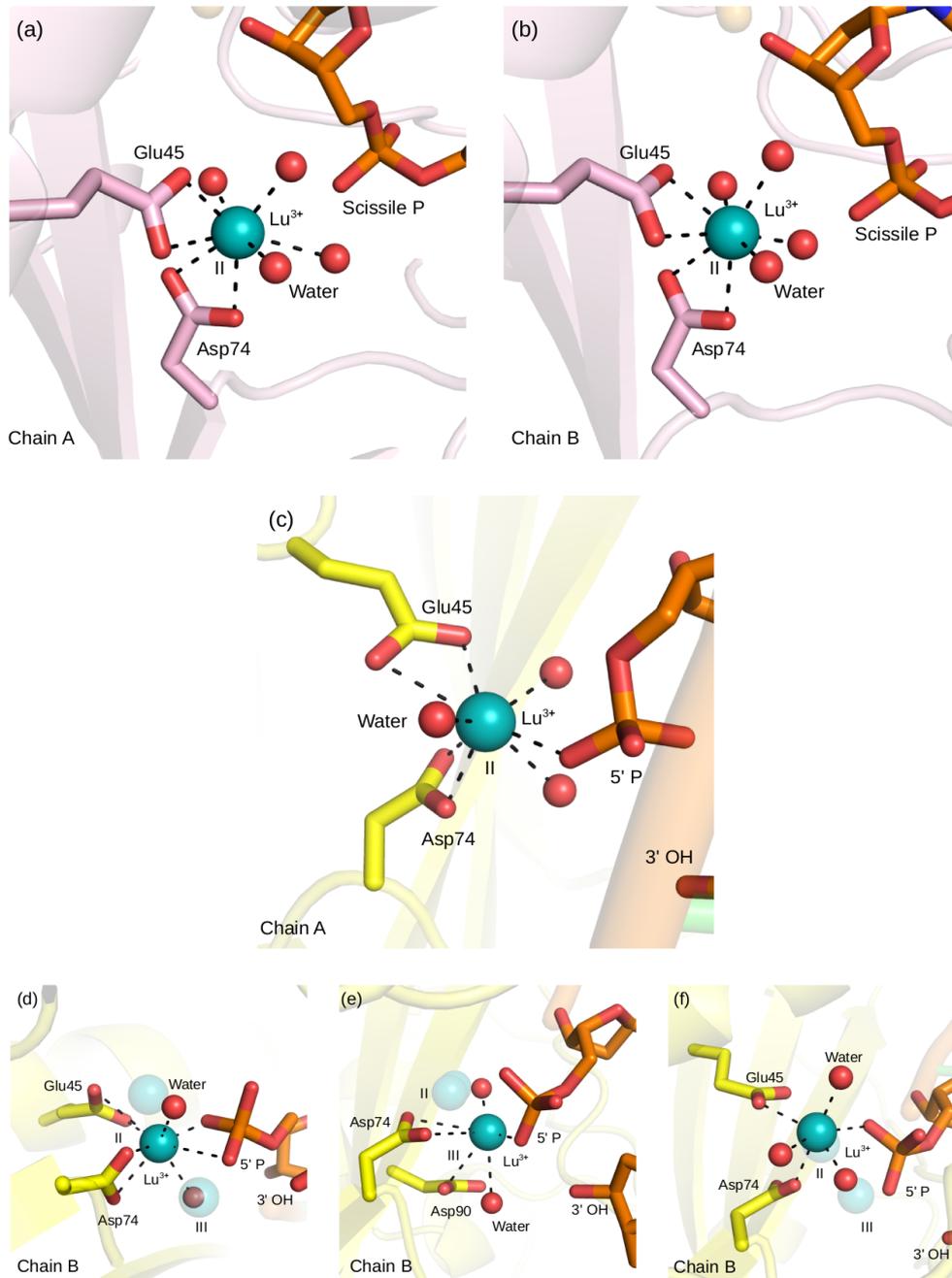
<sup>a</sup> Bases used to define the best-fit curvilinear axis are underlined.

<sup>b</sup> Vertical line at central TA step indicates DNA cleavage.

<sup>c</sup> DNA structural parameters were calculated using Curves+ (Lavery et al., 2009).

### 2.4.1 Local environment of Lu<sup>3+</sup> in the crystal structures

A survey of lanthanide-ions containing PDB structures, reports that octahedral geometry and bidentate ligation to carboxylate side-chains are the characteristic features of lanthanide coordination (Pidcock and Moore, 2001). The local environment of Lu<sup>3+</sup> in our EcoRV structures is typical of what has been observed for other lanthanide ions. Each Lu<sup>3+</sup> ion, in the two active sites of the EcoRV-substrate DNA complex (Figure 2.8 on page 33), occupies the Glu45/Asp74 site or site II (Horton and Perona, 2004). Both are coordinated by the two carboxylate oxygens each of Glu45 and Asp74 and by four oxygens from water molecules in an octahedral geometry. Similarly, the single Lu<sup>3+</sup> ion in chain A active site of the EcoRV-product complex (Figure 2.8 on page 33) occupies site II and binds in a bidentate way to the carboxylate oxygens of Glu45, Asp74 and to the three oxygens from water molecules and one of the oxygens of the 5' DNA phosphate. On the other hand, the chain B active site of the product complex has three Lu<sup>3+</sup> bound with different occupancies. One of them occupies site II (Glu45/Asp74) and the second Lu<sup>3+</sup> binds in Asp74/Asp90 site or site III (Horton and Perona, 2004). These site II and site III Lu<sup>3+</sup> were modeled with occupancies of 50% and 20% respectively. The site II Lu<sup>3+</sup> coordination sphere is completed by two oxygens each from Glu45, Asp74 and 5' phosphate group and by two oxygens from water molecules. Whereas, the site III Lu<sup>3+</sup> is coordinated by Asp74, Asp90, 5' DNA phosphate and three water molecules. The third Lu<sup>3+</sup> binds in a unique site, ~2 Å from site II, not seen in any other EcoRV-DNA-metal complex structures and was modeled with 50% occupancy. It is coordinated by two carboxylate oxygens of Glu45, one carboxylate oxygen of Asp74, one oxygen of 5' phosphate group and three oxygens from water molecules.



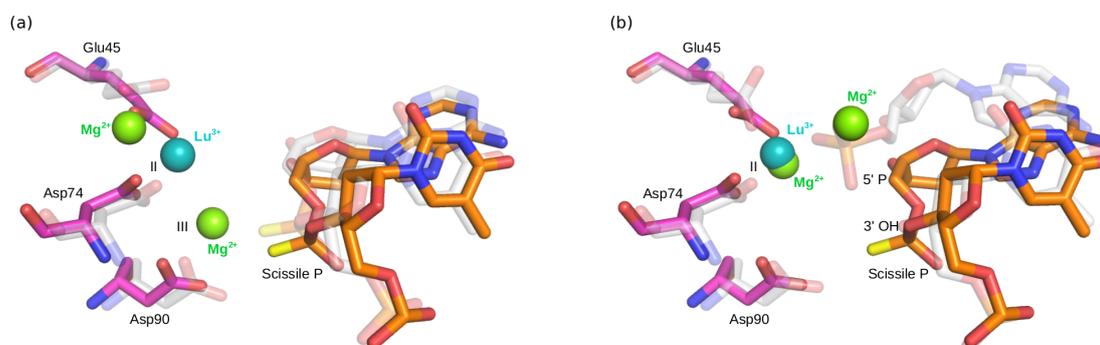
**Figure 2.8: Local environment of Lu<sup>3+</sup> in uncles and cleaved EcoRV-DNA-Lu<sup>3+</sup> complex crystal structures.** (a) and (b) Chain A and chain B active sites of the uncles complex (PDB ID 5F8A) respectively, showing Lu<sup>3+</sup> coordination to Glu45, Asp74 and 4 water molecules. (c) Chain A active site of the cleaved complex (PDB ID 5HLK) where Lu<sup>3+</sup> is coordinated to Glu45, Asp74, 3 water molecules and the 5' phosphate group. (d), (e) and (f) Chain B active site of the cleaved complex with 3 Lu<sup>3+</sup> ions bound with different occupancies. (d) Site II Lu<sup>3+</sup> (modeled with 50% occupancy) is coordinated to Glu45, Asp74, 2 water molecules and 5' phosphate group. (e) Site III Lu<sup>3+</sup> (modeled with 20% occupancy) is coordinated to Asp74, Asp90, 2 water molecules and 5' phosphate group. (f) The third Lu<sup>3+</sup> is ~2 Å away from site II and is coordinated to Glu45, Asp74, 3 water molecules and 5' phosphate group.

### 2.4.2 Comparison with EcoRV-DNA-Mg<sup>2+</sup> structures

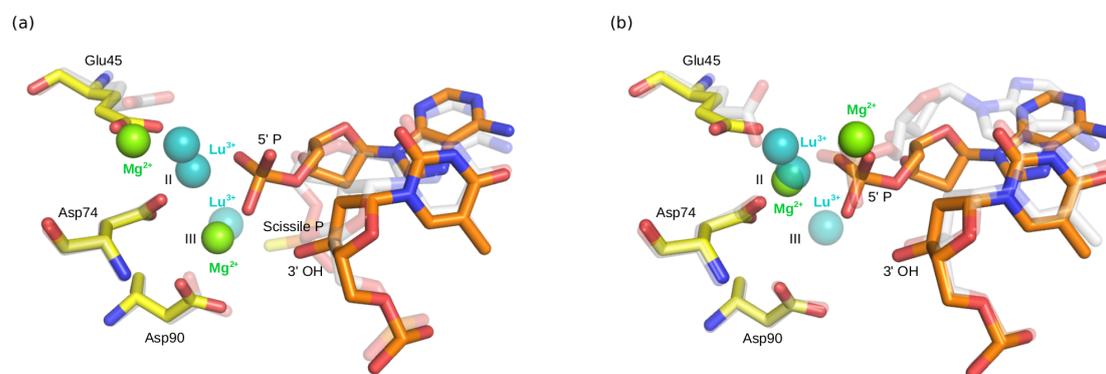
The crystal structure of EcoRV complexed with substrate DNA and Mg<sup>2+</sup> (PDB ID 1RVB) has 2 Mg<sup>2+</sup> ions bound in one of the two active sites. One of these Mg<sup>2+</sup> ions occupies Asp74/Asp90 site or site III (Horton and Perona, 2004) and is coordinated octahedrally by the scissile phosphodiester group, Asp74, Asp90 and three water molecules. In this structure, Mg<sup>2+</sup> was soaked into the crystals but it did not lead to DNA cleavage (Kostrewa and Winkler, 1995). On the other hand, the crystal structure of EcoRV-product DNA complex in the presence of Mg<sup>2+</sup> (PDB ID 1RVC) shows two metal ions bound identically in each of the two active sites. One of the Mg<sup>2+</sup> ion in each active site occupies Glu45/Asp74 site or site II (Horton and Perona, 2004) and is coordinated by Glu45, Asp74, 5' phosphate group and two water molecules. Superimposition of our uncleaved EcoRV-DNA-Lu<sup>3+</sup> structure (PDB ID 5F8A) with the structure of substrate complex with Mg<sup>2+</sup> (PDB ID 1RVB) shows Lu<sup>3+</sup> binds in a site intermediate between the two Mg<sup>2+</sup> ions (Figure 2.9 on page 36). However, when the substrate complex structure with Lu<sup>3+</sup> (PDB ID 5F8A) is superimposed with the product complex structure with Mg<sup>2+</sup> (PDB ID 1RVC), it is observed that Lu<sup>3+</sup> is coincident with the site II Mg<sup>2+</sup> of 1RVC (Figure 2.9 on page 36). In both superimpositions, the sidechains of Asp74 and Asp90 have very similar conformations but Glu45 shows ~2 Å shift in position. Similarly, the product complex structure with Lu<sup>3+</sup> (PDB ID 5HLK) when superimposed with the structure of substrate complex with Mg<sup>2+</sup> (PDB ID 1RVB) shows one of the Lu<sup>3+</sup> (20% occupancy) bound in a site coincident with the site III Mg<sup>2+</sup> of 1RVB (Figure 2.10 on page 37). In addition, an overlay of the product complex structures with Lu<sup>3+</sup> (PDB ID 5HLK) or Mg<sup>2+</sup> (PDB ID 1RVC) shows a second Lu<sup>3+</sup> ion occupying a site identical to the site II Mg<sup>2+</sup> (Figure 2.10 on page 37). Also, in these superimpositions, Asp74 and Asp90 sidechains have identical conformation but Glu45 shows a shift in position.

Although, DNA binding and NMR titration data suggest that two Lu<sup>3+</sup> bind per monomer, our crystal structure with the substrate complex shows only one Lu<sup>3+</sup> bound per monomer.

In the EcoRV structure with Mg<sup>2+</sup> (PDB ID 1RVB) that has two Mg<sup>2+</sup> ions per active site, only one Mg<sup>2+</sup> ion shows proper coordination geometry. The assignment of the second Mg<sup>2+</sup> in this structure is questionable since Mg<sup>2+</sup> has structure factors similar to a water molecule. An advantage of our Lu<sup>3+</sup> structure over the Mg<sup>2+</sup> structure is that it allows us to calculate anomalous difference maps so that we are certain of their presence and position. Possibly our crystal structure has only 1 Lu<sup>3+</sup> bound per active site because the crystal packing interactions do not allow the formation of optimal binding site for the second Lu<sup>3+</sup> ion whereas in solution 2 Lu<sup>3+</sup> bind per active site. Interestingly, our structure of the EcoRV product DNA complex (PDB ID 5HLK) has three Lu<sup>3+</sup> ions in one of the active site with different occupancy. Presence of 3 Lu<sup>3+</sup> ions, in addition to relatively poor electron density (for a 2.1 Å structure) of the active-site residues and DNA phosphate, suggest that we may have trapped a structure undergoing transition from uncleaved to cleaved DNA complex.



**Figure 2.9: Superimposition of the uncleaved or substrate EcoRV-DNA-Lu<sup>3+</sup> complex crystal structure (PDB ID 5F8A) with the structures of (a) substrate DNA (PDB ID - 1RVB) and (b) product DNA (PDB ID 1RVC) with Mg<sup>2+</sup>.** Since the two active sites are identical, only one is shown for simplicity. Lu<sup>3+</sup> structure is colored magenta (sidechains) and orange (DNA). Mg<sup>2+</sup> structure is colored gray (sidechains & DNA).



**Figure 2.10: Superimposition of the cleaved or product EcoRV-DNA-Lu<sup>3+</sup> complex crystal structure (PDB ID - 5HLK) with the structures of (a) substrate DNA (PDB ID - 1RVB) and (b) product DNA (PDB ID - 1RVC) with Mg<sup>2+</sup>. Only the Chain B active site of Lu<sup>3+</sup> structure is shown and it is colored yellow (sidechains) and orange (DNA). Mg<sup>2+</sup> structure is colored gray (sidechains & DNA).**

## 2.5 Conclusion

In this chapter, it is established that lanthanides ( $\text{Ln}^{3+}$ ) are versatile probes to study metal-induced changes in protein-DNA complexes. They do not support  $\text{Mg}^{2+}$  dependent catalysis in EcoRV, PvuII, Dicer, Klenow fragment and T4 DNA polymerase. Specifically  $\text{Lu}^{3+}$  is an excellent substitute for  $\text{Mg}^{2+}$  to study metal-induced structural and/or dynamic changes in EcoRV-DNA complex.  $\text{Lu}^{3+}$  competitively inhibits  $\text{Mg}^{2+}$  dependent DNA hydrolysis in EcoRV and stimulates EcoRV-DNA binding (cleavage-inhibition and DNA-binding assays, Jen-Jacobson lab).  $\text{Lu}^{3+}$  is similar in size to  $\text{Mg}^{2+}$  and binds with the same stoichiometry of 2 metal-ions per active-site as  $\text{Mg}^{2+}$  (NMR titration experiments, Dr. Kaustubh Sinha). Crystal structures of EcoRV-DNA- $\text{Lu}^{3+}$  complex with uncleaved (PDB ID 5F8A) and cleaved DNA (PDB ID 5HLK) show that  $\text{Lu}^{3+}$  exclusively binds in the active site and occupies a position similar to  $\text{Mg}^{2+}$ . In the next chapter, NMR assignments of EcoRV-DNA complex with and without  $\text{Lu}^{3+}$  are presented.

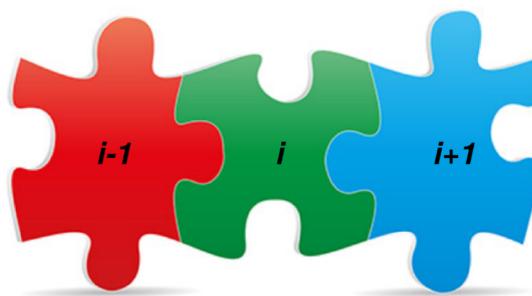
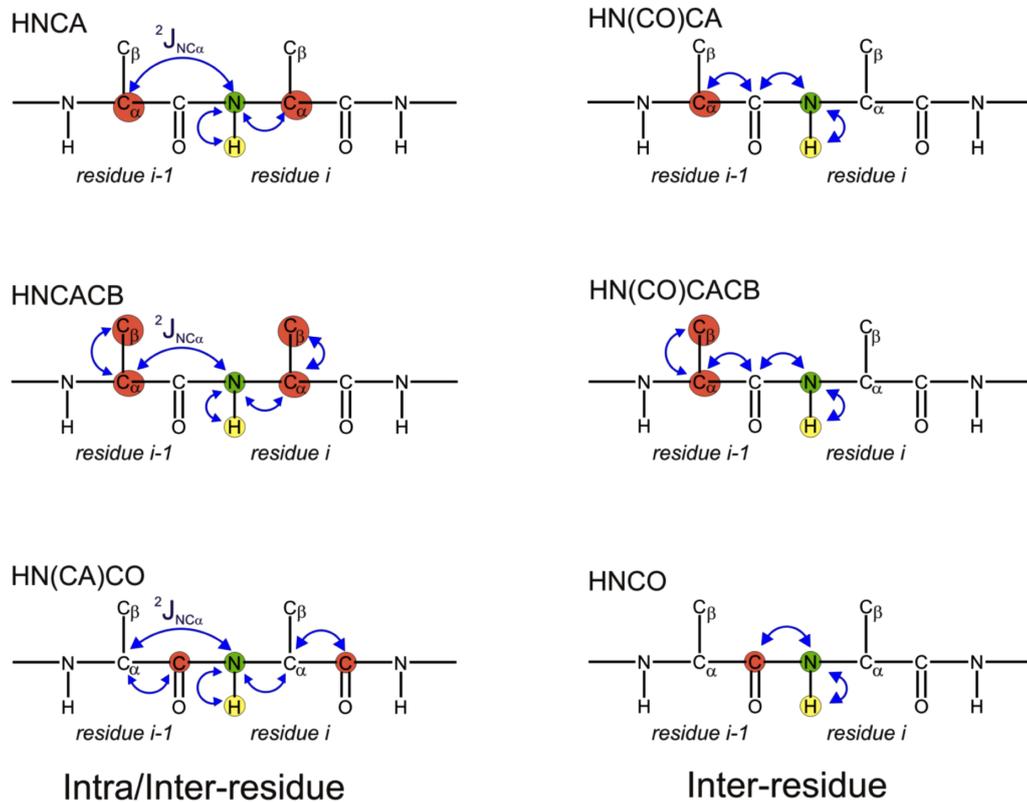
## Chapter 3

# Backbone amide and methyl side-chain assignments of EcoRV

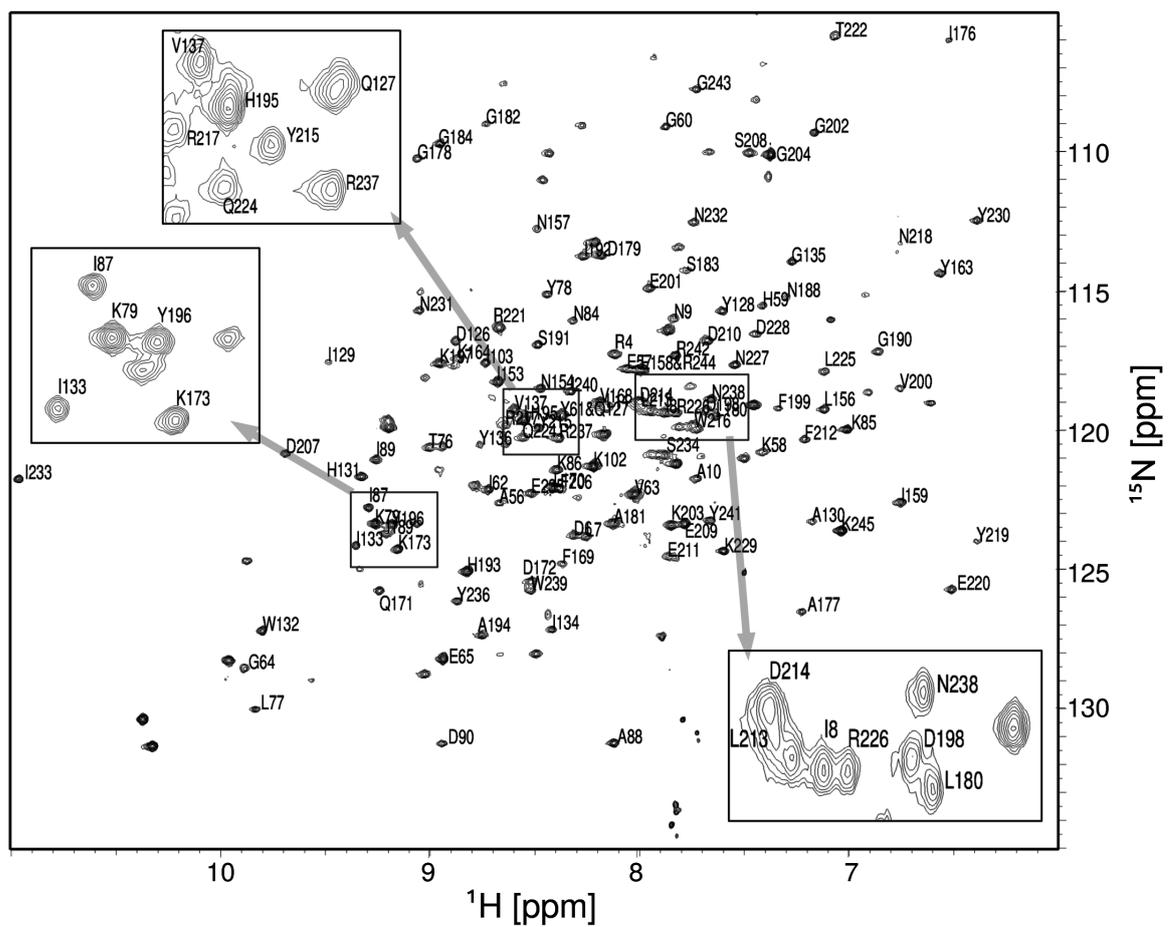
Pre-requisite for any in depth study by NMR spectroscopy is the assignment of the resonance peaks in the spectrum to the specific protein atoms. Especially in our studies, where we are interested in mapping metal-induced structural and/or dynamic changes in the EcoRV-DNA complex, the assignment of NMR resonance peaks with a good overall coverage is of central importance. For a comprehensive analysis of the effect of metal ions on the dynamics of EcoRV-DNA complex, we chose to focus on both mainchain amides and methyl side-chains. Backbone amides were previously assigned in the lab by Dr. Kaushtubh Sinha, where he employed traditional correlation experiments to assign NH resonances with and without  $\text{Lu}^{3+}$  independently. To assign the side chain methyl resonances, we used a combination of approaches - correlation NMR experiments, site-directed mutagenesis, paramagnetic relaxation enhancement (PRE) and 4D methyl-methyl nuclear overhauser enhancement (NOE) spectroscopy.

## 3.1 Backbone NH assignments

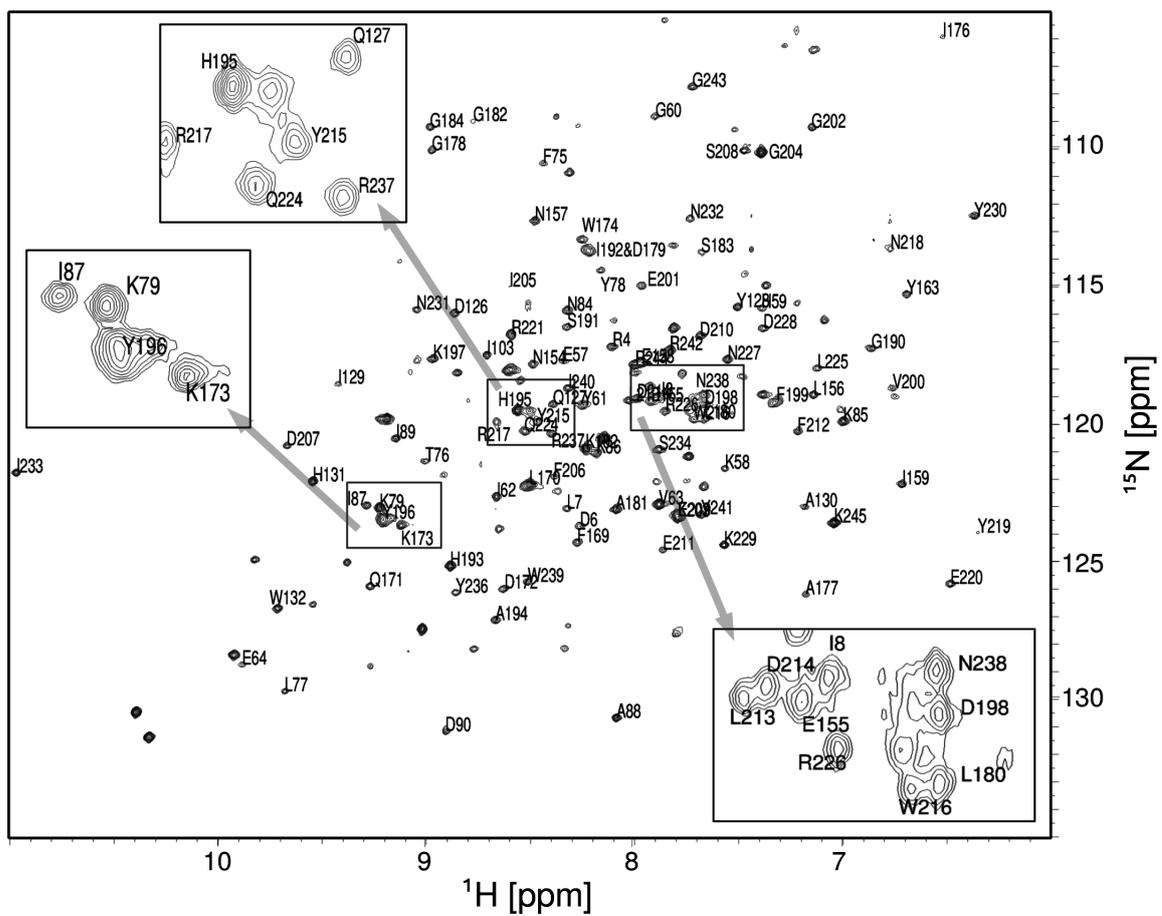
Mainchain NH resonances were assigned previously in the lab using the traditional triple resonance experiments (unpublished, Dr. Kaustubh Sinha). The assignment process involves correlating inter-residue resonances sequentially and is analogous to putting pieces of a jigsaw puzzle together as illustrated in Figure 3.1 on page 41. 170/245 NH resonances are observable in the  $^{15}\text{N}$ - $^1\text{H}$  spectrum and 120/170 (~70%) of these have been assigned for EcoRV-DNA complex with and without  $\text{Lu}^{3+}$  (Figure 3.2 on page 42 and Figure 3.3 on page 43). A number of amide resonances are missing in the spectrum because (a) they are part of the solvent exposed regions, not observable at pH 7.4 due to chemical exchange with the solvent, (b) they belong to protein regions that directly contact the DNA and are line broadened from dipolar interaction with the protonated DNA, (c) they were not protonated after growth in  $\text{D}_2\text{O}$  and (d) they are dynamic and are undergoing exchange broadening due to conformation exchange. As a result, a small number of the observable resonance peaks remain unassigned due to missing connectivities to the residues whose amide resonances are not observed. In spite of the incompleteness, the assigned NH resonances give a good coverage of the protein which is essential for studying dynamics (Figure 3.4 on page 44).



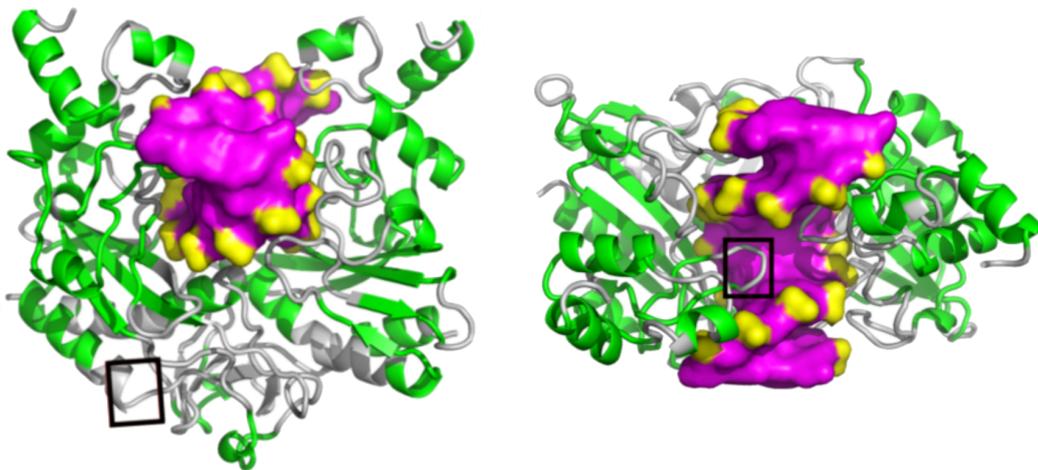
**Figure 3.1: Schematic of triple resonance experiments used for backbone NH assignments.** The arrows indicate “out and back” magnetization transfer from the amide proton. Combination of all these experiments was used to correlate resonances between  $i$  and  $i-1$  residues.



**Figure 3.2:**  $^1\text{H}$ - $^{15}\text{N}$  HSQC-TROSY spectrum of deuterated EcoRV-DNA complex without metal measured at 35°C and 600MHz.



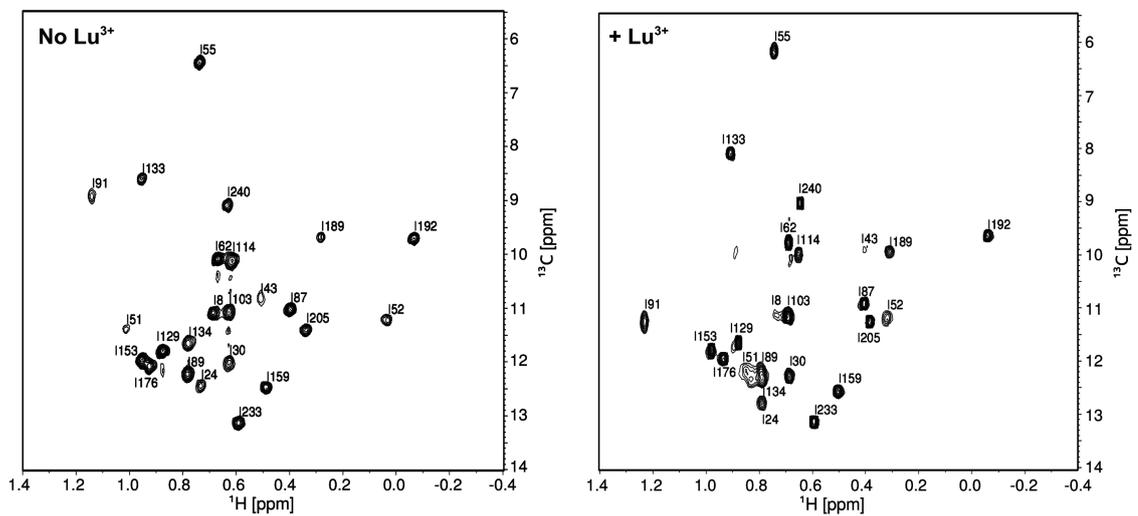
**Figure 3.3:**  $^1\text{H}$ - $^{15}\text{N}$  HSQC-TROSY spectrum of deuterated EcoRV-DNA- $\text{Lu}^{3+}$  complex measured at  $35^\circ\text{C}$  and 600MHz.



**Figure 3.4: Coverage map of assigned backbone NH in EcoRV-DNA complex.** Assigned mainchain residues are colored green. DNA is shown in magenta and the phosphates are yellow. Boxes indicate residues unobservable due to solvent exposure (left) or dipolar broadening from DNA (right).

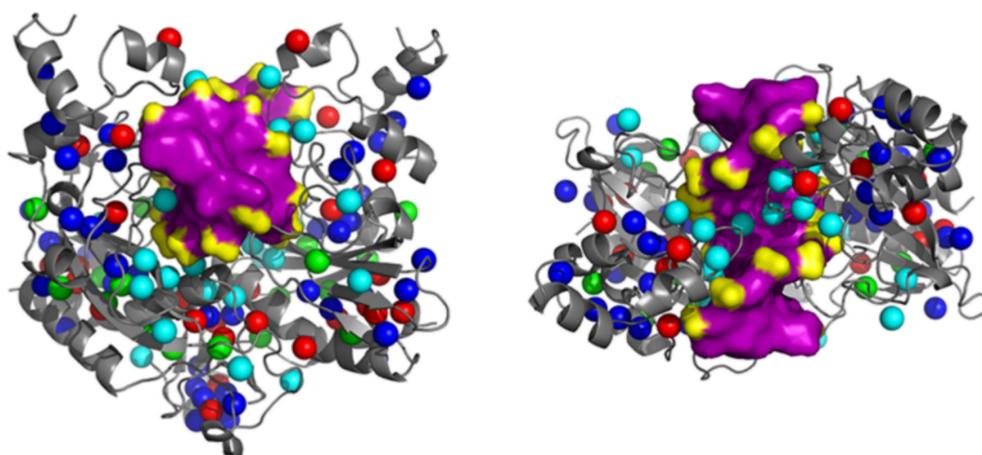
## 3.2 Methyl assignments

Methyl side-chains are buried in the protein core and also have favorable NMR spectral properties, making them excellent reporters of changes in dynamics (Igumenova et al., 2006; Ruschak and Kay, 2010). We focused on Ile, Val and Leu methyls as their labeling strategy in large, highly deuterated systems has been worked out well in previous studies (Gardner and Kay, 1997; Goto et al., 1999; Tugarinov et al., 2004; Tugarinov and Kay, 2005). Initially, the assignment was done using the conventional correlation spectroscopy experiments (Tugarinov and Kay, 2003). However, due to a number of missing backbone NH assignments (Figure 3.2 on page 42 and Figure 3.3 on page 43) the above strategy was insufficient in assigning all ILV methyls. Consequently, we had to employ a laborious approach of site-directed mutagenesis where each of the methyl ILV residues were mutated to a different residue. Conservative mutations (Ile→Val, Leu→Val and Val→Ala) were chosen to minimize the effect of mutation on the overall protein structure. PRE and 4D methyl-methyl NOE were used to assign additional ILV methyl resonances and to verify those assigned by correlation spectroscopy and site-directed mutagenesis. Using this multipronged approach, we obtained assignments for 90% of ILV methyls (24/26 Ile- $\delta$ -CH<sub>3</sub>, 12/14 Leu and 9/10 Val residues). Moreover, we were able to stereospecifically assign the pro-R-CH<sub>3</sub> and pro-S-CH<sub>3</sub> groups of Leu and Val residues by selectively labeling only the proS methyls with <sup>13</sup>CH<sub>3</sub> (Gans et al., 2010). The assignments of ILV methyls are shown in Figure 3.5 on page 46 and Figure 3.6 on page 47 and are also listed in the Appendix. The coverage map of assigned ILV methyls is shown in Figure 3.7 on page 48.



**Figure 3.5: Assignment of Ile- $\delta$ 1-CH<sub>3</sub>.**  $^1\text{H}$ - $^{13}\text{C}$  HMQC spectra of Ile- $\delta$ 1-CH<sub>3</sub> of the deuterated EcoRV-DNA complex with and without Lu<sup>3+</sup> measured at 35°C and 600MHz. Assigned resonance peaks are labeled and 24/26 Ile residues were assigned.  $^{13}\text{C}$  chemical shifts were referenced to TMS.





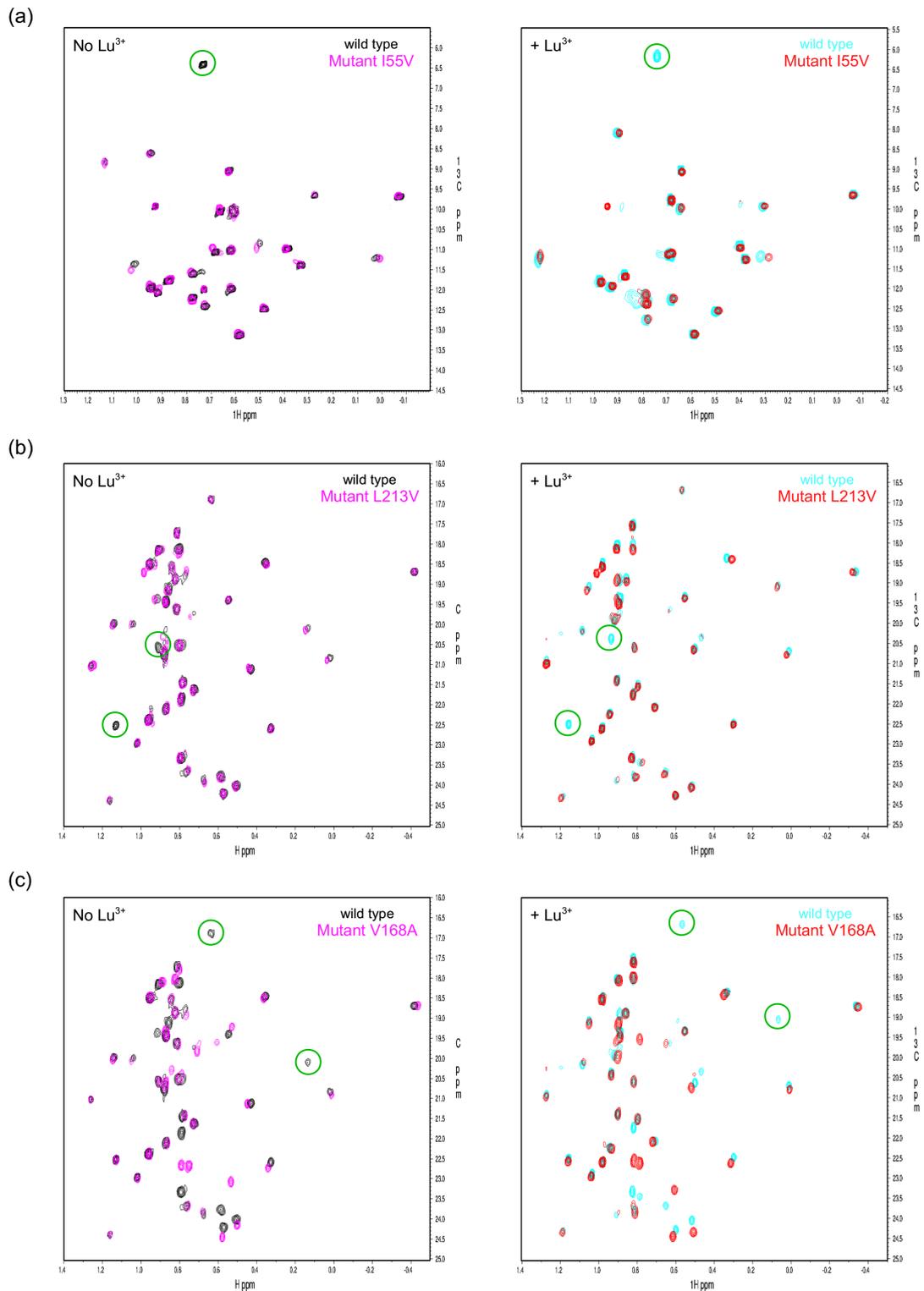
**Figure 3.7: Coverage map of assigned ILV-CH<sub>3</sub> in EcoRV-DNA complex.** Assigned Ile-δ1 methyls are shown as blue spheres. Assigned Val-γ2 and Leu-δ2 methyls are shown as green and red spheres respectively. Thr methyls are unassigned and shown as cyan spheres.

### 3.2.1 Correlation spectroscopy

Methyl side-chains are conventionally assigned by correlating the methyl resonances to the assigned backbone amide groups. In addition to the triple resonance experiments used for NH assignments, additional experiments (Materials and Methods) were performed to obtain correlations between the methyl resonances and the  $C\alpha$ ,  $C\beta$  and  $C\gamma$  resonances. However, incomplete mainchain NH assignments (Figure 3.2 on page 42 and Figure 3.3 on page 43) limited the use of correlation spectroscopy in assigning all ILV resonances and only a small number (11/50 ILV residues) were assigned using this method - I24, I91, I134, I159, I192, I205, I233, I240, V166, L7 and L77.

### 3.2.2 Site-directed mutagenesis

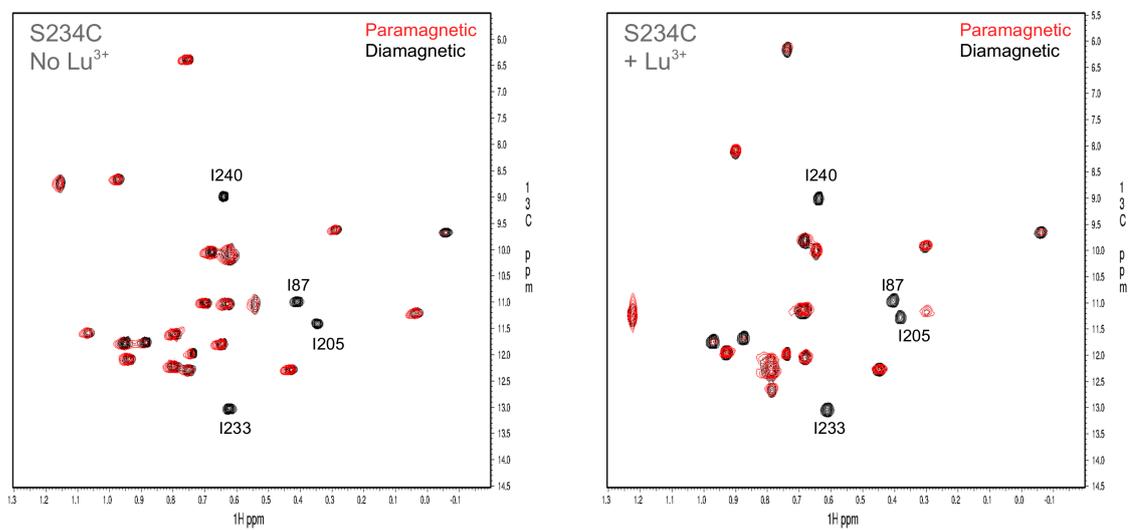
In order to supplement the correlation experiments and assign additional methyl resonances, we made single-site mutations for the following residues: (Ile→Val) - 8, 23, 30, 43, 51, 52, 55, 62, 87, 89, 103, 129, 133, 153, 176, 189, (Val→Ala) - 63, 122, 141, 137, 168, 175, 200 and (Leu→Val) - 148, 156, 170, 180, 213, 225. To assign them, the  $^1\text{H}$ - $^{13}\text{C}$  methyl spectra of the mutant proteins which showed a single missing resonance (Ile mutants) or a pair of missing resonances (Val or Leu mutants) was compared to the  $^1\text{H}$ - $^{13}\text{C}$  methyl spectra of the WT EcoRV-DNA complex with and without  $\text{Lu}^{3+}$ . Mutations resulted in minor changes in the remaining ILV resonances but that did not confound the assignment process. However, 4 of the Leu→Val mutant proteins (L11, L33, L40 and L46) precipitated during purification and 5 other mutant proteins (I114, I121, V20, V39 and L107) gave poor quality spectra. A few examples of assignment by site-directed mutagenesis for respective ILV mutants are shown in Figure 3.8 on page 50.



**Figure 3.8: Assignment of ILV methyls using site-directed mutagenesis.** Overlays of  $^1\text{H}$ - $^{13}\text{C}$  methyl spectra of wild type EcoRV-DNA complex with and without  $\text{Lu}^{3+}$  with that of ILV mutant complexes - (a) I55V, (b) L213V and (c) V168A. Green circles indicate the peaks missing in the mutant spectrum.  $^{13}\text{C}$  chemical shifts were referenced to TMS.

### 3.2.3 Paramagnetic relaxation enhancement (PRE)

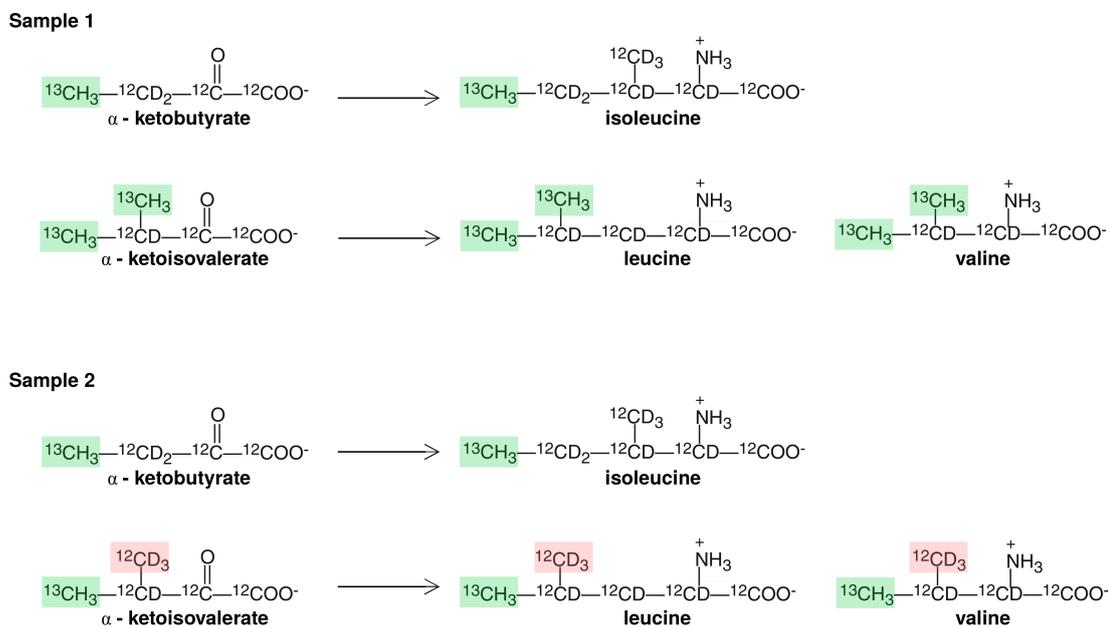
The availability of crystal structures of EcoRV-DNA complex with and without metal ions, facilitated the use of PRE and 4D NOE (Venditti et al., 2011) to assign additional ILV methyl resonances as well as corroborate the resonances that were assigned by correlation spectroscopy and site-directed mutagenesis. In PRE, a paramagnetic center (nitroxide spin-label in our case), carrying an unpaired electron is introduced into the complex by chemical conjugation. The unpaired electron influences the relaxation of the nearby nuclear spins (as measured by NMR spectroscopy) and causes them to relax faster. This PRE effect has a steep distance dependence ( $r^{-6}$ ) and causes the resonance peaks of residues that are close to the paramagnetic center (within 15 Å) to disappear in the NMR spectrum. Thus, based on the distances obtained from the crystal structure, the possible assignment of the disappeared peaks was narrowed down to a small set of ILV residues that are within 15 Å of the nitroxide labeling site. An example of the use of PRE to assign/verify ILV methyl side-chains is shown in Figure 4.9 on page 78, where the nitroxide was labeled at S234. PRE experiments were also used to study dynamics and are discussed in more detail in Chapter 4.



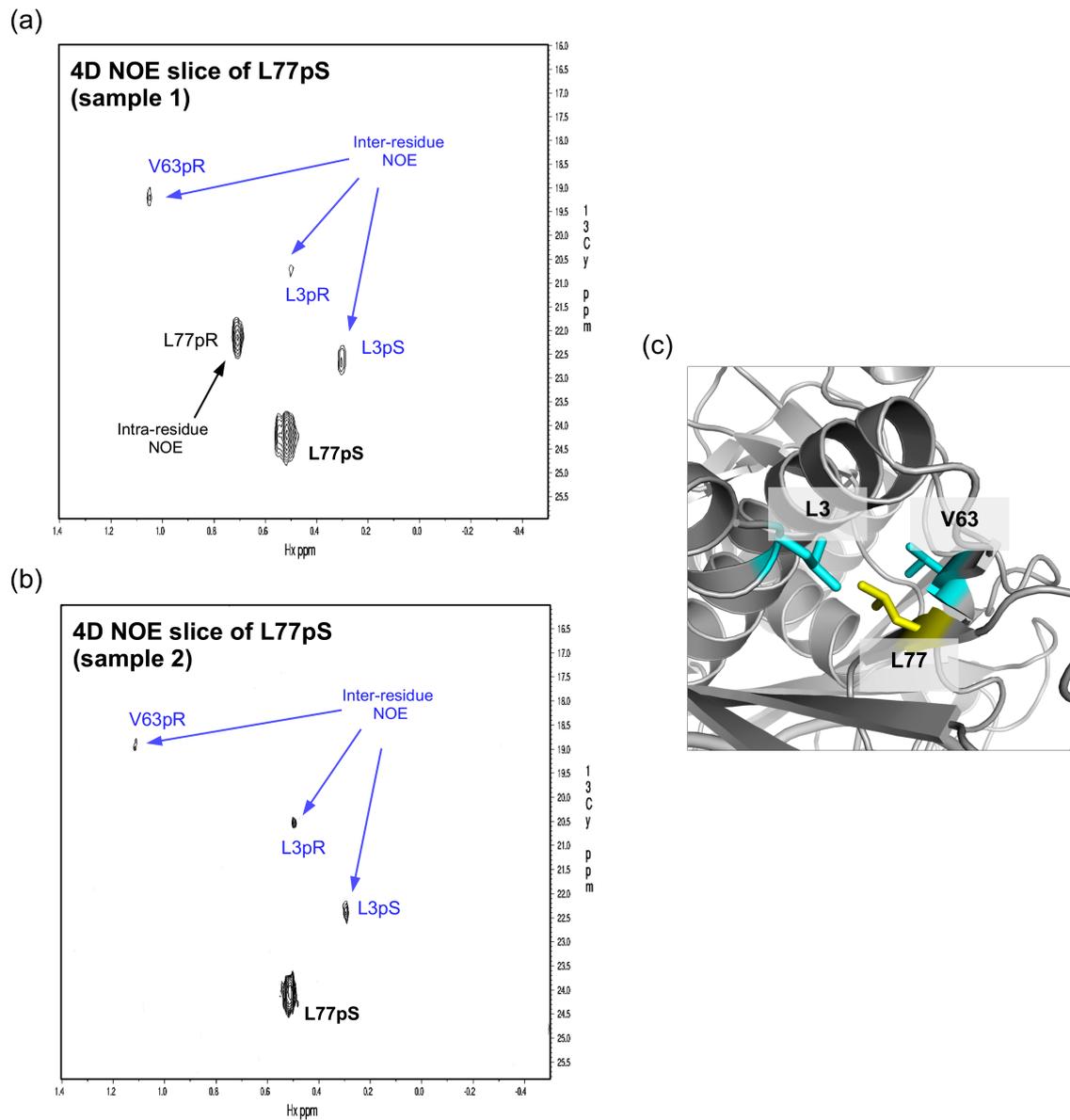
**Figure 3.9: PRE to verify ILV methyl assignments.** Overlays of paramagnetic and diamagnetic  $^1\text{H}$ - $^{13}\text{C}$  methyl spectra of EcoRV-DNA complex with and without  $\text{Lu}^{3+}$  with nitroxide labeled at residue 234. Methyl resonances of Ile residues that are within 15 Å of the nitroxide disappear in the paramagnetic sample. Assignment of these missing peaks were narrowed down to a small set of possible residues using the distances from the crystal structure.  $^{13}\text{C}$  chemical shifts were referenced to TMS.

### 3.2.4 4D methyl-methyl NOE

In the 4D methyl-methyl NOE experiments, the methyl side-chain residues are correlated through three-dimensional space. The 4D NOE spectrum of a particular methyl gives a strong self-peak but relatively weak cross-peaks for any methyl that is within 5-7 Å. The intensity of a methyl-methyl cross-peak depends on the distance between the methyls. Thus, with a known crystal structure, a particular methyl can be correlated to its corresponding neighbors. In addition, 4D NOE can also be used to identify methyl pairs of Leu and Val. When both proR and proS methyls of LV residues are simultaneously isotopically labeled, the NOE spectrum of each methyl group will show a strong intra-residue peak. For the 4D methyl-methyl NOE experiments of EcoRV-DNA complex, two different ILV  $^{13}\text{C}$ -methyl labeled samples were produced (Figure 3.10 on page 54). In the first sample, 4-  $^{13}\text{C}$ - $\alpha$ -ketobutyric acid and dimethyl- $^{13}\text{C}_2$ - $\alpha$ -ketoisovaleric acid were used as the precursors to label the ILV residues (Tugarinov and Kay, 2003). In this sample, both the proR and proS methyl groups of Leu & Val were  $^{13}\text{CH}_3$ -labeled to identify the methyl pairs, resulting in a strong intra-residue NOE peak in the 4D NOE spectrum. A second sample was made with 4-  $^{13}\text{C}$ - $\alpha$ -ketobutyric acid and 3-methyl- $^{13}\text{C}$ -3,4,4,4  $^2\text{H}_4$ - $\alpha$ -ketoisovaleric acid as precursors and resulted in Leu and Val side-chains labeled with  $^{13}\text{CH}_3$  and  $^{12}\text{CD}_3$  methyl groups. Intra-residue peaks were missing in the 4D NOE spectrum of this sample. This labeling strategy allows the detection of inter-residue NOEs over longer distances because with fewer vicinal protons the signal decay becomes more gradual. An example of the 4D NOE spectrum of L77pS is shown in Figure 3.11 on page 55, where the NOE of L77pS was used to identify its methyl pair (L77pR) and to assign/verify L3 and V63 residues.



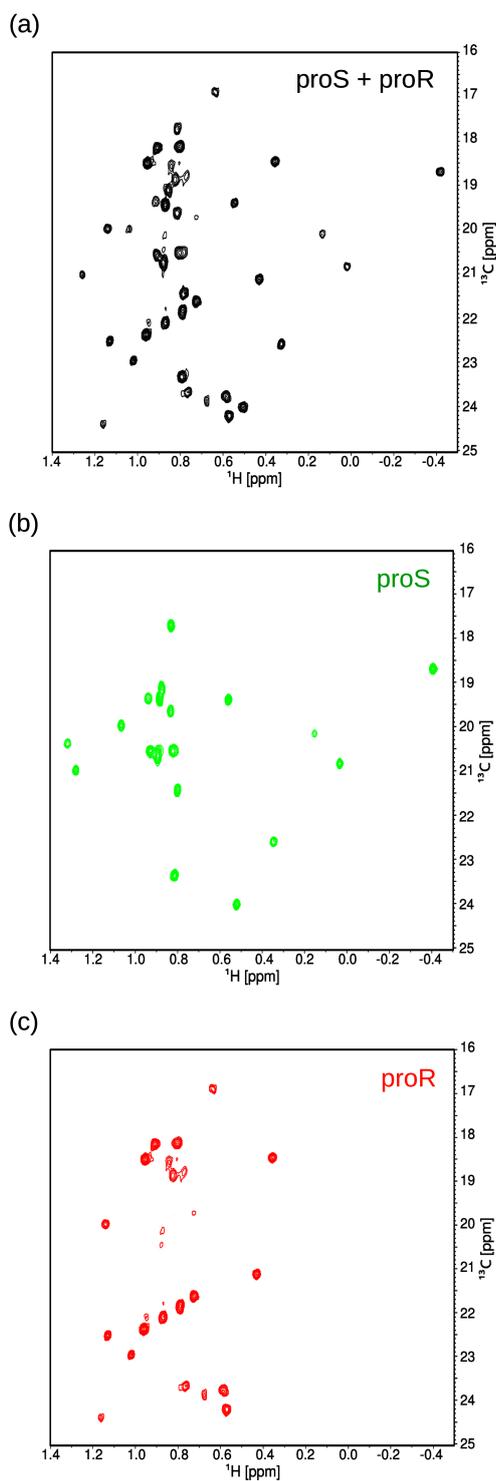
**Figure 3.10: Biosynthetic precursors used to make 4D methyl-methyl NOE samples.** Sample 1 - both methyl groups of Leu and Val were  $^{13}\text{CH}_3$ -labeled. Sample 2 - only one methyl group (50% proR & 50% proS) was  $^{13}\text{CH}_3$ -labeled.



**Figure 3.11: Example of assignment by 4D methyl-methyl NOE.** A slice of 4D NOE spectrum of L77pS. **(a)** Sample 1 - both methyl pairs were labeled, showing a strong intra-residue peak. **(b)** Sample 2 - only one methyl was labeled and so intra-residue peak is not observed. Peaks labeled in blue show inter-residue NOEs with L77pS. **(c)** L77 and its NOE partners (L3 and V63) in the EcoRV-DNA complex structure are shown as yellow and cyan sticks respectively.  $^{13}\text{C}$  chemical shifts were referenced to TMS.

### 3.2.5 Stereospecific assignments of Leu and Val

ProS methyl groups of Leu and Val were selectively  $^{13}\text{C}$  labeled by adding 2- $^{13}\text{CH}_3$ -methyl-4- $^2\text{H}_3$ -acetolactate during expression as described (Gans et al., 2010). In the  $^1\text{H}$ - $^{13}\text{C}$  HMQC spectrum of this sample, only the proS methyl resonances were observed (Figure 3.12 on page 57). A difference spectrum of the proS and proR  $^{13}\text{C}$ -methyl labeled sample and the proS only  $^{13}\text{C}$ -methyl labeled sample helped identify the proR resonances.



**Figure 3.12: Stereospecific methyl assignment of Leu and Val residues.** (a)  $^1\text{H}$ - $^{13}\text{C}$  HMQC spectrum of LV region where both proS and proR groups are  $^{13}\text{C}$ -labeled. (b)  $^1\text{H}$ - $^{13}\text{C}$  HMQC spectrum of LV region where only proS group was  $^{13}\text{C}$ -labeled using the labeling strategy of Gans et al., 2010. (c) Difference spectrum [(a)-(b)] to identify proR methyl groups.  $^{13}\text{C}$  chemical shifts were referenced to TMS.

### 3.3 Conclusion

120/170 backbone NH resonances and 45/50 ILV methyl side-chains have been assigned in the EcoRV-DNA complex with and without  $\text{Lu}^{3+}$ . Though not complete, together these assignments provide a good coverage of the complex for residue-specific analysis of metal-induced changes. In the next chapter, paramagnetic relaxation enhancement (PRE) to study structural and/or dynamic changes in EcoRV-DNA complex  $\pm \text{Lu}^{3+}$  is introduced and data from the PRE experiments are presented.

# Chapter 4

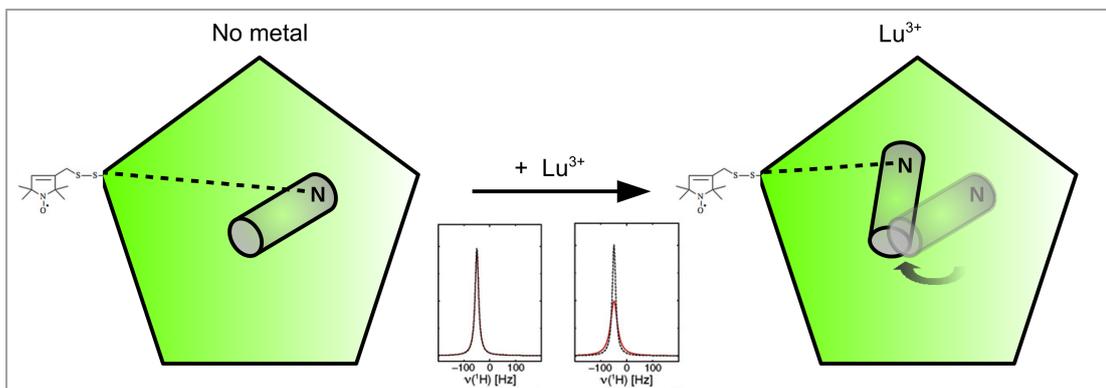
## Paramagnetic relaxation enhancement (PRE)

Paramagnetic relaxation enhancement (PRE) is a powerful tool to study collective domain motions of macromolecules or macromolecule complexes as it provides long-range distance information (up to 35 Å). In this chapter, first an overview of the use of PRE for studying metal-induced structural and dynamic changes in EcoRV-DNA complex is given. Second, our strategy for spin-labeling, identifying ideal background EcoRV mutant and the selection of spin-labeling sites are discussed. Third, the methods used for measuring PRE rates and back-calculating PREs from the crystal structures are explained. Lastly, the analysis of PRE data acquired for spin-label at S2C, K197C and S234C sites is presented.

### 4.1 Detecting domain motions by PRE

A schematic representation of how changes in the mean distance to the spin-label can be detected by PRE is shown in Figure 4.1 on page 61. A paramagnetic center is introduced in to the macromolecule of interest (protein in our case) by site-directed spin labeling (SDSL).

In the absence of ligand (metal ion in our case), if the nucleus (N) of the residue of interest lies outside the PRE range of the paramagnetic center, there will be no change in its intensity and the measured PRE rate will be essentially zero. However, when a ligand ( $\text{Lu}^{3+}$ ) is added and say it causes a collective motion of the domain, that our residue of interest is a part of, such that now it is within the PRE range of the paramagnetic center, its intensity will decrease ( $\Delta I \propto \frac{1}{r^6}$ ) and a PRE rate can be measured. The measured PRE rate can be used to approximate the new position that the domain has moved to because of ligand binding. Collective domain motions are expected to show correlated changes in measured PRE rates of its residues due to ligand binding. On the other hand, local motions where a small number of residues that move in perhaps uncorrelated ways may also be detected. A change in PRE can be due to two, potentially interrelated factors: i) the mean distance from the NMR spin can change, i.e. a true structural change ii) A change in dynamics may increase (or decrease) close approaches of the NMR spin to the nitroxide. It is possible to have changes in the closest approach (e.g change in dynamics) without necessarily changing the mean position. However, it is not possible to separate these two possibilities based on PRE data alone and additional experiments are required.



**Figure 4.1:** Schematic of how PRE will be used to detect changes in mean distance to the spin-label upon  $\text{Lu}^{3+}$  binding in EcoRV-DNA complex. N is the nucleus of interest (NH or  $\text{CH}_3$ ) and the nitroxide spin-label is the paramagnetic center.

## 4.2 Spin-labeling

In our PRE experiments, nitroxide radicals were the obvious choice as paramagnetic probes because metal-ion based probes would interfere with our studies of EcoRV-DNA-metal complex by binding to the intrinsic metal ion binding sites on the protein. The most widely used method to incorporate spin labels in a protein is by conjugation to a solvent-exposed cysteine residue, which can be engineered using site-directed mutagenesis (Hubbell et al., 2000; Todd et al., 1989). There are several types of nitroxide spin labels commercially available that can be conjugated to cysteine. Most commonly used are thiol-specific nitroxide spin labels such as MTSL and MTS (Figure 4.2 on page 63). However, one draw-back of these thiol-based spin labels is that they are not very stable over long periods of time due to disulfide-exchange reactions (Iwahara et al., 2007). Though, other nitroxide labels such as iodo or bromoacetamide derivatives are relatively more stable (Gillespie and Shortle, 1997; Ogawa and McConnell, 1967), they have low specificity and can react slowly with methionine, histidine and lysine residues. Maleimide derivatives have also been reported to be unstable (Kosen, 1989). On the other hand, conjugation by disulfide chemistry is easy to perform and provides high site-specificity (Battiste and Wagner, 2000; Dvoretzky et al., 2002; Gaponenko et al., 2008). Thus, we used the thiol-based spin-labels, MTSL [1-oxy-2,2,5,5-tetramethyl-d3-pyrroline-3-methyl)-methanethiosulfonate] for our paramagnetic samples and MTS [1-acetoxy-2,2,5,5-tetramethyl-d3-pyrroline-3-methyl)-methanethiosulfonate] for the diamagnetic samples. To circumvent the stability problem, all data was acquired within a week of sample preparation. MTS is structurally similar to MTSL (Figure 4.2 on page 63) and therefore is a better diamagnetic reference sample than the unlabeled protein. Although, some studies have also used ascorbic acid to reduce the nitroxide radical and generate the diamagnetic sample, our preliminary studies with spin-labeled EcoRV-DNA complex led to sample precipitation when ascorbic acid was added. The detailed protocol used for spin-labeling is given in Materials and Methods.

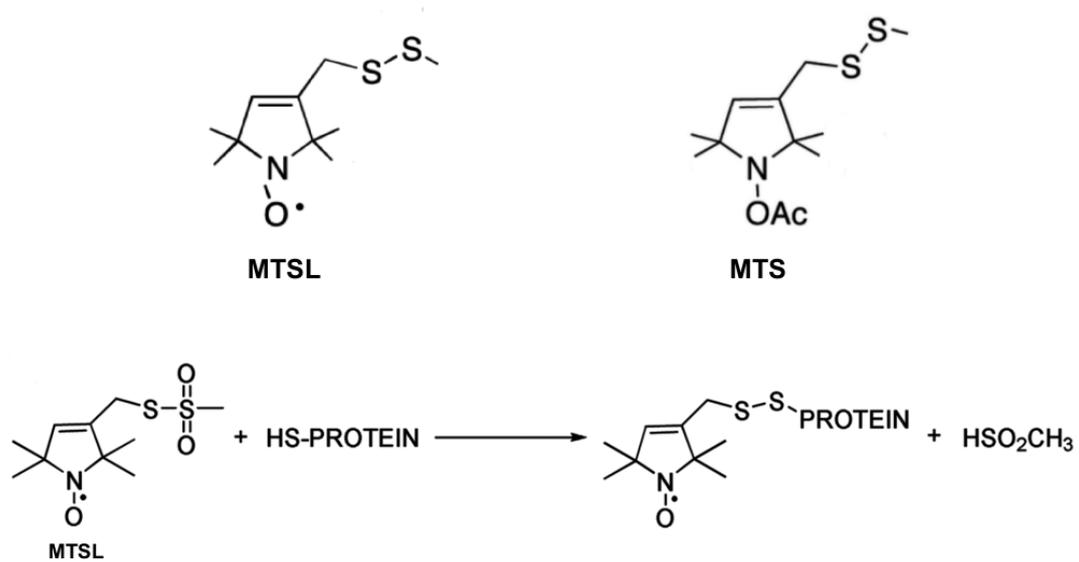
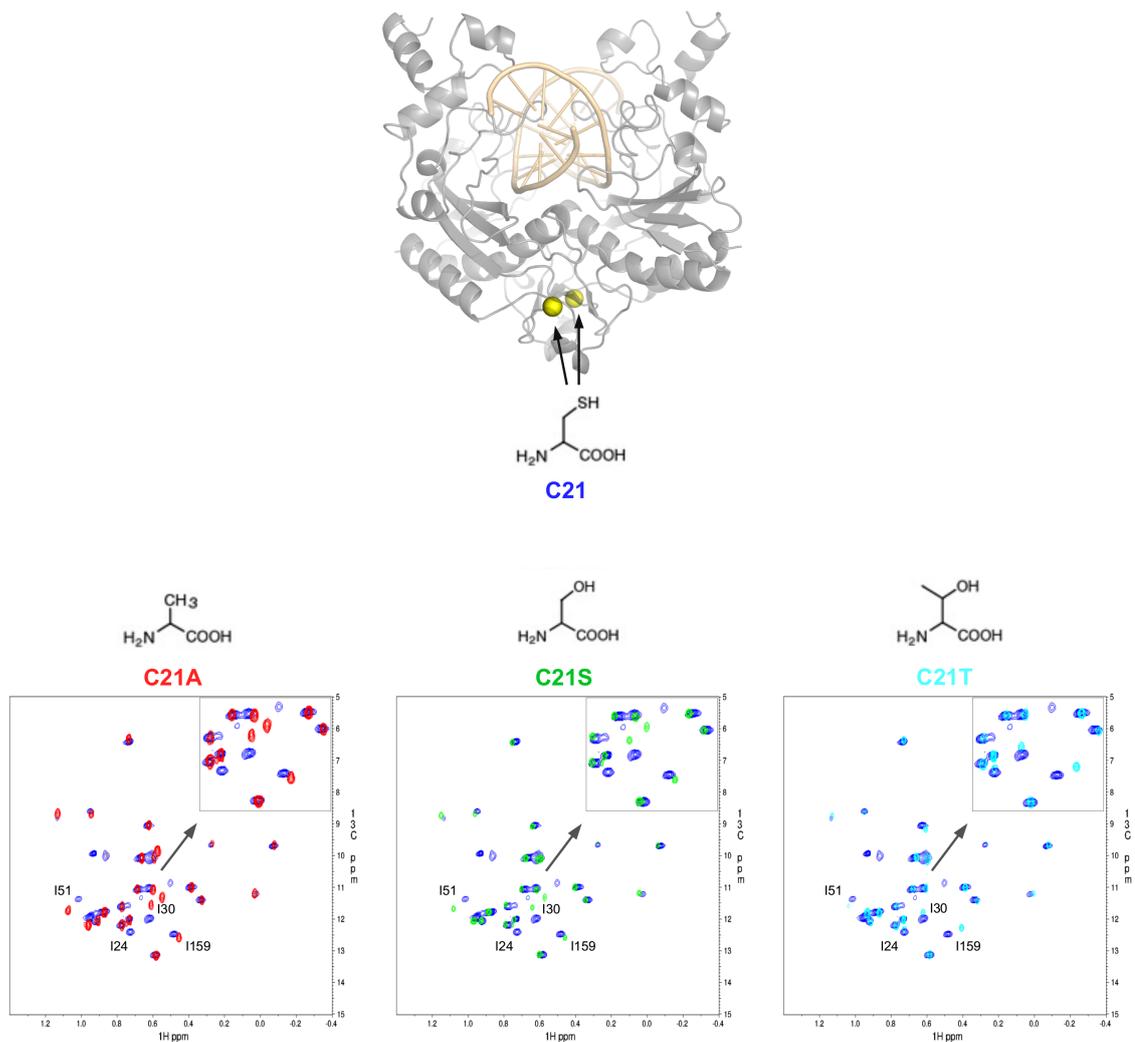


Figure 4.2: Structure and conjugation of MTSL and MTS spin-labels.

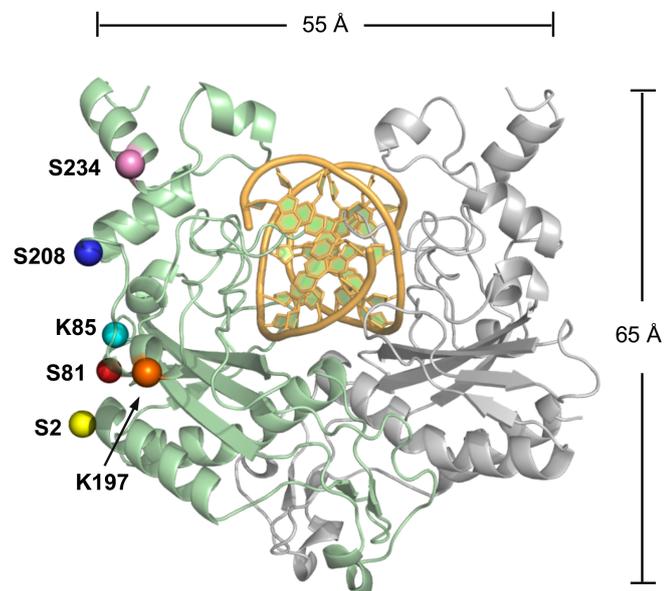
### 4.3 Cysteine-free EcoRV and nitroxide labeling sites

Before selecting sites for spin-labeling and mutating them to cysteine residues, it is important to remove any endogenous cysteines from the protein to prevent undesired labeling at these sites. EcoRV has a single cysteine (C21) which had to be mutated to a conserved residue. Three EcoRV mutants C21A, C21S and C21T were independently generated using site-directed mutagenesis. NMR spectra of each mutant protein-DNA complex was compared with the WT EcoRV-DNA complex to analyze any structural changes due to mutation (Figure 4.3 on page 66). C21T shows local (I30 and I159) and long-range (I51) effects but I30 and I51 movement is smallest in C21T relative to C21A and C21S. Large movement of I159 makes sense because it is next to C21 in the protein structure. Moreover, there is a small void between C21 and adjacent K161 that is perhaps best filled by the larger Thr than Ser or Ala. Thus, of all the three mutants, C21T EcoRV showed the least perturbation compared to WT and was chosen as the background mutant for PRE studies. C21T EcoRV binds the DNA with similar affinity as the WT enzyme (Jen-Jacobson lab) and is catalytically active (data not shown). The next step was to identify the sites to be mutated to cysteines for spin-labeling. The criteria for site-selection was that the site has to be solvent-exposed and should report within the same monomer. Solvent exposure is important to ensure that the labeling is straightforward and complete as even minor diamagnetic impurities (<5%) have been reported to significantly affect the PRE measurements (Iwahara et al., 2007). In addition, a solvent-exposed nitroxide group is less likely to perturb the protein structure. Moreover, since EcoRV-DNA complex is a homodimer, two sites will be labeled per complex. Therefore, to keep the data analysis simple, it was essential that each site reports within the same monomer. Conforming to these criteria, six sites (S2, S81, K85, K197, S208 and S234) were chosen to be mutated to cysteine individually in the C21T EcoRV-DNA complex for spin-labeling. However, two of the six sites (S81 and K85) resulted in insoluble protein, suggesting that these residues participate in interactions responsible for

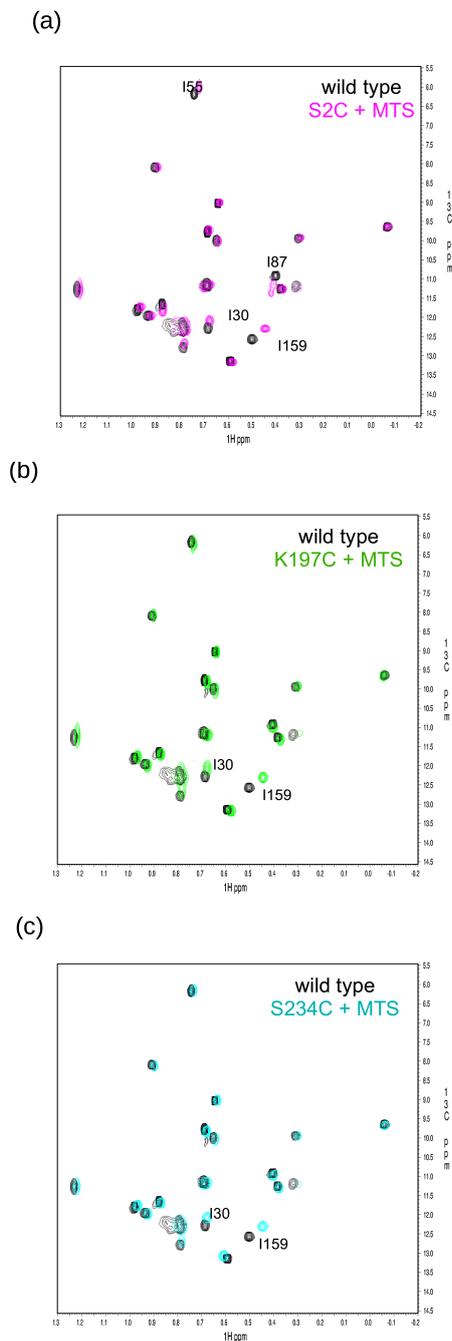
maintaining the EcoRV quarternary structure. Activity assays were performed on the spin labeled proteins to ascertain that labeling does not affect the DNA binding affinity or catalytic activity of EcoRV (data not shown). Also, NMR spectra of the diamagnetic samples (EcoRV-DNA complex with MTS) are identical to that of the WT EcoRV-DNA complex, further suggesting that spin-label does not alter the overall structure or DNA binding of EcoRV (Figure 4.5 on page 68).



**Figure 4.3: Cysteine-free EcoRV to prevent undesired spin-labeling.** Overlays of  $^1\text{H}$ - $^{13}\text{C}$  HMQC spectra of the Ile- $\delta$ -CH<sub>3</sub> region of WT EcoRV-DNA complex (blue) with the respective mutants (C21A - red, C21S - green and C21T - cyan). Peaks that move due to mutation are labeled.  $^{13}\text{C}$  chemical shifts were referenced to TSS.



**Figure 4.4: Sites mutated to cysteine for spin-labeling.** Only one monomer is labeled for clarity. The dimensions of EcoRV-DNA complex are indicated. PRE range of nitroxide spin-label is up to 25 Å.



**Figure 4.5: Spin-label does not affect the protein structure.** Overlays of  $^1\text{H}$ - $^{13}\text{C}$  methyl spectra of WT and diamagnetic (MTS-labeled) EcoRV-DNA complex with  $\text{Lu}^{3+}$ . WT spectra is shown in black. (a) S2C + MTS - magenta, (b) K197C + MTS - green and (c) S234C + MTS - cyan. In S2C + MTS, I55 and I87 show a shift in position, which is a local effect since these residues are very close to the S2C site. Apart from this, only I159 and I30 move in all spectra which is due to the background mutation C21T as seen earlier. All other diamagnetic spectra peaks are coincident with the WT peaks.  $^{13}\text{C}$  chemical shifts were referenced to TSS.

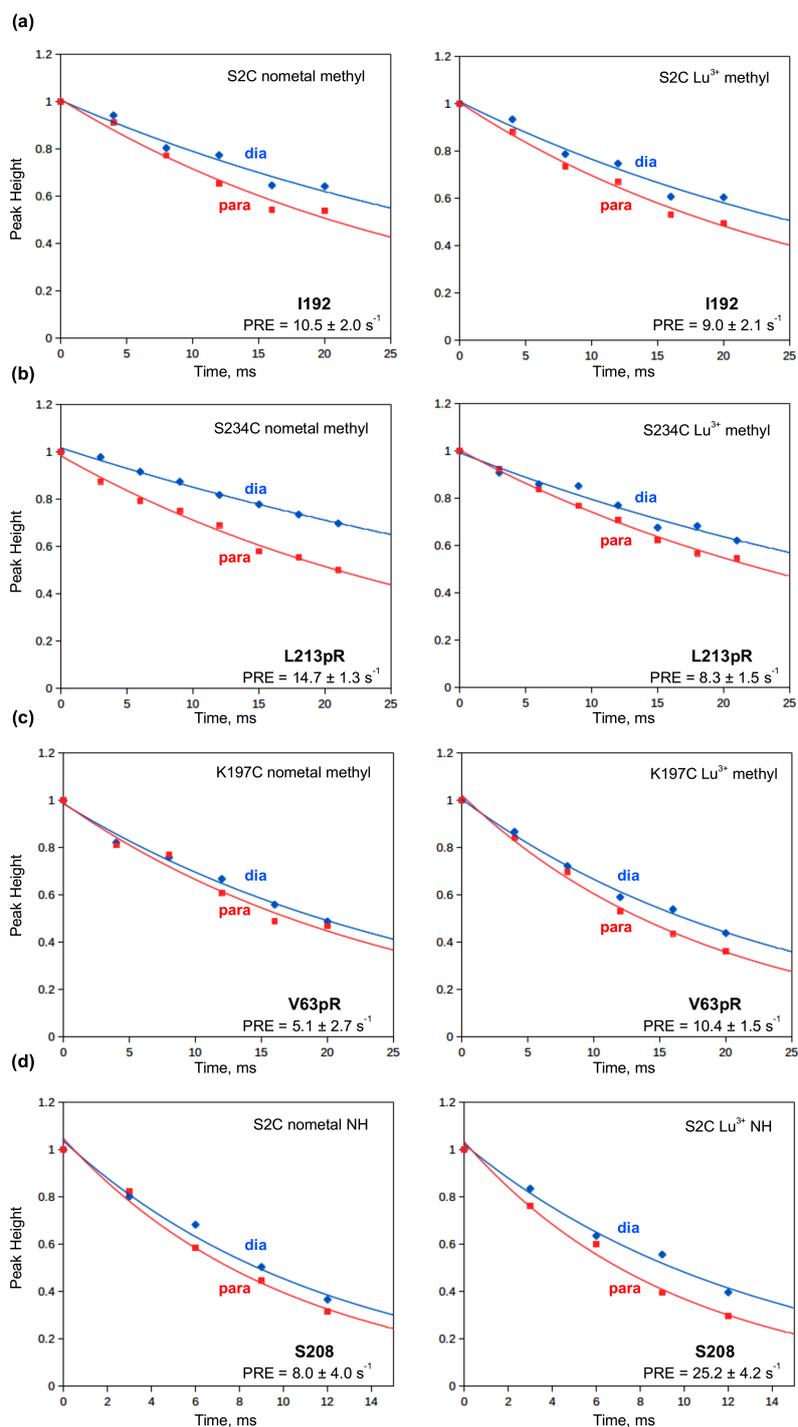
## 4.4 PRE measurements on EcoRV-DNA complex

PREs are measured on protons ( $^1\text{H}$ ) because of their high sensitivity due to their large nuclear gyromagnetic ratio.  $^1\text{H}$  PREs are of two types, longitudinal and transverse PREs represented by  $\Gamma_1$  and  $\Gamma_2$  respectively. However, longitudinal  $^1\text{H}$ -PRE ( $^1\text{H}\text{-}\Gamma_1$ ) is not very useful in providing distance restraints or detecting collective domain motions because it is sensitive to internal motions and unfavorable cross relaxation mechanisms. On the other hand, transverse  $^1\text{H}$ -PRE ( $^1\text{H}\text{-}\Gamma_2$ ) is much less affected by these factors and hence is often used (Iwahara et al., 2004b).  $^1\text{H}\text{-}\Gamma_2$  PRE rates were initially measured using a two-time-point method (Iwahara et al., 2007). However, there was variance in the rates calculated by this method when the experiments were repeated. Therefore, we shifted to a more reliable multi-time-point strategy in which the  $^1\text{H}$ - transverse relaxation rates for the diamagnetic and paramagnetic samples were obtained by fitting the peak intensities at multiple time points to a two-parameter single exponential function. The PRE rate was calculated as the difference in the transverse relaxation rates of the paramagnetic and the diamagnetic samples:

$$\Gamma_2 = R_{2,para} - R_{2,dia}$$

$^1\text{H}\text{-}\Gamma_2$  PRE rates were measured on  $^{15}\text{N}$ -attached and  $^{13}\text{C}$ -attached methyl protons of EcoRV-DNA complex with and without  $\text{Lu}^{3+}$  with nitroxide labeled at S2, K197, S208 and S234 (Appendix). All sites but S208C, gave good quality and well-dispersed spectra and therefore S208C was excluded from analysis. Data were collected at  $35^\circ\text{C}$  on a Bruker-BioSpin Avance 700 MHz NMR spectrometer equipped with cryogenic probe using the interleaved versions of the pulse programs reported by Clore and co-workers (Iwahara et al., 2004b, 2007, Appendix). A few examples of paramagnetic and diamagnetic  $^1\text{H}$ -transverse relaxation decay curves used to calculate PREs are shown in Figure 4.6 on page 71. The errors

in PRE rates are the errors in fitting to the single exponential function. For comparisons with the crystal structures, a global error was defined for all three sites. Maximum error was determined for each spin-label data set and the global error was calculated as an average of maximum errors of all three data sets. For NH and methyl side-chain  $^1\text{H}$ - $\Gamma_2$  PRE rates, the global errors were determined to be  $\pm 14 \text{ s}^{-1}$  and  $\pm 3.7 \text{ s}^{-1}$  respectively.



**Figure 4.6:**  $^1\text{H}$  transverse relaxation decay curves. Panels on the left are for nometal complex and the panels on right are for the complex with  $\text{Lu}^{3+}$ . The site labeled with nitroxide and the type of PRE rate (NH or methyl) is indicated at the top of each panel. Paramagnetic curve is colored red and the diamagnetic curve is colored blue. An example each for Ile, Leu & Val methyls and mainchain amide is shown. (a) I192, (b) L213pR, (c) V63pR and (d) S208. PRE was calculated as the difference in paramagnetic and diamagnetic rates ( $\Gamma_2 = R_{2,para} - R_{2,dia}$ ).

#### 4.4.1 Back-calculating PREs from crystal structures

Amide and methyl side-chain  $^1\text{H-}\Gamma_2$  PRE rates were back-calculated for each nitroxide spin-label position from the crystal structures of EcoRV-DNA complex with and without  $\text{Lu}^{3+}$  (PDB ID 5F8A and PDB ID 1B95 respectively). An ensemble approach, where the paramagnetic tag is represented by a collection of multiple conformers (Iwahara et al., 2004b) was employed and the PRE values were calculated as the sum of the PREs for the individual conformers (Anthis and Clore, 2015). We used a 3 conformer model to represent the flexible nitroxide group, which was tested to be sufficient to account for the conformational space sampled by the nitroxide. Simulated annealing protocol and PRE-pseudo-potential of XPLOR-NIH were used to refine the nitroxide group against the measured PRE data (Iwahara et al., 2007). The refinement was done independently for the nometal and with  $\text{Lu}^{3+}$  PRE data sets to account for any changes in nitroxide positions due to metal-ion binding. The scripts used for refinement were obtained from Dr. Marius Clore (<http://spin.niddk.nih.gov/clore/Software/software.html>) and the modified versions are given in the Appendix. The rotational correlation time ( $\tau_c$ ) for EcoRV-DNA complex was iterated from 25 to 35 ns to optimize the agreement between the experimental and back-calculated PREs. In addition, a protein correlation time calculator (<http://nickanthis.com/tools/tau>) was used to verify the  $\tau_c$  for EcoRV-DNA complex ( $\tau_c=32$  ns for a 68 kDa complex at 35°C). To evaluate the agreement between experimental PREs and those back-calculated from the crystal structures, a PRE Q-factor was used and is defined as (Iwahara et al., 2004b):

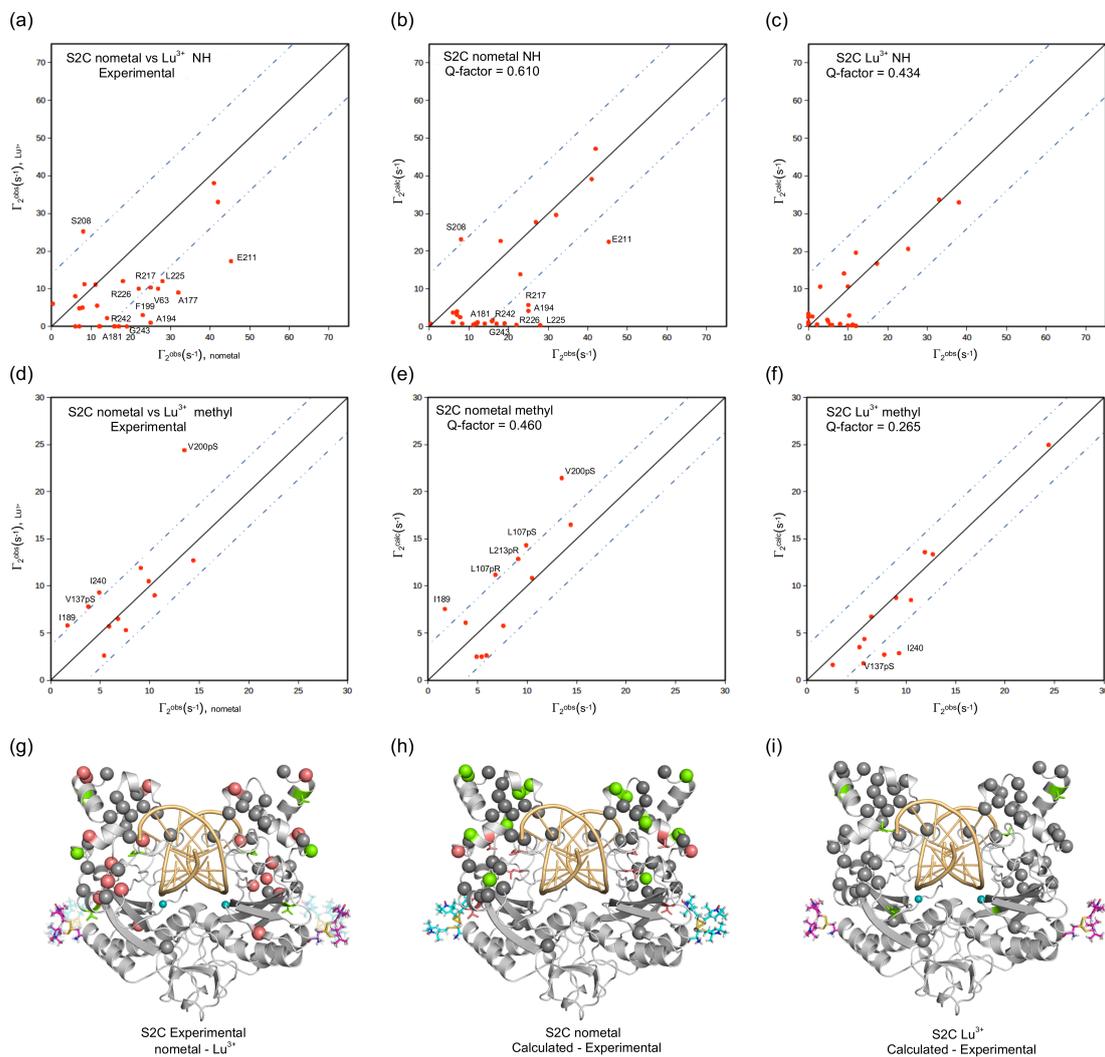
$$Q = \sqrt{\frac{\sum_i \{\Gamma_2^{obs}(i) - \Gamma_2^{calc}(i)\}^2}{\sum_i \Gamma_2^{obs}(i)^2}}$$

where  $\Gamma_2^{obs}(i)$  and  $\Gamma_2^{calc}(i)$  are the observed and calculated transverse PRE rates respectively for residue  $i$ . Lower Q-factors entail better agreement between experimental and back-calculated PRE data.

### 4.4.2 Spin-label at S2C

For EcoRV-DNA complex with spin-label at S2C, the amide and methyl side-chain  $^1\text{H-}\Gamma_2$  PRE correlation plots with and without  $\text{Lu}^{3+}$  are shown in Figure 4.7 on page 75. The correlation plots of experimental PREs with and without  $\text{Lu}^{3+}$  (panels a & d) show metal-induced changes in the PRE rates of some regions of the EcoRV-DNA complex. When compared to the crystal structures, the experimental and back-calculated PRE rates of the no-metal complex have poor agreement as indicated by the PRE Q-factor values (S2C:  $\text{Q-factor}_{\text{amide}} = 0.610$ ;  $\text{Q-factor}_{\text{methyl}} = 0.460$ ), implying that there are differences – either due to dynamics or true structural differences between the no-metal complex in solution and the crystal structure (panels b & e). On the other hand, Q-factor values for the complex with  $\text{Lu}^{3+}$  are lower (S2C:  $\text{Q-factor}_{\text{amide}} = 0.434$ ;  $\text{Q-factor}_{\text{methyl}} = 0.265$ ), indicating that the crystal structures better represent the complex with metal-ion in solution for the regions probed by S2C (panels c & f). Majority of the residues for which the back-calculated and experimental PREs of the no-metal complex show poor agreement (E211, R217, L225, R226, R242, G243, S208, I240 and V200), belong to the C-terminal subdomains (200-245) of EcoRV, suggesting that this region has a different mean position or it is more flexible in the absence of metal ion. Some regions of the C-terminal subdomains (Y219, R221, S223, Q224 and R226) make intricate contacts with the DNA phosphates and crystallographic studies suggest that the C-terminal region undergoes a disorder-to-order transition upon DNA binding (Perona and Martin, 1997). MD simulation studies on free and DNA bound EcoRV suggest that the two C-terminal domains of EcoRV are very dynamic in unbound state and come close together when bound to cognate DNA (Zahran et al., 2010). Moreover, H/D exchange studies, previously done in our lab (unpublished, Dr. Kaustubh Sinha), show that the amide exchange rates of Q224, R237 and G243 are reduced upon  $\text{Lu}^{3+}$  binding, further supporting the decrease in flexibility of the C-terminal subdomains due to metal-ion binding. Thus, our PRE data from S2C site can be interpreted to mean that the C-terminal

region is dynamic in solution and enjoys some degree of conformational flexibility even after binding to DNA, but becomes more compact upon metal-ion binding. To further verify this, and distinguish between a structural change or change in dynamics of the C-terminal region, residual dipolar coupling (RDC) experiments will be employed (discussed in the next chapter).

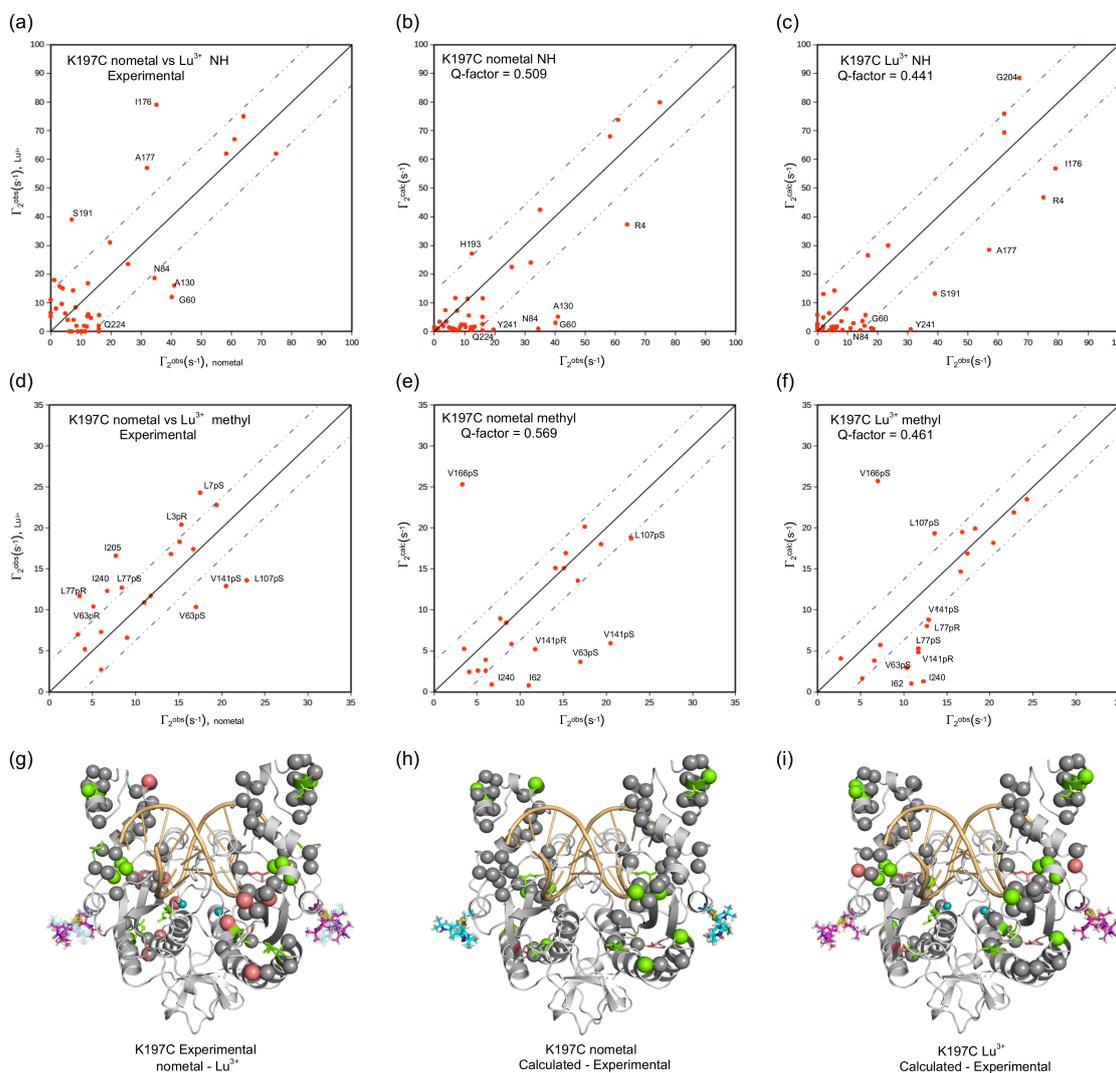


**Figure 4.7: PRE correlation plots for spin-label at S2C. (a) & (d)** Effect of  $\text{Lu}^{3+}$  binding on the observed  $^1\text{H}-\Gamma_2$  PRE rates for amide and methyl side-chain protons respectively. **(b) & (e)** Correlation between observed and back-calculated  $^1\text{H}-\Gamma_2$  PRE rates for amide and methyl-sidechain protons respectively of EcoRV-DNA complex (PDB ID 1B95). **(c) & (f)** Correlation between observed and back-calculated  $^1\text{H}-\Gamma_2$  PRE rates for amide and methyl-sidechain protons respectively of EcoRV-DNA- $\text{Lu}^{3+}$  complex (PDB ID 5F8A). Blue dotted line indicates the global error in PRE rate measurement. For NH and methyl side-chain, the global errors in  $^1\text{H}-\Gamma_2$  PRE rates were determined to be  $\pm 14 \text{ s}^{-1}$  and  $\pm 3.7 \text{ s}^{-1}$  respectively. Only residues outside the error line are labeled in all plots for clarity. **(g), (h) & (i)** PRE differences mapped on the structures of EcoRV-DNA complex. Red spheres/sticks indicate decrease in PRE rate, green spheres/sticks indicate increase in PRE rate and dark gray spheres indicate no change in PRE rate within the global error. Nitroxides are labeled in cyan (nometal) and magenta ( $\text{Lu}^{3+}$ ). **(g)** Differences in experimental PRE rates due to  $\text{Lu}^{3+}$  binding [nometal -  $\text{Lu}^{3+}$ ]. **(h) & (i)** Differences between back-calculated and experimental PREs for nometal and with  $\text{Lu}^{3+}$  complexes respectively [calculated - experimental].

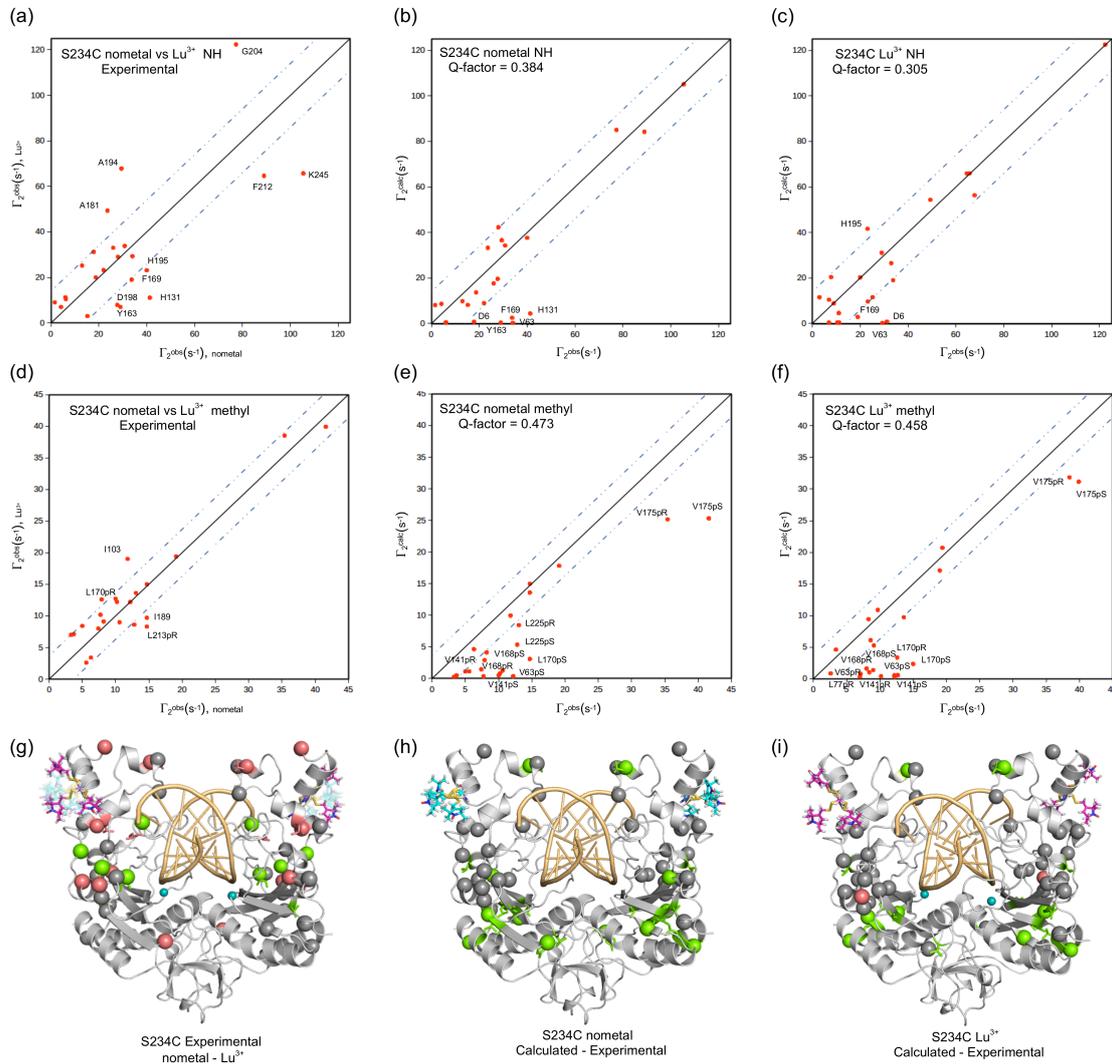
### 4.4.3 Spin-label at K197C and S234C

The amide and methyl side-chain  $^1\text{H-}\Gamma_2$  PRE correlation plots with and without  $\text{Lu}^{3+}$  for spin-label at K197C and S234C sites are shown in Figure 4.8 on page 77 and Figure 4.9 on page 78. Metal-induced changes are observed in both amide and methyl side-chain  $^1\text{H-}\Gamma_2$  PRE rates (panels a & d). However, the PRE analysis of these sites is confounding and inconclusive at present. K197C may be sampling different positions in  $\pm \text{Lu}^{3+}$  complexes as it is part of a small helix (196-201) and is flanked by large loop regions (177-195 and 202-209) on both sides. Furthermore, PRE rates of A194 & H195 (neighboring K197), as calculated from S2C site, suggests that K197 is more dynamic in the nometal complex compared to the complex with  $\text{Lu}^{3+}$ , implying that the space sampled by the spin-label at this site is different in the complexes with and without  $\text{Lu}^{3+}$ .

In the case of S234C, it is unclear if the agreement between back-calculated and experimental PREs is significantly different  $\pm \text{Lu}^{3+}$  (Figure 4.9 on page 78). S234C is part of the C-terminal region that shows changes in PREs, as probed from S2C site, and therefore also may be occupying a different positions in solution compared to crystal structures and  $\pm \text{Lu}^{3+}$ . Consequently, the changes in PRE rates due to metal-ion binding that we see or do not see from K197C and S234C sites may be a result of the changes in the spin-label positions. In order to verify the changes in PRE rates reported by these sites and compare with PREs back-calculated from crystal structures, it is important to assess their flexibility and accurately determine their positions in the nometal and  $\text{Lu}^{3+}$  complexes. RDC experiments (discussed in the next chapter) will be crucial in this regard as the bond vector restraints from RDCs will help us better model the conformational flexibility of the spin-label at K197C and S234C sites.



**Figure 4.8: PRE correlation plots for spin-label at K197C.** (a) & (d) Effect of  $\text{Lu}^{3+}$  binding on the observed  $^1\text{H}-\Gamma_2$  PRE rates for amide and methyl side-chain protons respectively. (b) & (e) Correlation between observed and back-calculated  $^1\text{H}-\Gamma_2$  PRE rates for amide and methyl-sidechain protons respectively of EcoRV-DNA complex (PDB ID 1B95). (c) & (f) Correlation between observed and back-calculated  $^1\text{H}-\Gamma_2$  PRE rates for amide and methyl-sidechain protons respectively of EcoRV-DNA- $\text{Lu}^{3+}$  complex (PDB ID 5F8A). Blue dotted line indicates the global error in PRE rate measurement. For NH and methyl side-chain, the global errors in  $^1\text{H}-\Gamma_2$  PRE rates were determined to be  $\pm 14 \text{ s}^{-1}$  and  $\pm 3.7 \text{ s}^{-1}$  respectively. Only residues outside the error line are labeled in all plots for clarity. (g), (h) & (i) PRE differences mapped on the structures of EcoRV-DNA complex. Red spheres/sticks indicate decrease in PRE rate, green spheres/sticks indicate increase in PRE rate and dark gray spheres indicate no change in PRE rate within the global error. Nitroxides are labeled in cyan (nometal) and magenta ( $\text{Lu}^{3+}$ ). (g) Differences in experimental PRE rates due to  $\text{Lu}^{3+}$  binding [nometal -  $\text{Lu}^{3+}$ ]. (h) & (i) Differences between back-calculated and experimental PREs for nometal and with  $\text{Lu}^{3+}$  complexes respectively [calculated - experimental].



**Figure 4.9: PRE correlation plots for spin-label at S234C.** (a) & (d) Effect of  $\text{Lu}^{3+}$  binding on the observed  $^1\text{H}-\Gamma_2$  PRE rates for amide and methyl side-chain protons respectively. (b) & (e) Correlation between observed and back-calculated  $^1\text{H}-\Gamma_2$  PRE rates for amide and methyl-sidechain protons respectively of EcoRV-DNA complex (PDB ID 1B95). (c) & (f) Correlation between observed and back-calculated  $^1\text{H}-\Gamma_2$  PRE rates for amide and methyl-sidechain protons respectively of EcoRV-DNA- $\text{Lu}^{3+}$  complex (PDB ID 5F8A). Blue dotted line indicates the global error in PRE rate measurement. For NH and methyl side-chain, the global errors in  $^1\text{H}-\Gamma_2$  PRE rates were determined to be  $\pm 14 \text{ s}^{-1}$  and  $\pm 3.7 \text{ s}^{-1}$  respectively. Only residues outside the error line are labeled in all plots for clarity. (g), (h) & (i) PRE differences mapped on the structures of EcoRV-DNA complex. Red spheres/sticks indicate decrease in PRE rate, green spheres/sticks indicate increase in PRE rate and dark gray spheres indicate no change in PRE rate within the global error. Nitroxides are labeled in cyan (nometal) and magenta ( $\text{Lu}^{3+}$ ). (g) Differences in experimental PRE rates due to  $\text{Lu}^{3+}$  binding [nometal -  $\text{Lu}^{3+}$ ]. (h) & (i) Differences between back-calculated and experimental PREs for nometal and with  $\text{Lu}^{3+}$  complexes respectively [calculated - experimental].

## 4.5 Discussion

Analysis of S2C PRE data suggests there are differences between the structure of the nometal complex in solution and the crystal structures. Specifically, the C-terminal region of EcoRV appears to be more flexible in solution in the nometal complex but becomes rigid upon metal-ion binding. However, the current PRE analysis has a few limitations. First, K197C and S234C data analysis is confounding due to flexibility of these sites and additional experiments such as RDCs will have to be performed to assess their flexibility. Second, we have not tested the data quality of each complex (nometal and Lu<sup>3+</sup>) with regards to the diamagnetic rates, which should be similar (within error) for all three sites for each residue that is included in the PRE analysis. Third, whether the differences between Q-factors of nometal and Lu<sup>3+</sup> complexes are significant has not been evaluated. In order to address these limitations, diamagnetic relaxation rates for each residue currently included in the PRE analysis will be compared for all three sites. Residues showing significant changes in diamagnetic relaxation rates within a data set will be removed and the PRE data will be reanalyzed. Also, to test if the differences in Q-factors are significant or not, errors will be propagated by generating random data sets using experimental errors.

# Chapter 5

## Summary, Conclusions and Future Directions

### 5.1 Summary and Conclusions

In this dissertation, we show that lanthanide ions ( $\text{Ln}^{3+}$ ) are excellent candidates for studying protein-nucleic acid interactions involving metal ions. Specifically, we show that  $\text{Lu}^{3+}$  inhibits  $\text{Mg}^{2+}$  dependent activity, binds with the same stoichiometry and occupies a site similar to  $\text{Mg}^{2+}$ . Also, we demonstrate that anomalous scattering of lanthanides can be used to unambiguously locate metal-binding sites by X-ray crystallography. Additionally, we have assigned 120/170 observable mainchain amide NMR resonances and 45/50 Ile, Leu and Val side-chain methyl resonances. Although not complete, these assignments provide a good overall coverage of the EcoRV-DNA complex, which is essential for dynamics studies. Moreover, the side-chain methyls of Leu and Val were stereospecifically assigned for a high resolution analysis of dynamic and/or structural changes. The assignments have laid the foundation for all future analyses of EcoRV-DNA complex by solution NMR spectroscopy.

In order to survey the metal-induced changes in structure and/or dynamics of EcoRV-DNA complex, we have used PRE with nitroxide spin-labels as the paramagnetic probes.

PRE data from spin-label at site S2C, suggest that there are differences in structure and/or dynamics between the nometal complex in solution and the crystal structure. Specifically, the C-terminal domain of EcoRV moves closer to S2C site, which is located at the other end of the protein. It is not clear at this time whether this represents a change in the structure of the enzyme or is a result of increased motion of the carboxy-terminus. PRE analysis of K197C and S234C sites is confounding at present, presumably because these sites are more dynamic in solution and they sample different locations with and without  $\text{Lu}^{3+}$ . Additional experiments such as residual dipolar coupling (RDC) and PRE with rigid probes such as dHis( $\text{Cu}^{2+}$ ) will have to be performed to assess the flexibility of these sites and accurately model the space sampled by them.

## 5.2 Future Directions

### 5.2.1 Residual dipolar coupling (RDC) experiments

Compared to other NMR experiments, residual dipolar couplings or RDCs provide a more global information about the geometry of a protein or protein complex. RDCs report on orientations of the inter-nuclear vectors relative to the external magnetic field. In RDC experiments, a protein or a protein complex is aligned weakly relative to the external magnetic field. The weak alignment as opposed to the isotropic Brownian motion in solution, provides an anisotropic condition such that the magnetic dipole-dipole interactions no longer average to zero. This results in an observable dipolar coupling, which can be measured as splitting in the NMR spectrum. RDCs can be used to probe both structure and dynamics of a macromolecule. A good agreement between the experimental and back-calculated RDCs (from crystal structures) signifies there are no major structural changes in solution. However, in the case they do not agree, an ensemble calculation of conformational states (Clare and

Schwieters, 2004) and various motional models can be used to account for the structural and/or dynamic differences in solution (Bouvignies et al., 2005). RDC experiments with multiple alignment media can help deconvolute a structural change from dynamics (Chen and Tjandra, 2012). Identical values of inter-nuclear vector orientations from multiple data sets imply that the particular vector is rigid. On the other hand if the values are different, it entails the inter-nuclear vector is dynamic.

For RDC studies of EcoRV-DNA complex, a detergent compatible alignment media will be used because our NMR samples contain 0.1% CHAPS, which is included in the NMR buffer to get a well dispersed spectrum for the amide resonances. The few detergent compatible alignment media available are stretched or compressed polyacrylamide gels, pinacyanol acetate or the more expensive DNA nanotubes and cellulose nanocrystals (Thiagarajan-Rosenkranz et al., 2015). Initially, RDC data will be acquired for backbone N-H bond vectors in EcoRV-DNA complex with and without  $\text{Lu}^{3+}$  using one of the above alignment media and will be compared to the RDCs predicted from the respective crystal structures. If they agree, we can discount any structural changes in solution and the changes observed in the PRE experiments can be attributed to collective domain motions. However, if the experimental and predicted RDCs do not agree, structural changes in solution may be present and a combination of ensemble calculation and various motional models will have to be employed to characterize them.

### **5.2.2 Inter-subunit distances and PRE with d(His) $\text{Cu}^{2+}$ - IDA**

The conformational flexibility of the nitroxide spin-label presents a challenge in reliable measurements of PRE rates or electron spin resonance (ESR) distances. In an effort to address this issue, Saxena and co-workers recently demonstrated the use of double-histidine (dHis)  $\text{Cu}^{2+}$ -binding motif as a rigid spin label in ESR distance measurements (Cunningham

et al., 2015). Two His sites are engineered at strategic positions such that their side-chains chelate a single  $\text{Cu}^{2+}$  ion. In addition, to prevent non-specific binding of  $\text{Cu}^{2+}$  in proteins with metal active-sites,  $\text{Cu}^{2+}$  is first complexed with iminodiacetate or IDA. To complement and verify the PRE data with nitroxide spin-label at S2C and S234C, d(His)- $\text{Cu}^{2+}$  IDA will be used as the paramagnetic probe at the same sites in EcoRV-DNA complex. For the diamagnetic samples, d(His)- $\text{Zn}^{2+}$  IDA will be used. Moreover, the d(His)- $\text{Cu}^{2+}$  IDA probe will be used to measure inter-subunit distances in EcoRV-DNA  $\pm\text{Lu}^{3+}$  complex using ESR spectroscopy.

### **5.2.3 Spin-labeling the DNA, PRE with paramagnetic lanthanides and DNA dynamics**

To expand our PRE coverage and report on EcoRV regions near the DNA binding interface, the DNA will be spin-labeled. DNA with phosphorothioate modification at the site selected for labeling will be reacted with 3-Iodomethyl-(1-oxy-2,2,5,5-tetramethylpyrroline) as described by Scholes and co-workers (Sun et al., 2012). In addition to metal-ion binding, the sequence of DNA triplet flanking the recognition sequence modulates EcoRV-DNA affinity (unpublished, Jen-Jacobson lab). Therefore, to study the effect of varying the flanking context in an EcoRV-DNA-metal complex, paramagnetic lanthanides will be used as PRE probes in addition to d(His)- $\text{Cu}^{2+}$  and spin-labeled DNA. Lanthanide ions are chemically similar, which makes it feasible to use multiple  $\text{Ln}^{3+}$  ions with very different paramagnetic properties in the same system.

Additionally, in order to study the metal-induced changes in dynamics from DNA's perspective, DNA will be isotopically labeled in an unlabeled protein background. To the best of our knowledge, there has been no studies on DNA dynamics in protein-DNA complexes and hence our work will be quite novel. Conditions for large-scale production of isotopically

---

labeled DNA have been worked out. We employed the PCR based ESRA (endonuclease-sensitive repeat amplification) approach pioneered by Gronenborn and co-workers (Louis et al., 1998). In ESRA, the template is self-primed which results in an exponential growth in DNA length carrying tandem repeats of the target sequence. Following PCR, the product is subjected to endonuclease cleavage to obtain the target sequence. Finally, the desired sequence is purified from a mix of dNTPs and uncleaved DNA using ion-exchange chromatography. Isotopically labeled dNTPs will be obtained from Silantes. Experimental details for large-scale production of isotopically labeled DNA are given in the Appendix.

# Chapter 6

## Materials and Methods

### 6.1 Expression of wild type and mutant EcoRV endonucleases

The codon-optimized synthetic gene encoding wild type EcoRV was obtained from DNA 2.0 (Welch et al., 2009). Mutations were generated using a site-directed mutagenesis protocol (QuickChange, Agilent Technologies) and were verified by complete double-strand sequencing of the gene (GENEWIZ, Inc). Wild type and mutant proteins were expressed from the pET22b(+) vector in C3013 T7 Express lysY/Iq E. coli cells (New England Biolabs), pre-transformed with the pmetB plasmid to constitutively express EcoRV methylase (Bougueleret et al., 1985). The cells, in PG media (Studier, 2005) containing 100 mg/L ampicillin and 50 mg/L kanamycin, were grown at 30°C to an A600 of 1.2-1.4, incubated at 42°C for 1 hr and induced with 1 mM isopropyl-β-D-thiogalactoside (IPTG) at -80°C. Isotopically labeled samples were prepared by growing the cells in 100 % D<sub>2</sub>O media with (<sup>15</sup>NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> as the sole source of nitrogen and <sup>13</sup>C deuterated glucose-D<sub>7</sub> (3 g/L) as the carbon source. ILV methyls were <sup>13</sup>C labeled by addition of the appropriate ketoacid precursors (Tugarinov and Kay, 2003) to the medium 1.5 hours prior to induction with IPTG for

18 hours. Isotopically labeled compounds were purchased from either Cambridge Isotopes or Sigma Aldrich.

## 6.2 Protein purification

Cell pellets were resuspended in lysis buffer (20 mM potassium phosphate, pH 8.0, 0.5% Triton X-100 and 10 mM EDTA), stirred on ice with a pinch of lysozyme and lysed by sonication. The lysate was centrifuged at 30,000 rpm for 40 minutes at 4°C. The supernatant was passed over a Q Sepharose (GE Healthcare) column and washed with buffer A (20 mM potassium phosphate, pH 8.0 and 10 mM EDTA). The flow-through containing EcoRV was passed over a SP C-50 column (Sigma Aldrich) and washed with buffer B (20mM potassium phosphate, pH 7.0, 100 mM NaCl and 10 mM EDTA). The protein was eluted using a salt gradient [buffer B, 0.1 → 1.5 M NaCl], concentrated and passed through a gel filtration column (Sephadex G-50; 20 mM MES, pH 6.5, 600 mM NaCl, 0.1% CHAPS and 1 mM EDTA). Typically, the yield was 100 mg protein from 1 L of bacterial culture, a 20 fold higher yield than using the non-optimized wild-type coding sequence (Bougueleret et al., 1985).

## 6.3 Oligodeoxynucleotide substrates

Purified single-stranded oligodeoxynucleotides were purchased from IDT (Integrated DNA Technologies) and resuspended in annealing buffer (10 mM HEPES, pH 7.5; 200 mM NaCl, 0.1 mM DTT) to a final concentration of 3 mM. DNA duplexes were formed by mixing equimolar amounts of single stranded species and annealing by incubating the oligodeoxynucleotides at 95°C for 5 min and slow cooling in a styrofoam box overnight. The annealed

duplex DNA was characterized using PAGE and UV spectroscopy and stored at  $-20^{\circ}\text{C}$ . For NMR samples, a 16 mer DNA substrate (5'-GCAAAGATATCTTTTCG-3') was used and the corresponding oligodeoxynucleotides purchased from IDT were purified by standard de-salting. In case of x-ray crystallography samples, an HPLC grade 12 mer DNA substrate (5'-AAAGATATCTTT-3') was used.

## 6.4 NMR sample preparation

Purified wild type and mutant EcoRV proteins were exchanged, unless indicated otherwise, into NMR buffer (20 mM HEPES, pH 7.4, 200 mM NaCl, 100  $\mu\text{M}$  EDTA and 0.1 % CHAPS). Metal-free EcoRV-DNA samples were prepared by stoichiometric (1 DNA: EcoRV dimer) addition of cognate DNA (5'-GCAAAGATATCTTTTCG-3') to a concentrated protein sample, followed by dialysis of the protein-DNA complex against NMR buffer. For EcoRV-DNA-Lu<sup>3+</sup> samples, the EDTA was removed from the concentrated protein by extensive dialysis, followed by the addition of 0.98 DNA equivalent per EcoRV dimer (slightly less than 1 DNA equivalent is added to minimize the amount of free DNA that can interact with Ln<sup>3+</sup>). Sufficient Lu<sup>3+</sup> was then added to saturate the metal binding sites at  $23^{\circ}\text{C}$ . Excess free Lu<sup>3+</sup> was carefully limited to prevent slow precipitation of Lu<sup>3+</sup>-DNA complexes. The concentration (EcoRV monomer) of samples were between 0.2 and 0.4 mM for all methyl assignment and PRE experiments.

### 6.4.1 Paramagnetic relaxation enhancement (PRE) samples

The single Cys residue (C21) in EcoRV was mutated to Thr and six single Cys mutations (S2C, S81C, K85C, K197C, S208C and S234C) were generated in the C21T background

using the Quickchange site-directed mutagenesis protocol. The cells were grown and expressed as described above. The protein was purified as described above, with 3mM DTT added to all the buffers. Prior to spin-labeling, DTT was removed by gel-filtration (G-25 Sephadex; 20 mM HEPES, pH 7.6, 200 mM NaCl and 1 mM EDTA ). MTSL (1-oxy-2,2,5,5-tetramethyl-d-3-pyrroline-3-methyl-methanethiosulfonate, Toronto Research Chemicals) or MTS (1-acetyl-2,2,5,5-tetramethyl-d-3-pyrroline-3-methyl-methanethiosulfonate, Santa Cruz Biotechnology) were used to generate the corresponding paramagnetic and diamagnetic samples of each mutant protein (Iwahara et al., 2004b; Volkov et al., 2006). Labeling was performed at a protein concentration of 0.2 mM (dimer) by first incubating the enzyme with 2 mM MTSL (or MTS) for 30 min at 4°C and then increasing the concentration of the labeling agent to 4 mM, followed by incubation for an additional 16 hours at 23°C. After completion of the reaction, the excess label was removed by gel filtration (Sephadex G-25; 20 mM HEPES, pH 7.4, 200 mM NaCl, 100  $\mu$ M EDTA and 0.1 % CHAPS). The labeled protein samples retained their ability to cut plasmid DNA at the cognate site (GATATC).

Ellman's reagent (Riddles et al., 1983) was used to test the labeling efficiency. First, a linear standard curve was generated using various amounts of cysteine hydrochloride monohydrate. Then, as a control the protein concentration of unlabeled protein sample was estimated from the free sulfhydryl concentration (in the range 0.1-1.0 mM) by fitting to the standard curve. This value was compared to the value obtained from UV absorbance (Abs 280 nm) and the two agreed. Next, the samples labeled with nitroxide were quantified for free sulfhydryl groups. There was practically no change observed in the absorbance (412 nm) of these samples suggesting no free sulfhydryls are available for reaction and the labeling is > 99% efficient. In addition, if the labeling was less than 99% efficient, the paramagnetic spectra should show many peaks relaxing slower than that expected from the crystal structures, which was not observed.

## 6.5 NMR spectroscopy

### 6.5.1 Methyl assignments

Ile, Leu and Val methyl side-chain assignments were obtained by a combination of correlation spectroscopy, site-directed mutagenesis, 4D methyl-methyl NOE and paramagnetic relaxation enhancement (PRE) experiments. Correlations between the methyl resonances and the  $C\alpha$ ,  $C\beta$  and  $C\gamma$  resonances were obtained using high sensitivity versions of the HMCM[CG]CBCA experiment (Tugarinov and Kay, 2003), as described previously (Sinha et al., 2013). Due to a number of missing mainchain amide assignments, it was necessary to supplement these assignments by site-directed mutagenesis, 4D methyl-methyl NOE and PRE measurements. Single-site mutations were generated for the following residues: 8, 23, 30, 43, 51, 52, 55, 62, 87, 89, 103, 129, 133, 153, 176, 189 (Ile→Val), 63, 122, 141, 137, 168, 175, 200 (Val→Ala), 148, 156, 170, 180, 213, 225 (Leu→Val). NMR data were acquired at 35°C on Bruker spectrometer operating at 600 MHz ( $^1\text{H}$ ) equipped with triple resonance cryoprobes and the methyl-HMQC pulse sequence used is given in the Appendix. The  $^1\text{H}$ - $^{13}\text{C}$  methyl spectra of the mutant proteins which showed a single missing resonance (Ile mutants) or a pair of missing resonances (Val or Leu mutants) with minor movement of the remaining ILV peaks were used for assignment purposes.

The wealth of structural information from crystallographic data (>30 EcoRV structures deposited in the PDB) enabled the use of 4D NOE and PRE (Venditti et al., 2011) to generate data to assign additional ILV methyl peaks and confirm the peaks already assigned. The 4D HMQC-NOE-HMQC methyl NOE pulse program was obtained from Dr. Marius Clore. Two differently labeled ILV  $^{13}\text{C}$ -methyl labeled samples were produced for the 4D NOE experiments. One sample was produced using 4- [ $^{13}\text{C}$ ]- $\alpha$ -ketobutyric acid (2-KB) and dimethyl- [ $^{13}\text{C}_2$ ]- $\alpha$ -ketoisovaleric acid as the precursors for ILV labeling (Tugarinov and Kay, 2003). The use of dimethyl- [ $^{13}\text{C}_2$ ]- $\alpha$ -ketoisovaleric acid results in all LV methyl

groups being  $^{13}\text{CH}_3$ -labeled, thus giving a strong intra-residue NOE peak in the 4D NOE spectrum. This enabled us to identify the methyl pairs for LV residues. The second sample was produced by adding 2-KB and 3-methyl- $^{13}\text{C}$ -3,4,4,4  $^2\text{H}_4$ --ketoisovaleric acid (Cambridge Isotope Laboratories, Inc.) as precursors. This resulted in LV side chains with  $^{-13}\text{CH}_3$  and  $^{-12}\text{CD}_3$  methyl groups. The NOE spectrum of this sample lacked the intra-residue peak and permitted detection of NOEs over longer distances, since the signal decays more slowly with fewer vicinal protons. To label specifically the proS methyl groups of Leu( $\delta^2$ - $\text{CH}_3$ ) and Val ( $\gamma^2$ - $\text{CH}_3$ ) with  $^{13}\text{C}$ , we expressed the protein in the presence of 2- $^{13}\text{CH}_3$ -methyl-4- $^2\text{H}_3$ -acetolactate (purchased from NMR-Bio) as described (Gans et al., 2010). The resulting methyl spectra were used to identify the proS methyl peaks. Peaks that were not labeled using this protocol were identified as proR.

### 6.5.2 PRE experiments

Spectra of the paramagnetic and diamagnetic protein-DNA complexes (with and without  $\text{Lu}^{3+}$ ) were acquired at  $35^\circ\text{C}$  on a 700 MHz Bruker spectrometer, with at least 64 accumulated scans per free induction decay (FID). Transverse relaxation rates ( $^1\text{H}-\Gamma_2$ ) for the  $^{15}\text{N}$  amide and  $^{13}\text{C}$ -attached methyl protons were measured using interleaved versions of the pulse sequence reported by Iwahara et al. (Iwahara et al., 2004b, 2007). Time delays used for methyl PRE experiments were - S2C and K197C,  $T = 0, 4, 8, 12, 16$  and  $20$  ms; S234C,  $T = 0, 3, 6, 9, 12, 15, 18,$  and  $21$  ms. Time delays used for all PRE amide experiments were  $T = 0, 3, 6, 9$  and  $12$  ms. NMRPipe (Delaglio et al., 1995) was used to process the spectra and NMRView (Johnson and Blevins, 1994) was used for rate analysis.  $^1\text{H}-\Gamma_2$  was obtained by fitting the peak intensities to a two-parameter single exponential function. The experimental PREs were calculated as the difference in relaxation rates of the paramagnetic and the corresponding diamagnetic samples. PREs from the crystal structure were back-

calculated with XPLOR-NIH (Iwahara et al., 2004b), using a 3 conformer representation of the attached nitroxide.

## 6.6 Crystal structures of EcoRV-DNA-Lu<sup>3+</sup> complexes

### 6.6.1 Crystallization and data collection

The EcoRV-DNA-Lu<sup>3+</sup> complex was generated by combining EcoRV (0.079 mM dimer) with a 12mer DNA duplex (5'-AAAGATATCTTT-3') (0.079 mM) in binding buffer (10 mM HEPES, 200 mM NaCl, 0.1 mM DTT, 1 mM EDTA, 10% glycerol; pH 7.5), incubating for one hour (22°C), followed by dialysis to remove EDTA, and addition of Lu<sup>3+</sup> (0.45 mM). Crystals were obtained by mixing 3  $\mu$ L of EcoRV-DNA-Lu<sup>3+</sup> complex solution with 3  $\mu$ L of reservoir solution (100 mM HEPES, 8% ethylene glycol, 4% PEG 8000, 10 % glycerol; pH 7.5) in a sitting drop tray with 500  $\mu$ L reservoir solution. To generate the EcoRV-cleaved DNA-Lu<sup>3+</sup> complex, EcoRV (0.165 mM dimer) was combined with a 12 mer DNA duplex (5'-AAAGATATCTTT-3') (0.165 mM) in binding buffer without EDTA but with trace amounts of catalytic cofactors Mg<sup>2+</sup> and Mn<sup>2+</sup> for 1 hour (22°C), followed by addition of Lu<sup>3+</sup> (0.79 mM). We verified complete DNA cleavage in solution, prior to the addition of the inhibitor Lu<sup>3+</sup> by replicating the incubation. Crystals of the cleaved complex were obtained with a similar procedure except the PEG 8000 concentration was 8 %. In both cases, small crystals appeared within 2 weeks and grew to ~0.2 mm after 5 weeks. Diffraction data were acquired on frozen crystals (100K) using a Rigaku FR-E generator with a Cu rotating anode. Data for the uncleaved complex were collected on a Rigaku RAXIS HTC detector, whereas a Saturn 944 CCD detector was used for cleaved complex.

### 6.6.2 Structure determination, model building and refinement

Data reduction and model building utilized software provided in the CCP4 software package (Winn et al., 2011). Indexing and integration of the reflections were performed with iMosflm (Battye et al., 2011). Scaling and truncation of data were performed using Scala (Evans, 2006) and cTruncate (Zwart, 2005). Initial phases were determined by molecular replacement using Phaser (McCoy et al., 2007); an EcoRV-DNA complex with 2 Ca<sup>2+</sup> ions (PDB ID 1B94) was used as the search model. Model building was performed using Coot (Emsley and Cowtan, 2004) and structures were refined with REFMAC5 (Murshudov et al., 2011). Maps were generated using FFT (Read and Schierbeek, 1988). The Lu<sup>3+</sup> ions were readily apparent in the initial map, showing strong density ( $>5\sigma$ ) at different locations from those of the Ca<sup>2+</sup> ions. To verify that the DNA was cleaved in the X-ray derived model of the EcoRV-cleaved DNA-Lu<sup>3+</sup>, we calculated omit maps, using the initial molecular replacement solution, but omitting the central TA bases in each strand. The resultant electron density was consistent with cleavage of the phosphodiester bond between these bases. Anomalous difference maps were calculated to confirm Lu<sup>3+</sup> binding. Structures were validated with Polygon (Urzhumtseva et al., 2009) and Rampage (Lovell et al., 2003).

# References

- Anthis, N. J. and Clore, G. M. (2015). Visualizing transient dark states by NMR spectroscopy. *Q Rev Biophys.*, 1:35–116.
- Baker, K. A., Hilty, C., Peti, W., Prince, A., Pfaffinger, P. J., Wider, G., Wüthrich, K., and Choe, S. (2006). NMR-derived dynamic aspects of N-type inactivation of a Kv channel suggest a transient interaction with the T1 domain. *Biochemistry*, 45(6):1663–72.
- Battiste, J. L. and Wagner, G. (2000). Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry*, 39(18):5355–65.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. W. (2011). iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D: Biological Crystallography*, 67(4):271–281.
- Bertini, I., Luchinat, C., and Piccioli, M. (2001). Paramagnetic Probes in Metalloproteins. *Methods in Enzymology*, 339:314–340.
- Bougueleret, L., Tenchini, M. L., Botterman, J., and Zabeau, M. (1985). Overproduction of the EcoR V endonuclease and methylase. *Nucleic Acids Research*, 13(11):3823–3839.
- Bouvignies, G., Bernadó, P., and Blackledge, M. (2005). Protein backbone dynamics from N-HN dipolar couplings in partially aligned systems: a comparison of motional models

- in the presence of structural noise. *Journal of Magnetic Resonance (San Diego, Calif. : 1997)*, 173(2):328–38.
- Bowen, L. M. and Dupureur, C. M. (2003). Investigation of restriction enzyme cofactor requirements: a relationship between metal ion properties and sequence specificity. *Biochemistry*, 42(43):12643–53.
- Brautigam, C. A., Sun, S., Piccirilli, J. A., and Steitz, T. A. (1999). Structures of normal single-stranded DNA and deoxyribo-3'-S-phosphorothiolates bound to the 3'-5' exonucleolytic active site of DNA polymerase I from *Escherichia coli*. *Biochemistry*, 38(2):696–704.
- Chen, K. and Tjandra, N. (2012). The use of residual dipolar coupling in studying proteins by NMR. *Topics in Current Chemistry*, 326:47–67.
- Clore, G. M. and Iwahara, J. (2009). Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chemical Reviews*, 109(9):4108–4139.
- Clore, G. M. and Schwieters, C. D. (2004). How much backbone motion in ubiquitin is required to account for dipolar coupling data measured in multiple alignment media as assessed by independent cross-validation? *Journal of the American Chemical Society*, 126(9):2923–38.
- Cunningham, T. F., Putterman, M. R., Desai, A., Horne, W. S., and Saxena, S. (2015). The double-histidine Cu<sup>2+</sup>-binding motif: A highly rigid, site-specific spin probe for electron spin resonance distance measurements. *Angewandte Chemie - International Edition*, 54(21):6330–6334.
- Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMR-

- Pipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, 6(3):277–93.
- Dupureur, C. M. (2005). NMR studies of restriction enzyme-DNA interactions: role of conformation in sequence specificity. *Biochemistry*, 44(13):5065–74.
- Dvoretzky, A., Gaponenko, V., and Rosevear, P. R. (2002). Derivation of structural restraints using a thiol-reactive chelator. *FEBS Letters*, 528(1):189–192.
- Emsley, P. and Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, 60(12 I):2126–2132.
- Engler, L. E., Welch, K. K., and Jen-Jacobson, L. (1997). Specific binding by EcoRV endonuclease to its DNA recognition site GATATC. *Journal of Molecular Biology*, 269(1):82–101.
- Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallographica Section D: Biological Crystallography*, 62(1):72–82.
- Frey, M. W., Frey, S. T., Horrocks, W. D., Kaboord, B. F., and Benkovic, S. J. (1996). Elucidation of the metal-binding properties of the Klenow fragment of Escherichia coli polymerase I and bacteriophage T4 DNA polymerase by lanthanide(III) luminescence spectroscopy. *Chemistry & Biology*, 3(5):393–403.
- Gans, P., Hamelin, O., Sounier, R., Ayala, I., Durá, M. A., Amero, C. D., Noirclerc-Savoie, M., Franzetti, B., Plevin, M. J., and Boisbouvier, J. (2010). Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. *Angewandte Chemie - International Edition*, 49(11):1958–1962.
- Gaponenko, V., Howarth, J. W., Gasmi-Seabrook, G., Yuan, J., Columbus, L., Hubbell, W. L., and Rosevear, P. R. (2008). Protein global fold determination using site-directed spin and isotope labeling. *Protein Science*, 9(2):302–309.

- Gardner, K. H. and Kay, L. E. (1997). Production and Incorporation of  $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^2\text{H}$  ( $^1\text{H}$ - $\delta^1$  Methyl) Isoleucine into Proteins for Multidimensional NMR Studies. *Journal of the American Chemical Society*, 119(32):7599–7600.
- Gillespie, J. R. and Shortle, D. (1997). Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *Journal of Molecular Biology*, 268(1):170–184.
- Goto, N. K., Gardner, K. H., Mueller, G. A., Willis, R. C., and Kay, L. E. (1999). A robust and cost-effective method for the production of Val, Leu, Ile ( $\delta^1$ ) methyl-protonated  $^{15}\text{N}$ -,  $^{13}\text{C}$ -,  $^2\text{H}$ -labeled proteins. *Journal of Biomolecular NMR*, 13(4):369–74.
- Hancock, S. P., Hiller, D. A., Perona, J. J., and Jen-Jacobson, L. (2011). The Energetic Contribution of Induced Electrostatic Asymmetry to DNA Bending by a Site-Specific Protein. *Journal of Molecular Biology*, 406(2):285–312.
- Hansen, D. F., Hass, M. A. S., Christensen, H. M., Ulstrup, J., and Led, J. J. (2003). Detection of short-lived transient protein-protein interactions by intermolecular nuclear paramagnetic relaxation: plastocyanin from *Anabaena variabilis*. *Journal of the American Chemical Society*, 125(23):6858–9.
- Hiller, D. A., Fogg, J. M., Martin, A. M., Beechem, J. M., Reich, N. O., and Perona, J. J. (2003). Simultaneous DNA binding and bending by EcoRV endonuclease observed by real-time fluorescence. *Biochemistry*, 42(49):14375–85.
- Hiller, D. A. and Perona, J. J. (2006). Positively charged C-terminal subdomains of EcoRV endonuclease: contributions to DNA binding, bending, and cleavage. *Biochemistry*, 45(38):11453–63.

- Horton, N. C. and Perona, J. J. (2000). Crystallographic snapshots along a protein-induced DNA-bending pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 97(11):5729–34.
- Horton, N. C. and Perona, J. J. (2004). DNA cleavage by EcoRV endonuclease: two metal ions in three metal ion binding sites. *Biochemistry*, 43(22):6841–57.
- Hubbell, W. L., Cafiso, D. S., and Altenbach, C. (2000). Identifying conformational changes with site-directed spin labeling. *Nature Structural Biology*, 7(9):735–739.
- Hubbell, W. L., López, C. J., Altenbach, C., and Yang, Z. (2013). Technological advances in site-directed spin labeling of proteins. *Current Opinion in Structural Biology*, 23(5):725–733.
- Igumenova, T. I., Frederick, K. K., and Wand, A. J. (2006). Characterization of the Fast Dynamics of Protein Amino Acid Side Chains Using NMR Relaxation in Solution. *Chemical Reviews*, 106(5):1672–1699.
- Iwahara, J. and Clore, G. M. (2006). Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature*, 440(7088):1227–1230.
- Iwahara, J., Schwieters, C. D., and Clore, G. M. (2004a). Characterization of nonspecific protein-DNA interactions by  $^1\text{H}$  paramagnetic relaxation enhancement. *Journal of the American Chemical Society*, 126(40):12800–8.
- Iwahara, J., Schwieters, C. D., and Clore, G. M. (2004b). Ensemble approach for NMR structure refinement against  $(^1\text{H})$  paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *Journal of the American Chemical Society*, 126(18):5879–96.

- Iwahara, J., Tang, C., and Marius Clore, G. (2007). Practical aspects of  $^1\text{H}$  transverse paramagnetic relaxation enhancement measurements on macromolecules. *Journal of Magnetic Resonance*, 184(2):185–195.
- Iwahara, J., Zweckstetter, M., and Clore, G. M. (2006). NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. *Proceedings of the National Academy of Sciences*, 103(41):15062–15067.
- Jen-Jacobson, L., Engler, L. E., Ames, J. T., Kurpiewski, M. R., and Grigorescu, A. A. (2000). Thermodynamic Parameters of Specific and Nonspecific Protein-DNA Binding. *Supramolecular Chemistry*, 12(2):143–160.
- Johnson, B. A. and Blevins, R. A. (1994). NMR View: A computer program for the visualization and analysis of NMR data. *Journal of Biomolecular NMR*, 4(5):603–14.
- Kaminker, I., Yagi, H., Huber, T., Feintuch, A., Otting, G., and Goldfarb, D. (2012). Spectroscopic selection of distance measurements in a protein dimer with mixed nitroxide and  $\text{Gd}^{3+}$  spin labels. *Physical Chemistry Chemical Physics*, 14(13):4355.
- Kosen, P. A. (1989). Spin labeling of proteins. *Methods in Enzymology*, 177:86–121.
- Kostrewa, D. and Winkler, F. K. (1995).  $\text{Mg}^{2+}$  binding to the active site of EcoRV endonuclease: a crystallographic study of complexes with substrate and product DNA at 2 Å resolution. *Biochemistry*, 34(2):683–96.
- Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D., and Zakrzewska, K. (2009). Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Research*, 37(17):5917–5929.
- Louis, J. M., Martin, R. G., Clore, G. M., and Gronenborn, A. M. (1998). Preparation of uniformly isotope-labeled DNA oligonucleotides for NMR spectroscopy. *The Journal of Biological Chemistry*, 273(4):2374–8.

- Lovell, S. C., Davis, I. W., Adrendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003). Structure validation by C alpha geometry: phi,psi and C beta deviation. *Proteins-Structure Function and Genetics*, 50(August 2002):437–450.
- Macrae, I. J., Zhou, K., Li, F., Repic, A., Brooks, A. N., Cande, W. Z., Adams, P. D., and Doudna, J. A. (2006). Structural basis for double-stranded RNA processing by Dicer. *Science (New York, N.Y.)*, 311(5758):195–8.
- Martin, A. M., Horton, N. C., Lusetti, S., Reich, N. O., and Perona, J. J. (1999). Divalent metal dependence of site-specific DNA binding by EcoRV endonuclease. *Biochemistry*, 38(26):8430–9.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, 40(4):658–674.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, 67(4):355–367.
- Ogawa, S. and McConnell, H. M. (1967). Spin-label study of hemoglobin conformations in solution. *Proceedings of the National Academy of Sciences of the United States of America*, 58(1):19–26.
- Otting, G. (2010). Protein NMR Using Paramagnetic Ions. *Annual Review of Biophysics*, 39(1):387–405.
- Perona, J. J. and Martin, A. M. (1997). Conformational transitions and structural deforma-

- bility of EcoRV endonuclease revealed by crystallographic analysis. *Journal of Molecular Biology*, 273(1):207–225.
- Pidcock, E. and Moore, G. R. (2001). Structural characteristics of protein binding sites for calcium and lanthanide ions. *Journal of Biological Inorganic Chemistry : JBIC : a publication of the Society of Biological Inorganic Chemistry*, 6(5-6):479–89.
- Pingoud, A., Fuxreiter, M., Pingoud, V., and Wende, W. (2005). Type II restriction endonucleases: structure and mechanism. *Cellular and molecular life sciences : CMLS*, 62(6):685–707.
- Read, R. J. and Schierbeek, A. J. (1988). A phased translation function. *Journal of Applied Crystallography*, 21(5):490–495.
- René, B., Masliah, G., Zargarian, L., Mauffret, O., and Fermandjian, S. (2006). General method of preparation of uniformly <sup>13</sup>C, <sup>15</sup>N-labeled DNA fragments for NMR analysis of DNA structures. *Journal of Biomolecular NMR*, 36(3):137–146.
- Riddles, P. W., Blakeley, R. L., and Zerner, B. (1983). Reassessment of Ellman's reagent. *Methods in Enzymology*, 91:49–60.
- Ruschak, A. M. and Kay, L. E. (2010). Methyl groups as probes of supra-molecular structure, dynamics and function. *Journal of Biomolecular NMR*, 46(1):75–87.
- Shannon, R. D. and IUCr (1976). Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallographica Section A*, 32(5):751–767.
- Sinha, K., Jen-Jacobson, L., and Rule, G. S. (2013). Divide and conquer is always best: Sensitivity of methyl correlation experiments. *Journal of Biomolecular NMR*, 56(4):331–335.

- Solomon, I. (1955). Relaxation Processes in a System of Two Spins. *Physical Review*, 99(2):559–565.
- Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification*, 41(1):207–234.
- Sun, Y., Zhang, Z., Grigoryants, V. M., Myers, W. K., Liu, F., Earle, K. A., Freed, J. H., and Scholes, C. P. (2012). The internal dynamics of mini c TAR DNA probed by electron paramagnetic resonance of nitroxide spin-labels at the lower stem, the loop, and the bulge. *Biochemistry*, 51(43):8530–8541.
- Tang, C., Ghirlando, R., and Clore, G. M. (2008a). Visualization of transient ultra-weak protein self-association in solution using paramagnetic relaxation enhancement. *Journal of the American Chemical Society*, 130(12):4048–56.
- Tang, C., Iwahara, J., and Clore, G. M. (2006). Visualization of transient encounter complexes in protein-protein association. *Nature*, 444(7117):383–6.
- Tang, C., Louis, J. M., Aniana, A., Suh, J.-Y., and Clore, G. M. (2008b). Visualizing transient events in amino-terminal autoprocessing of HIV-1 protease. *Nature*, 455(7213):693–6.
- Tang, C., Schwieters, C. D., and Clore, G. M. (2007). Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature*, 449(7165):1078–82.
- Thiagarajan-Rosenkranz, P., Draney, A. W., Smrt, S. T., and Lorieau, J. L. (2015). A Positively Charged Liquid Crystalline Medium for Measuring Residual Dipolar Couplings in Membrane Proteins by NMR. *Journal of the American Chemical Society*, 137(37):11932–11934.
- Thomas, M. P., Brady, R. L., Halford, S. E., Sessions, R. B., and Baldwin, G. S. (1999).

- Structural analysis of a mutational hot-spot in the EcoRV restriction endonuclease: a catalytic role for a main chain carbonyl group. *Nucleic Acids Research*, 27(17):3438–45.
- Todd, A. P., Cong, J., Levinthal, F., Levinthal, C., and Hubell, W. L. (1989). Site-directed mutagenesis of colicin E1 provides specific attachment sites for spin labels whose spectra are sensitive to local conformation. *Proteins: Structure, Function, and Genetics*, 6(3):294–305.
- Tugarinov, V., Hwang, P. M., and Kay, L. E. (2004). Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins. *Annual Review of Biochemistry*, 73:107–46.
- Tugarinov, V. and Kay, L. E. (2003). Ile, Leu, and Val Methyl Assignments of the 723-Residue Malate Synthase G Using a New Labeling Strategy and Novel NMR Methods. *Journal of the American Chemical Society*, 125(45):13868–13878.
- Tugarinov, V. and Kay, L. E. (2005). Methyl groups as probes of structure and dynamics in NMR studies of high-molecular-weight proteins. *ChemBioChem*, 6(9):1567–1577.
- Urzhumtseva, L., Afonine, P. V., Adams, P. D., and Urzhumtsev, A. (2009). Crystallographic model quality at a glance. *Acta Crystallographica Section D: Biological Crystallography*, 65(3):297–300.
- Vega, S., Abian, O., and Velazquez-Campoy, A. (2016). On the link between conformational changes, ligand binding and heat capacity. *Biochimica et Biophysica Acta - General Subjects*, 1860(5):868–878.
- Venditti, V., Fawzi, N. L., and Clore, G. M. (2011). Automated sequence- and stereo-specific assignment of methyl-labeled proteins by paramagnetic relaxation and methyl-methyl nuclear overhauser enhancement spectroscopy. *Journal of Biomolecular NMR*, 51(3):319–328.

- Vipond, I. B., Baldwin, G. S., and Halford, S. E. (1995). Divalent metal ions at the active sites of the EcoRV and EcoRI restriction endonucleases. *Biochemistry*, 34(2):697–704.
- Volkov, A. N., Worrall, J. A. R., Holtzmann, E., and Ubbink, M. (2006). Solution structure and dynamics of the complex between cytochrome c and cytochrome c peroxidase determined by paramagnetic NMR. *Proceedings of the National Academy of Sciences of the United States of America*, 103(50):18945–50.
- Welch, M., Govindarajan, S., Ness, J. E., Villalobos, A., Gurney, A., Minshull, J., and Gustafsson, C. (2009). Design parameters to control synthetic gene expression in *Escherichia coli*. *PloS One*, 4(9):e7002.
- Winkler, F. K., Banner, D. W., Oefner, C., Tsernoglou, D., Brown, R. S., Heathman, S. P., Bryan, R. K., Martin, P. D., Petratos, K., and Wilson, K. S. (1993). The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *The EMBO Journal*, 12(5):1781–95.
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*, 67(4):235–242.
- Zahran, M., Daidone, I., Smith, J. C., and Imhof, P. (2010). Mechanism of DNA recognition by the restriction enzyme EcoRV. *Journal of Molecular Biology*, 401(3):415–432.
- Zwart, P. H. (2005). Anomalous signal indicators in protein crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 61(11):1437–1448.

# **Part I**

## **Appendix**

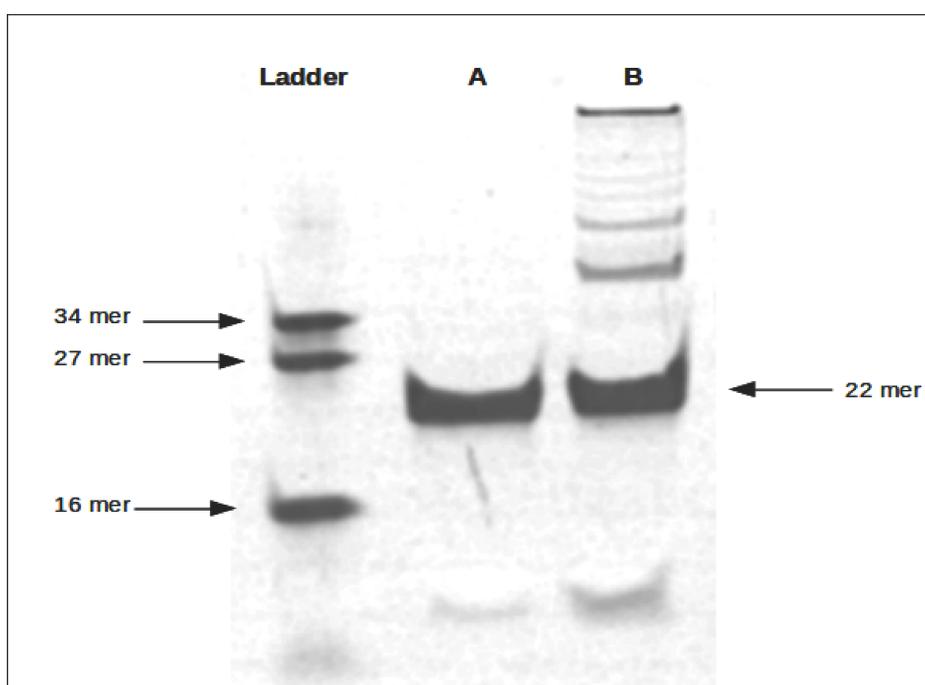
---

## 1. Production of isotopically labeled DNA substrate for DNA dynamics

The template DNA used for PCR based ESRA approach (Louis et al., 1998):

**CTGGCAAAGATATCTTTTCGCAGCTGGCAAAGATATCTTTTCGCAG**

The sequence in bold letters is the PvuII site. PCR conditions were optimized and the extension times were gradually increased as the DNA length increased to maximize the yield (René et al., 2006). PCR product was digested with PvuII followed by ion exchange chromatography to obtain 22 mer with the target sequence (**CTGGCCGTGATATCACGCGCAG**). Figure 1 on page 106, shows the PCR product after digestion with PvuII in lane B and the purified 22 mer in lane A. The final yield was ~1 mg of 22 mer from 10 mL PCR reaction, which is sufficient for a 0.2 mM (dimer) 400  $\mu$ L EcoRV-DNA complex NMR sample.



**Figure 1: Production of isotopically labeled DNA.** 20 % Native PAGE gel, EtBR stained. **Lane B:** PCR product after digestion with PvuII. **Lane A:** Purified isotopically labeled DNA after ion exchange chromatography following PvuII digestion.

## 2. Methyl assignment

For all methyl spectra,  $^{13}\text{C}$  chemical shifts were referenced to TMS. In order to compare them with chemical shifts referenced with DSS, 2.5 ppm needs to be added to the carbon shifts.

**EcoRV-DNA nometal complex**

<b>Assignment</b>	<b><sup>1</sup>H_PPM</b>	<b><sup>13</sup>C_PPM</b>
I8	11.084	0.686
I24	12.443	0.733
I30	12.013	0.627
I43	10.811	0.509
I51	11.388	1.013
I52	11.222	0.032
I55	6.435	0.738
I62	10.084	0.669
I87	11.022	0.397
I89	12.224	0.782
I91	8.925	1.141
I103	11.07	0.628
I114	10.119	0.615
I129	11.794	0.875
I133	8.598	0.954
I134	11.655	0.779
I153	11.973	0.953
I159	12.474	0.487
I176	12.081	0.925
I189	9.678	0.282
I192	9.698	-0.066
I205	11.406	0.338
I233	13.126	0.591
I240	9.093	0.635
L3pR	21.128	0.431
L3pS	22.591	0.327
L7pR	21.844	0.789
L7pS	23.331	0.793
L11pR	23.783	0.586
L77pR	21.618	0.725
L77pS	24.018	0.506
V63pR	19.987	1.139
V63pS	19.631	0.815
L107pR	23.673	0.762
L107pS	20.525	0.801
L148pR	22.378	0.963
L148pS	20.754	0.88
L156pR	22.102	0.869
L156pS	21.447	0.782
L170pR	24.208	0.57
L170pS	20.84	0.019
L180pR	24.406	1.162
L180pS	21.024	1.259

---

Continued ..

<b>Assignment</b>	<b><sup>1</sup>H_PPM</b>	<b><sup>13</sup>C_PPM</b>
L213pR	22.535	1.133
L213pS	20.584	0.91
L225pR	22.961	1.019
L225pS	19.445	0.87
V122pR	18.874	0.825
V122pS	20.381	1.298
V137pR	18.122	0.801
V137pS	19.11	0.856
V141pR	18.508	0.956
V141pS	17.74	0.815
V166pR	18.556	0.842
V166pS	19.989	1.04
V168pR	16.895	0.633
V168pS	20.103	0.135
V175pR	18.457	0.356
V175pS	18.694	-0.418
V200pR	18.15	0.909
V200pS	19.402	0.547

**EcoRV-DNA Lu<sup>3+</sup> complex**

<b>Assignment</b>	<b><sup>1</sup>H_PPM</b>	<b><sup>13</sup>C_PPM</b>
I8	11.138	0.738
I24	12.778	0.79
I30	12.277	0.687
I43	9.901	0.404
I52	11.198	0.32
I55	6.166	0.744
I62	9.766	0.689
I87	10.916	0.406
I91	11.264	1.232
L11	23.682	0.652
L40	20.257	1.193
L46	19.644	0.63
V20	19.09	0.591
V20	18.649	0.554
I103	11.159	0.692
I114	9.999	0.652
I129	11.654	0.878
I133	8.105	0.907
I134	12.274	0.791
I153	11.807	0.98
I159	12.582	0.503
I176	11.969	0.935
I189	9.937	0.311
I192	9.654	-0.058
I205	11.246	0.385
I233	13.136	0.594
I240	9.031	0.646
L3pR	20.628	0.5
L3pS	22.479	0.299
L7pR	21.739	0.819
L7pS	23.342	0.825
V39	18.293	0.956
L33pR	21.436	0.166
L33pS	20.343	0.468
L77pR	22.068	0.707
L77pS	24.04	0.516
V63pR	19.095	1.052
V63pS	19.954	0.92
L107pR	23.798	0.817
L107pS	20.599	0.817
L148pR	22.59	0.98
L148pS	21.415	0.903
L156pR	22.23	0.943

---

Continued ..

<b>Assignment</b>	<b><sup>1</sup>H_PPM</b>	<b><sup>13</sup>C_PPM</b>
L156pS	21.534	0.796
L170pR	24.283	0.599
L170pS	20.692	0.012
L180pR	24.296	1.185
L180pS	20.926	1.272
L213pR	22.502	1.157
L213pS	20.384	0.935
L225pR	22.901	1.035
L225pS	19.378	0.888
L40pR	23.918	0.91
V122pR	18.878	0.858
V122pS	20.313	1.268
V137pR	18.04	0.823
V137pS	18.941	0.892
V141pR	18.523	0.981
V141pS	17.571	0.821
V166pR	19.737	0.893
V166pS	20.172	1.086
V168pR	16.698	0.567
V168pS	19.049	0.067
V175pR	18.368	0.335
V175pS	18.708	-0.336
V200pR	18.041	0.904
V200pS	19.333	0.552

### 3. Backbone NH <sup>1</sup>H-Transverse relaxation rate constants and PRE rates

#### S2C amide

Residue	Dia		Para		PRE	
	$R_{2,dia}$ , s <sup>-1</sup>	error	$R_{2,para}$ , s <sup>-1</sup>	error	$\Gamma_2$ , s <sup>-1</sup>	error
Val63	91.6	2.1	64.7	4.4	26.9	4.4
Lys102	68.1	3.5	52.3	3.2	15.8	3.5
Tyr163	71	8	62.7	3.1	8.3	8
Ala177	108	7	76	10	32	10
Leu180	71	9	52.4	2.6	7	5
Ala181	75.2	3.4	66.6	3.5	16	4
Gly184	87	3.5	75	7	12	7
His193	79.3	0.7	69	6	6	2
Ala194	74	10	49	7	25	10
His195	58	7	54.4	3.4	7	4
Phe199	87	5	64	8	23	8
Gly202	133	9	92	12	41	12
Gly204	70	8	51	6	18	6
Ser208	92	5	77.3	4.5	8	4
Asp210	109	5	67	6	42	6
Glu211	119.1	4.5	73.8	5	45.3	5
Arg217	114	16	89	14	25	10
Tyr219	96.7	3.7	88.9	4.1	7.8	4.1
Glu220	85.2	4.9	79.2	4.3	6	4.9
Arg221	71.7	3.7	71.4	3	0.3	3.7
Leu225	108	6	80	6	28	6
Arg226	95.4	4.9	80.3	2.7	22	5
Lys229	73.1	2.1	62	8	11.1	8
Tyr230	74	7	57	5	14	5
Asn231	97	14	85	6	12	14
Asn232	89.2	4.6	77	9	12.2	9
Arg242	68	6	62	6	17	6
Gly243	93	5	74	2.5	19	5
Lys245	61	1.8	49.5	1.5	11.5	1.8

**S2C amide**

Residue	Dia		Para		PRE	
	$R_{2,dia}$ s <sup>-1</sup>	error	$R_{2,para}$ s <sup>-1</sup>	error	$\Gamma_2$ s <sup>-1</sup>	error
Val63	109	10	85.4	4.4	10	6
Lys102	56.1	2.6	58.5	3	0	3
Tyr163	64.2	3.4	53	6	11.2	6
Ala177	93	9	85	6	9	8
Leu180	80.8	4.9	76	1	4.8	4.9
Ala181	64.3	2.1	72.3	4.9	0	4.9
Gly184	90	6	94.7	2.3	0	6
His193	79.3	3.3	84	7	0	4
Ala194	79	10	78	7	1	10
His195	51.6	4	57.6	1.9	0	4
Phe199	66.4	1.3	67.1	2	3	1
Gly202	117	7	79	12	38	12
Gly204	59.8	4.7	48.4	2.9	12	5
Ser208	98.1	4.2	72.9	3.9	25.2	4.2
Asp210	101	5	67.8	3.5	33	4
Glu211	106.9	4	89.6	3.6	17.3	4
Arg217	112	8	101.7	3.6	10.3	8
Tyr219	107	15	95	16	5	12
Glu220	74	7	72.9	4.1	8	6
Arg221	76	6	66	1.4	6	1
Leu225	105	5	87	5	12	3
Arg226	85	9	69	10	10	7
Lys229	81.1	3.7	70	6	11.1	4
Tyr230	74.2	3.6	72	8	2.2	8
Asn231	83.5	2.6	89	9	0	9
Asn232	82	5	84.1	2.1	0	5
Arg242	69	5	69.3	1.1	0	4
Gly243	91	6	91.7	3.6	0	6
Lys245	55.2	2.7	49.7	3.1	5.5	3.1

**K197C amide**

Residue	Dia		Para		PRE	
	$R_{2,dia}$ , s <sup>-1</sup>	error	$R_{2,para}$ , s <sup>-1</sup>	error	$\Gamma_2$ , s <sup>-1</sup>	error
Arg4	139	9	78.2	2.6	64	10
Asp6	127	10	68.7	1	58.3	10
His59	103	10	80.3	3.2	16	11
Gly60	109.2	3.7	69	5	40.2	5
Val63	63.4	1.7	50	2.1	13.4	2.1
Thr76	112	6	84	7	16	9
Asn84	115	7	80.5	3.8	34.5	7
Lys85	65.8	1.7	60.9	2.3	4.9	2.3
Lys86	70	5	64.9	3.4	3	3
Ile103	148	11	73.2	3.7	74.8	11
Asp126	65.9	3.6	71.2	4.6	0	4.6
Tyr128	82.4	2.5	73.8	3.9	8.6	3.9
Ile129	85.4	1.3	83.6	2.4	1.8	2.4
Ala130	109	18	68	7	41	18
His131	99.8	5	83.7	2.5	16.1	5
Tyr163	74.6	4	67	6	7.6	6
Lys173	89.3	2.5	63.6	1.7	25.7	2.5
Ile176	123	5	87.9	3	35.1	5
Ala177	113.9	3.1	81.9	4	32	4
Gly178	86.1	1.6	75	6	11.1	6
Leu180	75.4	2.6	63.9	4.8	11.5	4.8
Ser183	94	2.8	85	1	9	2.8
Ser191	85.7	4.5	78.7	2.9	7	4.5
His193	82	5	69.6	1.7	12.4	5
Gly204	105	8	44	4.8	61	8
Ser208	78	4.9	74.3	0.8	3.7	4.9
Trp216	85.7	1.7	75.6	3.5	10.1	3.5
Arg217	90.2	1.1	79.8	3	10.4	3
Asn218	110	6	103.8	3.9	6.2	6
Tyr219	88.2	4.1	95	5	0	5
Glu220	74.7	2	67.7	1.2	7	2
Gln224	83	4.8	62.2	1.6	16	2
Leu225	94.6	3.6	88.9	2.9	5.7	3.6
Arg226	88.6	0.9	76.1	4.8	12.5	4.8
Ile233	75	8	76	7	4	9
Tyr236	85.1	4.1	72.9	2.8	12.2	4.1
Arg237	64.6	3.6	68.1	0.6	0	3.6
Asn238	71	6	69.8	1.5	1.2	6
Trp239	86.2	4.6	74.2	5	12	5
Tyr241	94.9	3.8	75.2	5	19.7	5
Arg242	73	8	64.7	3.4	8.3	8
Gly243	88	6	80.5	4.7	7.5	6

**K197C amide**

Residue	Lu <sup>3+</sup> Dia		Para		PRE	
	$R_{2,dia}$ s <sup>-1</sup>	error	$R_{2,para}$ s <sup>-1</sup>	error	$\Gamma_2$ s <sup>-1</sup>	error
Arg4	140	13	84.1	2.8	75	4
Asp6	133	10	80.1	4.2	62	14
His59	84	5	77	9	2	2
Gly60	103	7	91.5	4.2	12	6
Val63	90.7	4.2	86	8	4.7	8
Thr76	96	9	126	7	0	7
Asn84	96	3.9	77.4	4	18.6	4
Lys85	70.9	2.9	64.6	2	6.3	2.9
Lys86	95	7	79.3	2	15.7	7
Ile103	163	20	88.4	3.4	62	18
Asp126	72.2	3.3	66.9	2.6	5.3	3.3
Tyr128	77	5	76	5	2	4
Ile129	88.5	4.1	81	6	8	5
Ala130	91	10	75	6	16	10
His131	81.3	2.8	75.6	3.3	5.7	3.3
Tyr163	72.4	3.9	64.7	4.8	4	5
Lys173	94.9	2.5	71.4	1.5	23.5	2.5
Ile176	203	14	124	16	79	16
Ala177	134.9	4.5	85	8	57	5
Gly178	72.1	4.2	72	11	2	7
Leu180	67.3	1.9	79	6	0	6
Ser183	77.5	4.4	97	9	0	9
Ser191	117	3.4	78	13	39	13
His193	78.9	3.4	62.1	1.5	16.8	3.4
Gly204	114.4	4.5	47.4	3.4	67	4.5
Ser208	84.5	4	74.9	4.2	9.6	4.2
Trp216	83	1.8	83.5	3.5	0	3.5
Arg217	90.8	4.4	107	6	0	6
Asn218	76	2.4	86.4	2.6	0	2.6
Tyr219	96	8	89	6	11	5
Glu220	71.1	1.6	72	6	0	6
Gln224	74.8	2.9	74	4	0.8	4
Leu225	90.7	4.1	91	8	4	4
Arg226	79.5	3.4	75	6	6	2
Ile233	90	9	64.6	3.4	15	5
Tyr236	83.6	3.9	78.3	2.6	5.3	3.9
Arg237	76.5	2.3	70	5	6.5	5
Asn238	89.9	4.5	72	6	17.9	6
Trp239	78.7	3.1	77	10	1.7	10
Tyr241	101	8	74.2	4.5	31	5
Arg242	77.6	3.4	69.2	4	8.4	4
Gly243	90.3	3	76	3	14.3	3

**S234C amide**

Residue	Dia		Para		PRE	
	$R_{2,dia}$ , s <sup>-1</sup>	error	$R_{2,para}$ , s <sup>-1</sup>	error	$\Gamma_2$ , s <sup>-1</sup>	error
Asp6	91.8	3.9	74	5	17.8	5
Val63	84.8	2.4	50.8	3.1	34	3.1
Lys79	83.3	3.5	75	7	6	4
Asn84	84.3	3.2	78.1	4.3	6.2	4.3
Lys102	76.3	2	57.6	1.1	18.7	2
Ile103	116	8	93.4	4.6	26	6
His131	130.8	3.1	89.5	1.2	41.3	3.1
Tyr163	93	3.2	86	9	29	3
Phe169	109.2	4.1	75.5	4.6	33.7	4.6
Asp172	111.7	3	78	6	22	6
Ala177	118	6	87.2	3.9	30.8	6
Ala181	92.6	0.6	69	7	23.6	7
Ala194	95.5	3.5	66.1	3.9	29.4	3.9
His195	104	6	64	6	40	6
Tyr196	87.7	4.5	75.1	3.6	13	4
Asp198	121	8	93.3	4.5	27.7	8
Gly204	129.6	4.6	52.3	1.8	77.3	4.6
Asp210	98	6	70	6	28	6
Phe212	178	8	78.1	3.6	89	7
Arg221	81	2.2	79.4	1.9	1.6	2.2
Gln224	115	6	99.8	4.6	15.2	6
Leu225	107.2	4.8	103	6	4.2	6
Lys245	162	9	56.6	2	105.4	9

**S234C amide**

Residue	Dia		Para		PRE	
	$R_{2,dia}$ s <sup>-1</sup>	error	$R_{2,para}$ s <sup>-1</sup>	error	$\Gamma_2$ s <sup>-1</sup>	error
Asp6	116	8	84.8	4.5	31.2	8
Val63	101	6	71.7	3	29.3	6
Lys79	78.7	2.6	67.5	1.5	11.2	2.6
Asn84	96.1	3.8	85.6	4.2	10.5	4.2
Lys102	81	3.8	56.8	2.7	20	3
Ile103	114.2	3.4	81.2	1.8	33	3.4
His131	83.1	3.4	72	6	11.1	6
Tyr163	74	11	52	15	7	4
Phe169	91	12	76	10	19	3
Asp172	97.3	3.9	74.1	2.7	23.2	3.9
Ala177	130	6	96.2	1.7	33.8	6
Ala181	115.4	3.7	66.1	2.1	49.3	3.7
Ala194	156	11	88.2	3.1	67.8	11
His195	85.6	3.9	62.5	2.7	23.1	3.9
Tyr196	101.1	4.1	75.9	4.3	25.2	4.3
Asp198	104.2	3.5	96.3	1.1	7.9	3.5
Gly204	168.4	3.1	46.1	2.2	122.3	3.1
Asp210	100.3	3.7	71.3	4.1	29	4.1
Phe212	150	10	85.4	3.8	64.6	10
Arg221	86	5	72.8	3.5	9	6
Gln224	92	1.6	89	5	3	5
Leu225	119	7	128	6	7	3
Lys245	114	10	48.3	4.4	65.7	10

## 4. Side-chain methyl $^1\text{H}$ -Transverse relaxation rate constants and PRE rates

### S2C methyl

Residue	Dia		Para		PRE	
	$R_{z,dia}, \text{s}^{-1}$	error	$R_{z,para}, \text{s}^{-1}$	error	$\Gamma_2, \text{s}^{-1}$	error
I62	24.7	2.1	10.3	2.1	14.4	2.1
I103	41	2.7	33.4	1.5	7.6	2.7
L107pR	40.3	2.8	33.5	1.5	6.8	2.8
L107pS	36.8	2.4	26.9	1.5	9.9	2.4
V137pS	26.9	1.1	23.1	0.9	3.8	1.1
V166pS	29.9	1.6	24.5	1.3	5.4	1.6
L180pS	28.2	1.7	22.3	1.1	5.9	1.7
I189	29.6	3.3	27.9	1.6	1.7	3.3
I192	34.8	2	24.3	1.9	10.5	2
V200pS	40.8	1.4	27.3	1.2	13.5	1.4
L213pR	26.1	1.7	17	2	9.1	2
I240	18.8	3	13.9	2.5	4.9	3

**S2C methyl**

<b>Lu<sup>3+</sup></b>	<b>Dia</b>		<b>Para</b>		<b>PRE</b>	
<b>Residue</b>	<b>R<sub>2,dia</sub>, s<sup>-1</sup></b>	<b>error</b>	<b>R<sub>2,para</sub>, s<sup>-1</sup></b>	<b>error</b>	<b>Γ<sub>2</sub>, s<sup>-1</sup></b>	<b>error</b>
I62	22.5	1.5	9.8	1.9	12.7	1.9
I103	28.1	0.7	22.8	1.6	5.3	1.6
L107pR	42.2	1.9	35.7	2.9	6.5	2.9
L107pS	40.8	1.6	30.3	1.8	10.5	1.8
V137pS	32.6	0.7	24.8	0.8	7.8	0.8
V166pS	37.1	0.6	34.5	1.3	2.6	1.3
L180pS	28.1	1.4	22.4	1.4	5.7	1.4
I189	32.5	2	26.7	1.8	5.8	2
I192	36.7	1.6	27.7	2.1	9	2.1
V200pS	47.3	2.9	22.9	1.3	24.4	2.9
L213pR	28.5	1.6	16.6	1	11.9	1.6
I240	24	2	14.7	3.4	9.3	3.4

**K197C methyl**

<b>Nometal</b>	<b>Dia</b>		<b>Para</b>		<b>PRE</b>	
<b>Residue</b>	$R_{2,dia}, s^{-1}$	<b>error</b>	$R_{2,para}, s^{-1}$	<b>error</b>	$\Gamma_2, s^{-1}$	<b>error</b>
L3pR	31.9	3.2	16.6	1.7	15.3	3.2
L7pS	44.7	3.9	27.2	1.4	17.5	3.9
I55	37.4	1.7	18	1.1	19.4	1.7
I62	24.2	0.8	13.2	1.1	11	1.1
V63pR	39.8	2.7	34.7	1.6	5.1	2.7
V63pS	41.3	1.9	24.3	2.7	17	2.7
L77pR	34.3	2.9	30.8	2.6	3.5	2.9
L77pS	36.9	2.4	28.5	2.5	8.4	2.5
I91	43.2	3.2	26.5	2.2	16.7	3.2
L107pR	45.4	1	31.3	1.3	14.1	1.3
L107pS	51.6	2.5	28.7	1.5	22.9	2.5
I133	42.2	2.1	27.1	1.1	15.1	2.1
V141pR	27.2	1.4	15.45	0.39	11.75	1.4
V141pS	38	2.6	17.5	1.2	20.5	2.6
V166pS	35.5	1.3	32.2	1.3	3.3	1.3
L180pR	38	1.3	32	2.2	6	2.2
I189	33.5	2	27.5	1.3	6	2
I205	31.7	2	24	1.3	7.7	2
L213pR	20.7	2	16.6	1.7	4.1	2
I233	22.7	2.3	13.7	1.8	9	2.3
I240	20.7	1.7	14	2.1	6.7	2.1

**K197C methyl**

<b>Lu<sup>3+</sup></b>	<b>Dia</b>		<b>Para</b>		<b>PRE</b>	
<b>Residue</b>	<b>R<sub>2,dia</sub>, s<sup>-1</sup></b>	<b>error</b>	<b>R<sub>2,para</sub>, s<sup>-1</sup></b>	<b>error</b>	<b>Γ<sub>2</sub>, s<sup>-1</sup></b>	<b>error</b>
L3pR	47.8	2.2	27.4	1.8	20.4	2.2
L7pS	53.9	2.4	29.6	0.7	24.3	2.4
I55	44.7	0.9	21.9	2.8	22.8	2.8
I62	22.5	0.7	11.6	1.5	10.9	1.5
V63pR	51.6	1.5	41.2	1.3	10.4	1.5
V63pS	54.04	0.24	43.7	3.6	10.34	3.6
L77pR	44.5	1.7	32.8	3.7	11.7	3.7
L77pS	41	3	28.3	2.2	12.7	3
I91	40.7	0.7	23.3	0.5	17.4	0.7
L107pR	57	2.4	40.2	1.7	16.8	2.4
L107pS	49.1	1.2	35.5	3.8	13.6	3.8
I133	40	2.3	21.7	1	18.3	2.3
V141pR	25.9	1.2	14.2	1.2	11.7	1.2
V141pS	31.8	0.8	18.9	0.7	12.9	0.8
V166pS	46.8	2.1	39.8	1.9	7	2.1
L180pR	39	2.3	31.7	1.6	7.3	2.3
I189	33.9	1.9	31.2	3	2.7	3
I205	39.6	1.3	23	0.9	16.6	1.3
L213pR	23	1.7	17.8	1.1	5.2	1.7
I233	21.8	1.7	15.2	1.7	6.6	1.7
I240	27	1	14.7	2	12.3	2

**S234C methyl**

Residue	Dia		Para		PRE	
	$R_{z,dia}$ , s <sup>-1</sup>	error	$R_{z,para}$ , s <sup>-1</sup>	error	$\Gamma_2$ , s <sup>-1</sup>	error
L7pS	28.4	1.1	23.4	1.2	5	1.2
V63pR	38.6	1.1	35.3	1	3.3	1.1
V63pS	37.9	1.9	25.7	1.5	12.2	1.9
L77pR	28.1	1.3	24.4	2	3.7	2
L77pS	36	1.4	25.8	1.1	10.2	1.4
I91	31.6	1.4	26	1.4	5.6	1.4
I103	36.2	1.5	24.4	2.1	11.8	2.1
L107pR	50.8	1.7	31.7	1.2	19.1	1.7
L170pR	31.9	1.2	24	1.5	7.9	1.5
I133	33.5	1.2	27.2	1.3	6.3	1.3
V137pS	31.1	2	22.9	2	8.2	2
V141pR	23.1	0.5	15.4	0.9	7.73	0.9
V141pS	28.8	1.6	18.8	0.9	10	1.6
V168pR	43.4	1.7	32.8	0.6	10.6	1.7
V168pS	45.4	3.4	38	1.2	7.4	3.4
L170pS	36.8	2.3	22.1	1.3	14.7	2.3
V175pR	59.5	2.3	24.1	1.1	35.4	2.3
V175pS	70.6	3.4	29	2.1	41.6	3.4
I189	43.4	2.8	28.7	2.2	14.7	2.8
L213pR	32.4	1.3	17.7	0.5	14.7	1.3
L225pR	29.7	1.4	16.9	0.8	12.8	1.4
L225pS	26.8	1.2	13.75	0.36	13.05	1.2

**S234C methyl**

<b>Lu<sup>3+</sup></b>	<b>Dia</b>		<b>Para</b>		<b>PRE</b>	
<b>Residue</b>	<b>R<sub>2,dia</sub>' s<sup>-1</sup></b>	<b>error</b>	<b>R<sub>2,para</sub>' s<sup>-1</sup></b>	<b>error</b>	<b>Γ<sub>2</sub>' s<sup>-1</sup></b>	<b>error</b>
L7pS	33.6	0.9	25.2	1	8.4	1
V63pR	52.9	2.3	45.9	2.9	7	2.9
V63pS	50	2.4	37.8	1.7	12.2	2.4
L77pR	39.4	1.5	32.3	2.1	7.1	2.1
L77pS	41.1	1.9	28.9	1.4	12.2	1.9
I91	23.6	1.4	21	1	2.6	1.4
I103	41.5	2.1	22.5	1	19	2.1
L107pR	57.9	2.1	38.5	1.5	19.4	2.1
L170pR	38.9	2	26.3	1.3	12.6	2
I133	22.9	0.9	19.5	1.3	3.4	1.3
V137pS	34.6	2.3	25.5	1.7	9.1	2.3
V141pR	27.3	0.7	17.13	0.35	10.17	0.7
V141pS	30.7	1.6	18	1.3	12.7	1.6
V168pR	74.2	4.7	52.7	2	9	3.7
V168pS	47.4	2.3	47.8	2.5	8	2.8
L170pS	40.7	1.4	25.7	1.2	15	1.4
V175pR	64.3	2	25.8	1.5	38.5	2
V175pS	76.4	2.1	36.5	2.2	39.9	2.2
I189	41.4	2.3	31.7	1.9	9.7	2.3
L213pR	30.4	0.7	22.1	1.5	8.3	1.5
L225pR	28.6	1	20	1.2	8.6	1.2
L225pS	30	1.5	16.4	0.9	13.6	1.5

## **Part II**

### **Appendix**

# 1. Pulse programs

Pulse programs of  $^1\text{H}$ - $^{13}\text{C}$  methyl HMQC used for methyl assignment by site-directed mutagenesis and interleaved versions of pulse programs used for measuring  $^1\text{H}$ - $\Gamma_2$  PRE rates for backbone amides and methyl side-chains are given:

## 1. Methyl-HMQC

```

;file name - hmqcetgp.ks
;avance-version (02/07/15)
;HMQC
;2D H-1/X correlation via heteronuclear zero and double quantum
; coherence
;phase sensitive using Echo/Antiecho gradient selection
;with decoupling during acquisition
;using shaped pulses for inversion on f2 - channel

#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

#define LABEL_CN

"p2=p1*2"

"p4=p3*2"

"d0=3u"

"d2=1s/(cnst2*2)"

"d11=30m"

"DELTA1=d2-p16-8u"

# ifdef LABEL_CN
"DELTA=larger(p2,p22)+d0*2"
"DELTA2=p16+d16+d0+larger(p2,p22)/2+p4/2-p14/2-4u" ; -p4 replaces -p14
# else
"DELTA=p2+d0*2"
"DELTA2=p16+d16+d0+p2/2+p4/2-p14/2-4u" ; -p4 replaces -p14
# endif /*LABEL_CN*/

1 ze
2 d1 do:f2
3 p1 ph1
  d2
  DELTA2 p12:f2
  4u
  (p4 ph5):f2 ;(p14:sp3 ph5:r):f2 ;(p4 ph5):f2
  4u
  DELTA2 UNBLKGRAD ;p12:f2
  (p3 ph3):f2
  d0

# ifdef LABEL_CN
  (center (p2 ph2) (p22 ph1):f3 )

```

```

# else
  (p2 ph2)
# endif /*LABEL_CN*/

d0
p16:gp1*EA
d16
  (p4 ph4):f2
  DELTA
p16:gp1*-1*EA
d16
  (p3 ph4):f2
  4u
p16:gp2
  DELTA1 p112:f2
  4u BLKGRAD
  go=2 ph31 cpd2:f2
  d1 do:f2 mc #0 to 2
  F1EA(igrad EA, id0 & ip3*2 & ip5*2 & ip31*2)
exit

ph1=0
ph2=0 0 2 2
ph3=0 2
ph4=0 0 0 0 2 2 2 2
ph5=0
ph31=0 2 0 2 2 0 2 0

;p10 : 120dB
;p11 : f1 channel - power level for pulse (default)
;p12 : f2 channel - power level for pulse (default)
;p13 : f3 channel - power level for pulse (default)
;p112: f2 channel - power level for CPD/BB decoupling
;sp3: f2 channel - shaped pulse 180 degree
;p1 : f1 channel - 90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p3 : f2 channel - 90 degree high power pulse
;p4 : f2 channel - 180 degree high power pulse
;p14: f2 channel - 180 degree shaped pulse for inversion
;p16: homospoil/gradient pulse [1 msec]
;p22: f3 channel - 180 degree high power pulse
;d0 : incremented delay (2D) [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d2 : 1/(2J(XH))
;d11: delay for disk I/O [30 msec]
;d16: delay for homospoil/gradient recovery
;cnst2: = J(XH)
;in0: 1/(2 * SW(X)) = DW(X)
;nd0: 2
;NS: 2 * n
;DS: 16
;td1: number of experiments
;FnMODE: echo-antiecho
;cpd2: decoupling according to sequence defined by cpdprg2
;pcpd2: f2 channel - 90 degree pulse for decoupling sequence

;use gradient ratio: gp 1 : gp 2
; 80 : 40.2

;for z-only gradients:
;gpz1: 80%
;gpz2: 40.2%

```

```
;use gradient files:  
;gpnam1: SINE.100  
;gpnam2: SINE.100
```

```
                                ;preprocessor-flags-start  
;LABEL_CN: for C-13 and N-15 labeled samples start experiment with  
; option -DLABEL_CN (eda: ZGOPTNS)  
                                ;preprocessor-flags-end
```

```
;$Id: hmqcetgp,v 1.1.2.1 2002/07/16 12:59:33 ber Exp $
```

## 2. $^1\text{H}$ - $\Gamma_2$ PRE for $^{13}\text{C}$ -attached protons

```

;file name - pre_hmqc.gr

;sample should be in 100% D2O buffer, no water suppression
; For measurement of T2 of carbon attached protons using HMQC
; Quad detection by States-ttqi
;
; Modified by GR from original sequence by J. Iwahara
; JACS 126, 5879-5896
;
; Measured in an interleaved manner
; l8 : number of time points
; For l8=2/4 (two or four timepoints), you can split the spectra
; using the "splitinvnoe" /"splitinv_4" command
; this command will overwrite n & n+1 expt
; acquires Re*L8 and Im * L8
;
;
; in11 (interval of T-delays)*0.5
; Default set to 2m (see below)
;
; l1 = td1/2*l8
;
; programed by Junji Iwahara
; 6/13/03
;
#include <Avance.incl>
#include <Grad_sys.incl>
#include <Grad.incl>
#include <Delay.incl>
;
;CHANNEL DEFINITIONS
;H1 f1 o1p on water
; p1 90 H at p11 on f1 (hard 90)
; p11 power level for hard 90
; p10 zero power
;
;C-ME f2 o2p 19 ppm (methyl region)
; p2 90 C at p12 on F2 (hard 90)
; p12 pwr level for 90 C
; p112 power for carbon decoupling
; pcpd2 90 deg pulse length at p112
; cpdprg2 GARP-1
;
;Define increment for proton relaxation:

"in11=2m"
"in12=in11"

;GRADIENT TIMES
; x y z
"ptime1=1000u" ;0 0 40 G/cm G_a purge C13
"ptime2=500u" ;25 0 50 G/m G-b proton/carbon 180
"grex=150u"

define delay taua ;1st HMQC Pol transfer
define delay taub ;2nd HMQC Pol transfer

#define INTERLEAVE

"d2=3.7m" ;=1/2J_CH

"taua=d2-gtime2-grec" ;Polarization transfer
"taub=d2-gtime2-grec-20u-400u" ;Polarization transfer
"d11=3u" ;1st relax period (T/2)

```

```

;"d12=d11"
"d12=d11+0.6366*p1"

"l1 = td1/2*18"

;for aliasing in 13C
;"d0=3u"
"d0=(in0*0.5)-2.0*(p2/3.14)-p1"

;pulse sequence.

  30m BLKGRAD ;Allow locking
1 ze
2 60m do:f2
3 29m
4 30m
5 20u
6 20u do:f2 ;

  20u p11:f1 p12:f2 ;
  d1 ;
  50u UNBLKGRAD ;Prevent lock perturbation
  (p2 ph20):f2
  gtime1 gron1 ;g_a kill steady-state 13C
  grec groff
;Proton spin echo
  (p1 ph20):f1 ; -> Hy
  d11
  (p1*2 ph20):f1
  d12
;Begin HMQC
  gtime2 gron2 ;pi clean
  grec groff
  taua
  (p2 ph1):f2
  d0
  (p1*2 ph2):f1
  d0
  (p2 ph20):f2
  gtime2 gron2 ;pi clean
  grec groff
  400u BLKGRAD ;allow locking
  taub
  20u p112:f2 ;
  go=2 ph31 cpds2:f2 ;Decouple
  45m do:f2 wr #0 if #0 zd

#ifdef INTERLEAVE
  3.5m rd11
  3.5m rd12
  15m ip1
  lo to 4 times 2
  15m id0
  15m ip31*2 ; State-TPPI
  lo to 5 times l1 ;
  30m BLKGRAD ; allow locking
exit

ph1=0 2

```

```
ph2=0 0 1 1 2 2 3 3
;--- Receiver Phase ---
ph31=0 2 2 0
;----Constant Phase ---
ph20=0
ph21=1
ph22=2
ph23=3
```

### 3. $^1\text{H}-\Gamma_2$ PRE for $^{15}\text{N}$ -attached protons

```

;file name - trosypepwg.yzksgr
;This is a working sequence for 1H-T2(PRE)TROSY-HSQC
;Reference:
;JMR 136, 92-101 (1999) : pepwg
;Iwahara 2007 JMR : PRE
;trosy pep expt Fig. 1b <--- Different from the previous seq by Yiping(1a)
;phases work for Brukers and axial peak is removed via STATES-TPPI
;uses the 3-9-19 watergate

;2D 15N TROSY HSQC for 15N labeled protein
;first coded by Yiping on 6/25/03
;converted to single FID collection by GR July 2013

;THIS IS AN INTERLEAVED VERSION:
; delays :d8, d8+(in4*4), d8+(in4*4*2), d8+(in4*4*3) , ....
; FOR L8=2, use "split2"
; L8=4, USE "split4"

#include <Avance.incl>
#include <Grad_sys.incl>

#define INTERLEAVE

;CHANNEL DEFINITIONS
; 600a
;H f1 o1p 4.697ppm proton observe frequency
; p1 10.26u 90 H at p11 on f1
; p11 -3dB H pwr
;N f2 o2p 116 ppm nitrogen observe frequency
; p2 37u 90 N at p12 on f2
; p12 -3dB pwr for N 90
;
;Set nd0 = 2
;
;TIME DOMAIN INFORMATION
;t1-Nitrogen evolution
; d0 = initial delay
; in0 = dwell time
; nd0 = 2
; l0 = number of complex points
define delay grecl
"grecl=90u"
"d0 = 0.5*in0 - (p2*0.6366 + p1*0.6366 +2*grecl)*0.5" ;
;"d0 = 0.5*in0 - p2*0.6366 - p1*0.6366"
"l0=td1/2"
;GRADIENTS
; gpx gpy gpz
"gttime1=400u" ;0 0 10
;"gttime2=320u" ;0 0 8
"gttime3=400u" ;0 0 10
"gttime4=400u" ;0 0 24
;"gttime5=320u" ;0 0 -8
"gttime6=300u" ;0 0 15
"gttime7=300u" ;0 0 10

"grec=150u"

"iset = 10u"

define delay tau1
define delay tau2
define delay tau3

"d8 = 2212u"

```

```

"d4 = d8*0.25-gtime1-grec" ; d8 = initial delay T for PRE
"d5 = d8*0.25-gtime7-grec"

"d2 = 2.7m" ;1/(4JNH)
"d18 = 110u" ;watergate ref time
"d19 = d18 - p2"
"tau1 = d2 - gtime1 - grec - p1 * 0.6366 - p2"
"tau2 = d2 - gtime3 - grec - p1 * 0.6366 - p2"
"tau3 = d2 - gtime4 - grec - p1 * 2.385 - d18 * 5"
;"d21 = (p2 - p1) " ;pulse center delay
"in4=1.5m" ; total delay = d8+(in4*4)
"in5=in4"
"l1 = td1/(2*18)"
"l20=0" ;TROSY SWITCH
;-----
1 ze
2 90m
20 10m
   iset
3 iset LOCKH_OFF
4 iset
   d1 p11:f1
   iset p12:f2
   iset LOCKH_ON
   (p2 ph20):f2 ; nitrogen purge
   gtime6 gron6
   grec groff
   (p1 ph20):f1 ; first proton pulse
   gtime1 gron1
   grec groff
   d4
   (p2*2 ph20):f2 ; N180
   gtime7 gron7
   grec groff
   d5
   tau1
   (p1*2 ph4):f1
   gtime7 gron7
   grec groff
   d5
   (p2*2 ph20):f2
   d4
   tau1
   gtime1 gron1
   grec groff
;----- TROSY SWITCH_-----
if "l20 %2 ==1"
{
  (center (p1 ph23):f1 (p2 ph1):f2) ;ph1 = 3120
}
else
{
  (center (p1 ph23):f1 (p2 ph11):f2) ;ph11= 1320
}

   iset gron2
   ;grec groff
   ;(p2 ph1):f2
   d0
   grecl groff
   iset gron5
   d0
   grecl groff
   (p1 ph2):f1
   gtime3 gron3
   grec groff

```

```
tau2
(center (p1*2 ph20):f1 (p2*2 ph20):f2)
tau2
gtime3 gron3
grec groff
(p1 ph20):f1 (p2 ph5):f2
gtime4 gron4
grec groff
tau3
(p1*0.231 ph23):f1
d18*2
(p1*0.692 ph23):f1
d18*2
(p1*1.462 ph23):f1
d19
(p2*2 ph20):f2
d19
(p1*1.462 ph21):f1
d18*2
(p1*0.692 ph21):f1
d18*2
(p1*0.231 ph21):f1
tau3
gtime4 gron4
grec groff
(p2 ph3):f2
go=2 ph31
75m wr #0 if #0 zd

iset iu20 ;TROSY switch
iset ip2*2
iset ip3*2
lo to 4 times 2

#ifdef INTERLEAVE
7m id4
8m id5
lo to 20 times l8
#endif
7m rd4
8m rd5

iset id0
iset ip1*2 ;States-TTPI
iset ip11*2 ;States-TTPI
iset ip31*2
lo to 3 times l1
iset LOCKH_OFF
exit

ph1 = 3 1 2 0 ;N TROSY switch
ph11= 1 3 2 0 ;N TROSY switch
ph3 = 0 ;N
ph13= 2 ;N
ph5 = 3 ;N
ph2 = 3 ;H
ph12= 1 ;H
ph4 = 3 ;H
ph31= 2 0 3 1 ;R
;Constant Phases
ph20 = 0
ph21 = 1
ph22 = 2
ph23 = 3
```

## 2. Scripts for refining nitroxide position and back-calculating PREs

Scripts for back-calculating PREs using XPLOR-NIH were obtained from Dr. Marius Clore (<http://spin.niddk.nih.gov/clore/Software/software.html>). Scripts used for refining the nitroxide position and back-calculating PREs from S2C site are listed here:

1. **Adding nitroxide group to the site of labeling in the PDB.** Input file: pdb file “\*\_start.pdb” where the residue type for the site of labeling is changed to “CYSP” for each atom. Command: “xplor -py buildtag.py”

```
#buildtag.py
from pdbTool import *
import protocol
import regularize
protocol.loadPDB('s2c_1B95_start.pdb')

notKnown=AtomSel("not known")
toFix=AtomSel("resn CYSP")

try:
    regularize.addUnknownAtoms_fast(verbose=1,maxFixupIters=50)
except protocol.CovalentViolation:
    pass

try:
    regularize.fixupCovalentGeom(sel=toFix,useVDW=1,verbose=1)
except protocol.CovalentViolation:
    pass

xplor.command("write psf output=s2c_1B95_tagged.psf end")
PDBTool("s2c_1B95_tagged.pdb").write()
```

2. **Generating a 3 conformer representation of the nitroxide group.** Input file: pdb file “\*\_tagged.pdb” that is generated after running the buildtag.py script. Command: “xplor make3conf.inp”

```
#make3conf.inp

structure @s2c_1B95_tagged.psf
end

set echo=on message=on end

coor @s2c_1B95_tagged.pdb

vector do (segid "ZP00") (resn CYSP)

for $an in (1 2) loop alts

evaluate ($al = "ZP0" + encode($an))

duplicate
  segid = $al
  select = ( segid ZP00 )
end

end loop alts

write psf output=s2c_1B95_ready.psf end
write coor output=s2c_1B95_ready.pdb end

stop
```

3. **Fitting nitroxide positions to PRE data.** Input file: “s2c\_refine.tbl”. Command: “xplor -py fit.py > fit.out”. Output: in ./fit/ folder .pdb and .viols files will be generated.

```
#fit.py

import protocol
protocol.initStruct('s2c_1B95_ready_3conf.psf')
protocol.initParams('protein')
protocol.initCoords('s2c_1B95_ready_3conf.pdb')

outFilename = "./step1/fit/SCRIPT_STRUCTURE.pdb" # ouput directory
numberOfStructures=10
startingStructure=0

from xplorPot import XplorPot

#can use this for analyzing terms
from simulationTools import *
#print analyze(XplorPot('IMPR'))

from avePot import AvePot

from potList import PotList
potList = PotList()

from simulationTools import MultRamp, StaticRamp, InitialParams

rampedParams=[]
highTempParams=[]

protocol.initNBond()

# Uncomment these to use RAMA potential

#protocol.initRamaDatabase('protein')
#potList.append( AvePot(XplorPot,"RAMA") )

# PRE restraints

from prePotTools import create_PREPot

# Setup of backbone 1H PRE restraints
pre = create_PREPot("PRE","s2c_refine.tbl","normal") #input file

def runSBmode():
    print 'configuring SB mode'
    from prePotTools import setupSBmode
    setupSBmode(pre)
    pre.setTcType("fix")
    pre.setEquType("sb")
    pre.setTauC(32.0)
    pre.setRlxType("r2dd")
    pre.setSqn( 0.5 )
    pre.setGfac( 2 )
    pre.setFreqI( 700.133 )
    #pre.setFunType("square")
    pre.setSclType("const")
    pre.calcEnergy()
    pass
    return

def runSBMFmode():
    print 'configuring SBMF mode'
    from prePotTools import setupSBMFmode
```

```

    setupSBMFmode(pre)
    pre.setTcType("fix")
    pre.setTiType("fix")
    pre.setTtType("fix")
    pre.setEquType("sbmf")
    pre.setSbmfType("taut")
    pre.setTauC(32.0)
    pre.setTauT(0.000)
    pre.setTauI(0.000)
    pre.setRlxType("r2dd")
    pre.setSqn( 0.5 )
    pre.setGfac( 2 )
    pre.setFreqI( 700.133 )
    #pre.setFunType("square")
    pre.setSclType("const")
    pre.calcEnergy()
    pass
    return

potList.append(pre)
rampedParams.append( MultRamp(0.01,1, "pre.setScale( VALUE )" ) )
rampedParams.append( StaticRamp("runSBMFmode()" ) )
#highTempParams.append( StaticRamp("runSBmode()" ) )

potList.append( AvePot(XplorPot,"VDW" ) )
rampedParams.append( StaticRamp("protocol.initNBond()" ) )
rampedParams.append( MultRamp(0.9,0.8,
    "xplor.command('param nbonds repel VALUE end end' )" ) )
rampedParams.append( MultRamp(.004,4,
    "xplor.command('param nbonds rcon VALUE end end' )" ) )
# nonbonded interaction only between CA atoms
highTempParams.append( StaticRamp("protocol.initNBond(cutnb=100,
    rcon=0.004,
    tolerance=45,
    repel=1.2,
    onlyCA=1)" ) )

potList.append( AvePot(XplorPot,"BOND" ) )
potList.append( AvePot(XplorPot,"ANGL" ) )
potList['ANGL'].setThreshold( 5 )
rampedParams.append( MultRamp(0.4,1,"potList['ANGL'].setScale(VALUE)" ) )
potList.append( AvePot(XplorPot,"IMPR" ) )
potList['IMPR'].setThreshold( 5 )
rampedParams.append( MultRamp(0.1,1,"potList['IMPR'].setScale(VALUE)" ) )

# Uncomment to use RAMA potential

#rampedParams.append( MultRamp(0.01,1,"potList['RAMA'].setScale(VALUE)" ) )

#set mass and fric

from atomAction import SetProperty

protocol.massSetup()

# IVM setup

from ivm import IVM
dyn = IVM()

# initialize ivm topology for torsion-angle dynamics

dyn.fix(' (name N or name HN or name CA or name HA or name C or name O) or not (resn CYSB or resn ILE or resn LEU

protocol.torsionTopology(dyn)

```

```

#interaction potential for residues that are allowed to move
def setConstraints():
    xplor.command("""
        constraints
        ! interaction (all) (all)
        inter = (segid "A") (resn ILE or resn LEU or resn VAL or resid 3:10 or resid 55 or resid 58 or resid 59)
        inter = (segid "A") (resn CYS)
        inter = (segid "ZP00") (segid "ZP00")
        inter = (segid "ZP01") (segid "ZP01")
        inter = (segid "ZP02") (segid "ZP02")
        inter = (segid "ZP03") (segid "ZP03")
        inter = (segid "ZP04") (segid "ZP04")
        end
    """)
    return

rampedParams.append( StaticRamp("setConstraints()") )
highTempParams.append( StaticRamp("setConstraints()") )

# object which performs simulated annealing
#
from simulationTools import AnnealIVM
init_t = 3000. # Need high temp and slow annealing to converge
cool = AnnealIVM(initTemp =init_t,
                 finalTemp=25,
                 tempStep =12.5,
                 ivm=dyn,
                 rampedParams = rampedParams)

def calcOneStructure(loopInfo):
    """ this function calculates a single structure, performs analysis on the
    structure, and then writes out a pdb file, with remarks.
    """

    #read in starting structure - for analysis
    #prevFilename='refine_%d.pdb' % (loopInfo.structNum)
    #protocol.initCoords(prevFilename)
    #FinalParams( rampedParams )
    #potList.calcEnergy()

    # initialize parameters for high temp dynamics.
    InitialParams( rampedParams )
    # high-temp dynamics setup - only need to specify parameters which
    # differ from initial values in rampedParams
    InitialParams( highTempParams )

    # high temp dynamics
    #
    protocol.initDynamics(dyn,
                         potList=potList, # potential terms to use
                         bathTemp=init_t,
                         initVelocities=1,
                         finalTime=10, # stops at 10ps or 5000 steps
                         numSteps=5000, # whichever comes first
                         numSteps=1, # for testing
                         printInterval=100)

    dyn.setETolerance( init_t/100 ) #used to det. stepsize. default: t/1000
    dyn.run()

    # initialize parameters for cooling loop
    InitialParams( rampedParams )

    # initialize integrator for simulated annealing
    #

```

---

```
protocol.initDynamics(dyn,
                      potList=potList,
                      numSteps=2000, #at each temp: 100 steps or
                      finalTime=.2 , # .2ps, whichever is less
                      printInterval=100)

# perform simulated annealing
#
cool.run()

# final torsion angle minimization
#
protocol.initMinimize(dyn,
                     printInterval=50)
dyn.run()

#do analysis and write structure

enePre = pre.calcEnergy()
qPre = pre.Qfactor()
rPre = pre.rms()
tcPre = pre.tc() * 1.0e9
print pre.info()
print pre.showRestrains(0)

loopInfo.writeStructure(potList)

pass

from simulationTools import StructureLoop, FinalParams
StructureLoop(numStructures=numberOfStructures,
              startStructure=startingStructure,
              pdbTemplate=outFilename,
              structLoopAction=calcOneStructure,
              ).run()
```

```
!s2c_refine.tbl

! HN & CH3 1H-PRE
!
! [700.133MHz]
!
!S2C

#NH & CH3 - removed resi with PRE less than 5 (outside PRE range of nitroxide, 25A)

#format - assign (paramagnetic atom) (reporter atom) PRE error

#amide

assign (resid 2 and name NS1) (resid 63 and name HN) 26.9 4.4
assign (resid 2 and name NS1) (resid 102 and name HN) 15.8 3.5
assign (resid 2 and name NS1) (resid 163 and name HN) 8.3 8
assign (resid 2 and name NS1) (resid 177 and name HN) 32 10
assign (resid 2 and name NS1) (resid 180 and name HN) 7.0 5
assign (resid 2 and name NS1) (resid 181 and name HN) 16 4
assign (resid 2 and name NS1) (resid 184 and name HN) 12 7
assign (resid 2 and name NS1) (resid 193 and name HN) 6 2
assign (resid 2 and name NS1) (resid 194 and name HN) 25 10
assign (resid 2 and name NS1) (resid 195 and name HN) 7 4
assign (resid 2 and name NS1) (resid 199 and name HN) 23 8
assign (resid 2 and name NS1) (resid 202 and name HN) 41 12
assign (resid 2 and name NS1) (resid 204 and name HN) 18 6
assign (resid 2 and name NS1) (resid 208 and name HN) 8.0 4
assign (resid 2 and name NS1) (resid 210 and name HN) 42 6
assign (resid 2 and name NS1) (resid 211 and name HN) 45.3 5
assign (resid 2 and name NS1) (resid 217 and name HN) 25 10
assign (resid 2 and name NS1) (resid 219 and name HN) 7.8 4.1
assign (resid 2 and name NS1) (resid 220 and name HN) 6.0 4.9
assign (resid 2 and name NS1) (resid 221 and name HN) 0.3 3.7
assign (resid 2 and name NS1) (resid 225 and name HN) 28 6
assign (resid 2 and name NS1) (resid 226 and name HN) 22 5
assign (resid 2 and name NS1) (resid 229 and name HN) 11.1 8
assign (resid 2 and name NS1) (resid 230 and name HN) 14 5
assign (resid 2 and name NS1) (resid 231 and name HN) 12 14
assign (resid 2 and name NS1) (resid 232 and name HN) 12.2 9
assign (resid 2 and name NS1) (resid 242 and name HN) 17.0 6
assign (resid 2 and name NS1) (resid 243 and name HN) 19 5
assign (resid 2 and name NS1) (resid 245 and name HN) 11.5 1.8

#methyl

assign (resid 2 and name NS1) (resid 103 and name HD#) 7.6 2.7
assign (resid 2 and name NS1) (resid 189 and name HD#) 1.7 3.3
assign (resid 2 and name NS1) (resid 192 and name HD#) 10.5 2
assign (resid 2 and name NS1) (resid 240 and name HD#) 4.9 3
assign (resid 2 and name NS1) (resid 62 and name HD#) 14.4 2.1
assign (resid 2 and name NS1) (resid 107 and name HD1#) 6.8 2.8
assign (resid 2 and name NS1) (resid 107 and name HD2#) 9.9 2.4
assign (resid 2 and name NS1) (resid 180 and name HD2#) 5.9 1.7
assign (resid 2 and name NS1) (resid 213 and name HD1#) 9.1 2
assign (resid 2 and name NS1) (resid 137 and name HG2#) 3.8 1.1
assign (resid 2 and name NS1) (resid 166 and name HG2#) 5.4 1.6
assign (resid 2 and name NS1) (resid 200 and name HG2#) 13.5 1.4
```

4. **Calculating PREs for residues not included in the refinement.** Input file: “s2c\_zeros.tbl” (format is same as s2c\_refine.tbl). Command: “xplor -py fit.py > fit.out”. Output: in ./calc/ folder .pdb and .viols files will be generated.

```
#calc.py

import protocol
protocol.initStruct('s2c_1B95_ready_3conf.psf')
protocol.initParams('protein')

outFilename = "./step200/calc/calc_2_amide.pdb"
numberOfStructures=1
startingStructure=0

from xplorPot import XplorPot

#can use this for analyzing terms
from simulationTools import *
#print analyze(XplorPot('IMPR'))

from avePot import AvePot

from potList import PotList
potList = PotList()

from simulationTools import MultRamp, StaticRamp, InitialParams

rampedParams=[]
highTempParams=[]

protocol.initNBond()

# Uncomment these to use RAMA potential

#protocol.initRamaDatabase('protein')
#potList.append( AvePot(XplorPot,"RAMA") )

# PRE restraints

from prePotTools import create_PREPot

# Setup of backbone 1H PRE restraints
pre = create_PREPot("PRE","s2c_nometal_calc_amide_step200.tbl","normal")

def runSBmode():
    print 'configuring SB mode'
    from prePotTools import setupSBmode
    setupSBmode(pre)
    pre.setTcType("fix")
    pre.setEquType("sb")
    pre.setTauC(32.0)
    pre.setRlxType("r2dd")
    pre.setSqn( 0.5 )
    pre.setGfac( 2 )
    pre.setFreqI( 700.133 )
    #pre.setFunType("square")
    pre.setScIType("const")
    pre.calcEnergy()
    pass
    return

def runSBMFmode():
    print 'configuring SBMF mode'
    from prePotTools import setupSBMFmode
    setupSBMFmode(pre)
```

```

pre.setTcType("fix")
pre.setTiType("fix")
pre.setTtType("fix")
pre.setEquType("sbmf")
pre.setSbmfType("taut")
pre.setTauC(32.0)
pre.setTauT(0.000)
pre.setTauI(0.000)
pre.setRlxType("r2dd")
pre.setSq( 0.5 )
pre.setGfac( 2 )
pre.setFreqI( 700.133 )
#pre.setFunType("square")
pre.setScIType("const")
pre.calcEnergy()
pass
return

potList.append(pre)
rampedParams.append( MultRamp(0.01,1, "pre.setScale( VALUE )" ) )
rampedParams.append( StaticRamp("runSBMFmode()" ) )
#highTempParams.append( StaticRamp("runSBmode()" ) )

potList.append( AvePot(XplorPot,"VDW" ) )
rampedParams.append( StaticRamp("protocol.initNBond()" ) )
rampedParams.append( MultRamp(0.9,0.8,
    "xplor.command('param nbonds repel VALUE end end' )" ) )
rampedParams.append( MultRamp(.004,4,
    "xplor.command('param nbonds rcon VALUE end end' )" ) )
# nonbonded interaction only between CA atoms
highTempParams.append( StaticRamp("protocol.initNBond(cutnb=100,
    rcon=0.004,
    tolerance=45,
    repel=1.2,
    onlyCA=1)" ) )

potList.append( AvePot(XplorPot,"BOND" ) )
potList.append( AvePot(XplorPot,"ANGL" ) )
potList['ANGL'].setThreshold( 5 )
rampedParams.append( MultRamp(0.4,1,"potList['ANGL'].setScale(VALUE)" ) )
potList.append( AvePot(XplorPot,"IMPR" ) )
potList['IMPR'].setThreshold( 5 )
rampedParams.append( MultRamp(0.1,1,"potList['IMPR'].setScale(VALUE)" ) )

# Uncomment to use RAMA potential

#rampedParams.append( MultRamp(0.01,1,"potList['RAMA'].setScale(VALUE)" ) )

#set mass and fric

from atomAction import SetProperty

protocol.massSetup()

# IVM setup

from ivm import IVM
dyn = IVM()

# initialize ivm topology for torsion-angle dynamics

#dyn.fix(' (name N or name HN or name CA or name HA or name C or name O) or not (resn CYSP)' )
dyn.fix(' (name N or name HN or name CA or name HA or name C or name O) or not (resn CYSP or resn ILE or resn LEU

protocol.torsionTopology(dyn)

```

---

```

def setConstraints():
    xplor.command("""
        constraints
        ! interaction (all) (all)
            inter = (segid " ") (segid "A")
            inter = (segid "A") (resn ILE)
            inter = (segid "A") (resn VAL)
            inter = (segid "A") (resn LEU)
            inter = (segid "A") (resn MET)
            inter = (segid "ZP00") (segid "ZP00")
            inter = (segid "ZP01") (segid "ZP01")
            inter = (segid "ZP02") (segid "ZP02")
            inter = (segid "ZP03") (segid "ZP03")
            inter = (segid "ZP04") (segid "ZP04")
        end
    """)
    return
rampedParams.append( StaticRamp("setConstraints()") )
highTempParams.append( StaticRamp("setConstraints()") )

# object which performs simulated annealing
#
from simulationTools import AnnealIVM
init_t = 3000. # Need high temp and slow annealing to converge
cool = AnnealIVM(initTemp =init_t,
                 finalTemp=25,
                 tempStep =12.5,
                 ivm=dyn,
                 rampedParams = rampedParams)

def calcOneStructure(loopInfo):
    """ this function calculates a single structure, performs analysis on the
        structure, and then writes out a pdb file, with remarks.
    """

    #read in starting structure - for analysis
    prevFilename='./step200/fit/fit_2.pdb' #% (loopInfo.structNum)
    protocol.initCoords(prevFilename)
    FinalParams( rampedParams )
    potList.calcEnergy()

    #do analysis and write structure

    enePre = pre.calcEnergy()
    qPre = pre.Qfactor()
    rPre = pre.rms()
    tcPre = pre.tc() * 1.0e9
    print pre.info()
    print pre.showRestrains(0)

    loopInfo.writeStructure(potList)

    pass

from simulationTools import StructureLoop, FinalParams
StructureLoop(numStructures=numberOfStructures,
              startStructure=startingStructure,
              pdbTemplate=outFilename,
              structLoopAction=calcOneStructure,
              ).run()

```

5. **Grab parameters (Energy, Violations and Q-factors).** Input files: ./fit/.pdb or ./calc/.pdb files generated from fit.py or calc.py. Command: “./Grab\_par.awk > step1\_par”

```
#Grab_par.awk

BEGIN{

    fileroot = "./step1/fit/fit"
    numfiles = 10

    Qfactor = 0
    Energy = 0
    Viol = 0

    for(i = 0; i < numfiles; i++){
        filename[i]= fileroot+"_i".pdb"
        printf "\n\n"
        printf filename[i]

        while(getline < filename[i] > 0){

            if($2=="summary" && $3=="total"){
                Energy = $4
                printf " Energy:\t"
                printf Energy
                printf "\t"
                Viol = $5
                printf " Violations:\t"
                printf Viol
                printf "\t"
            }
            if($2=="PRE" && $3=="PRE"){
                Qfactor = $5
                printf " Qfac: "
                printf Qfactor
                printf "\t"
            }
        }close(filename[i])
    }
}
```

6. **Grab PREs amide.** Input files: ./fit/.pdb or ./calc/.pdb files generated from fit.py or calc.py. Command: “./Grab\_PRE\_amide.awk > step1\_pre\_amide”

```
#Grab_PRE_amide.awk

BEGIN{

    fileroot = "./step1/fit/fit"
    numfiles = 10

    AA = 0
    Resid = 0
    PRE = 0

    for(i = 0; i < numfiles; i++){
        filename[i]= fileroot+"_i".pdb.viols"
        printf "\n\n"
        printf filename[i]
        printf "\n"

        while(getline < filename[i] > 0){
            if($3=="(ZP00" && $5=="CYSP"){
                AA = $9
                Resid = $8
            }
            if($2=="(ZP00" && $4=="CYSP"){
                AA = $8
                Resid = $7
            }
            if($1=="(ZP02" && $3=="CYSP"){
                PRE = $6
                print (AA,Resid "\t" PRE)
            }
        }

        }close(filename[i])
    }
}
```

7. **Grab PREs methyl.** Input files: ./fit/.pdb or ./calc/.pdb files generated from fit.py or calc.py. Command: “./Grab\_PRE\_methyl.awk > step1\_pre\_methyl”

```
#Grab_PRE_methyl.awk

BEGIN{

    fileroot = "./step1/fit/fit"
    numfiles = 10

    AA = 0
    Resid = 0
    PRE = 0

    for(i = 0; i < numfiles; i++){
        filename[i]= fileroot_"i".pdb.viols"
        printf "\n\n"
        printf filename[i]
        printf "\n"

        while(getline < filename[i] > 0){

            if($1=="(ZP02" && $3=="CYSP"){
                AA = $7
                #printf " Energy:\t"
                #printf AA
                #printf "\t"
                Resid = $6
                #printf " Violations:\t"
                #printf Resid
                #printf "\t"
                PRE = $10
                #printf " Violations:\t"
                print (AA,Resid "\t" PRE)
                #printf "\t"
            }
        }close(filename[i])
    }
}
```