STRUCTURAL AND FUNCTIONAL STUDIES OF DNA NUCLEASES: SGRAI AND MK0566

By

Santosh Shah

_____

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

WITH A MAJOR IN BIOCHEMISTRY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2013

THE UNIVERSITY OF ARIZONA

GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Santosh Shah entitled “Structural and Functional Studies of DNA Nucleases: SgrAI and Mk0566” and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

_____Date: December 12, 2013

Dr. Nancy C. Horton

_____Date: December 12, 2013

Dr. William R. Montfort

_____Date: December 12, 2013

Dr. Matthew H. J. Cordes

_____Date: December 12, 2013

Dr. Megan M. McEvoy

_____Date: December 12, 2013

Dr. Indraneel Ghosh

Final approval and acceptance of this dissertation is contingent upon the candidate’s submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

_____Date: December 12, 2013

Dissertation Director: Dr. Nancy C. Horton

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that an accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: SANTOSH SHAH

ACKNOWLEDGEMENTS

First of all, I would like to thank my wonderful, hardworking and inspirational mentor **Dr. Nancy C. Horton** with great respect and regards. She has always been very supportive and encouraging throughout my graduate career. Her hardwork and dedication keeps me motivated and encourages me to work more and perform better to achieve the potential that she sees in me. I would like to thank her for spending a great amount of time teaching me all the laboratory techniques, as well as presentation and writing skills. I really appreciate all the moral support she has provided so far.

I would also like to thank my great committee members who have extended their support throughout my graduate career. I would like to thank **Dr. William R. Montfort** for always willing to talk about crystallography and the moral support he has provided over my stay here. I would like to thank **Dr. Matthew H. J. Cordes** and his lab members for allowing me to use his instruments and borrow chemicals when we were in short supply, and being a great neighbor. I would like to thank **Dr. Megan M. McEvoy** for providing me with the plasmids and competent cells, and making my day with her cheerful personality every time I come across her. I would like to thank **Dr. Indraneel Ghosh** for giving me an opportunity to learn cloning and for his moral support.

I would like to thank **Dr. Chad K. Park** from Analytical Biophysics Core Facility for always being there to troubleshoot any instrument or enrich me with tons of alternative ideas that I could try. I would also like to thank him for conducting, analyzing, and processing the analytical ultracentrifugation data. I would also like to thank **Dr. Sue Roberts** and **Dr. Andrzej Weichsel** from Macromolecular Crystallography Core Facility at the University of Arizona for helping me solve and refine the crystal structure of K96A SgrAI. I would like to thank our collaborators **Dr. Jurate Bitinaite** from New England Biolabs for providing us with purified WT SgrAI enzyme, and **Dr. Alka Agrawal** for constructing the K96A mutant and providing us with purified enzyme. I would also like to thank **Dr. Nancy C. Horton** for collecting, solving, and refining the crystal structure of SgrAI WT bound to 22-1HT, and construction of figures. I would also like to thank **Dr. Pete Dunten** from Stanford Synchrotron Radiation Lightsource for helping us refine the crystal structure of Mk0566. I would like to thank **Amanda Stiteler** for the cleavage assays and purified Mk0566, and **Dr. Andrea Babic** for purified Mk0566.

I would like to thank **Jane Dugas, Olivia Mendoza** and **Lori Boyd** for all their help figuring out the administrative requirements and always being there to answer any questions that I had.

I would also like to thank my close friends **Marina Asiedu, Katie Holso, Varuni Livera, Niraj Pant, Joey Klebba, Rahul Purohit, Samuel Jayakanthan, Zachery Campbell, Alberto Roscon, Michael Pham, Christian Roessler, Swapna Arvind,** and **Jonathan Sanchez** for being there for me.

Finally, I would like to thank my family members from the bottom of my heart who have sacrificed all their life bringing me up, making me feel special and treating me like a prince: **Eng. Dukha Shah, Mrs. Dukhiya Devi, Mr. Ram Pratap Panjiyar, Mrs. Urmila Devi Panjiyar, Dr. Pooja Panjiyar, Rita Shah, Asha Shah, Rahul Mahato, Sudhir Mahato,** and **Barsha Mahato.**

DEDICATION

I would like to dedicate this dissertation to my parents Eng. Dukha Shah and Mrs. Dukhiya Devi, my lovely wife Dr. Pooja Panjiyar, my mentor Dr. Nancy C. Horton and to the scientific community

TABLE OF CONTENTS

TABLE OF CONTENTS – *Continued*

TABLE OF CONTENTS – *Continued*

TABLE OF CONTENTS – *Continued*

TABLE OF CONTENTS – *Continued*

LIST OF FIGURES

LIST OF FIGURES - *Continued*

LIST OF FIGURES - *Continued*

LIST OF TABLES

LIST OF ABBREVIATIONS

Ala	Alanine
ATP	Adenosine triphosphate
B-factor	Crystallographic temperature factor
β-Me	2-mercaptoethanol
BPB	Bromophenol blue
$Ca(OAc)_2$	Calcium acetate
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
ds	Double strand
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
FPLC	Fast protein liquid chromatography
GTP	Guanosine triphosphate
HPLC	High performance liquid chromatography
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
KOAc	Potassium acetate
LB	Luria-Bertani
Lys	Lysine
MBP	Maltose binding protein
$MgCl_2$	Magnesium chloride
MWM	Molecular weight marker
NaCl	Sodium Chloride
NER	Nucleotide excision repair
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein Data Bank
PEG	Polyethylene glycol
PMSF	Phenylmethylsulfonyl fluoride
PNK	Polynucleotide kinase
RE	Restriction endonuclease
RMSD	Root mean square deviation
SAM	*S*-Adenosyl methionine
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
S_N2	Substitution nucleophilic bi-molecular
ss	Single strand
Tris-OAc	Tris acetate
vdW	van der Waals
XCFF	Xylene cyanol FF

ABSTRACT

DNA nucleases are essential for various biological functions such as replication, recombination, and repair. Restriction endonucleases (REs) are excellent model system for the investigation of DNA recognition and specificity. SgrAI is a type IIF RE that cuts an 8 base pair primary sequence. In addition to its primary cleavage activity it also cleaves secondary sequences, but only appreciably in the presence of the primary sequence. The longer flanking DNA exhibits much greater activated DNA cleavage by SgrAI (>1000 fold activation by secondary site). Interestingly, the asymmetric cleavage seen in one of the two types of secondary site DNA is lost upon activation of SgrAI, suggesting a loss of communication between DNA recognition and activity upon specificity expansion. The structure of SgrAI bound to 22-1HT supports the cryoelectron microscopy structure of activated, oligomeric SgrAI highlighting the significance of the contacts made by the flanking DNA and the role played by N-terminal domain contacts in forming the run-on oligomer. The biological study suggests that the run-on oligomer formation sequesters the host DNA from being cleaved by the activated SgrAI complex. The DNA sequence binding, cleavage preference, and the structure of K96A SgrAI were determined. Unexpectedly, this mutation did not alter the structure of the enzyme, nor did it result in an enzyme lacking sequence preference at the 7^{th} position. Instead, the largest effect of the mutation appears to be in making the enzyme more specific such that it fails to cleave either type of secondary site. It may be that the K96 side chain is required to distort the non YG sequences (specifically GG and TC) of secondary site DNA for proper positioning in the enzyme active site upon activation and specificity expansion. The

crystal structure of Mk0566, XPG homologue from *M. kandleri,* was solved to 2.48 Å resolution and was found to be very similar to that of human FEN-1 and to other archaeal FEN-1/XPG homologues. These results suggest that the main biological role of Mk0566 is in DNA replication; however, they do not preclude involvement in a modified form of nucleotide excision repair.

CHAPTER 1

INTRODUCTION

Restriction Endonucleases

DNA binding proteins are very essential in the functioning of living cells. Various biological functions such as replication, recombination, and repair are carried out by these DNA nucleases. These nucleases are very efficient and specific. Failure to perform specific tasks could lead to different diseases including cancer, rapid aging (Thoms *et al.* 2007), autoimmunity (Nagata 2007), and repeat expansion diseases (Fry *et al.* 1999). Restriction endonucleases (REs) are highly specific enzymes which cleave DNA at or near specific recognition sequence. They are part of restriction modification (R-M) systems that defend their host against foreign phage invasion (Arber *et al.* 1962, Pingoud *et al.* 2005). The host DNA is protected by methylation of N4 or C5 at cytosine or N6 at adenine. Initially, REs were studied by Werner Arber and Daisy Dussoix (Arber *et al.* 1962). The first sequence specific RE, HindII, was discovered in Hamilton Smith's lab at Johns Hopkins Medical School (Kelly *et al.* 1970). Shortly afterwards, EcoRI and EcoRII were discovered in Herb Boyer's lab at the University of California, San Francisco (Yoshimori *et al.* 1972). The Nobel Prize for Physiology or Medicine was awarded to Daniel Nathans, Werner Arber, and Hamilton O. Smith in 1978 for their pioneering work on R-M systems.

There are four types of restriction endonucleases (Figure 1.1). Type I REs have three different subunits, HsdM, HsdR and HsdS, that are responsible for modification,

restriction and sequence recognition, respectively. They require ATP, Mg^{2+} and *S*-Adenosyl methionine (SAM) for their activity. They cut the DNA distal to the recognition sites, about half-way between two sites (Murray 2000, Pingoud *et al.* 2005). Type II REs recognize palindromic sequences and cut DNA within the recognition sequence. They require Mg^{2+} as cofactor but they do not require ATP or GTP (Roberts *et al.* 1993). The core consists of a five-stranded mixed β-sheet flanked with α-helices (Venclovas *et al.* 1994, Kovall *et al.* 1999). Type III REs consists of two subunits only: modification and restriction. They require ATP and Mg^{2+} for activity and are stimulated by SAM. They cut DNA close to one of the recognition sites (Bourniquel *et al.* 2002). Similarly, Type IV REs recognize and cleave methylated DNA. They are not part of the restriction modification system. They have two different subunits for DNA recognition and cleavage. They require GTP and Mg^{2+} for DNA cleavage, and they cleave close to one of the two RmC sites (McCleland *et al.* 2004).

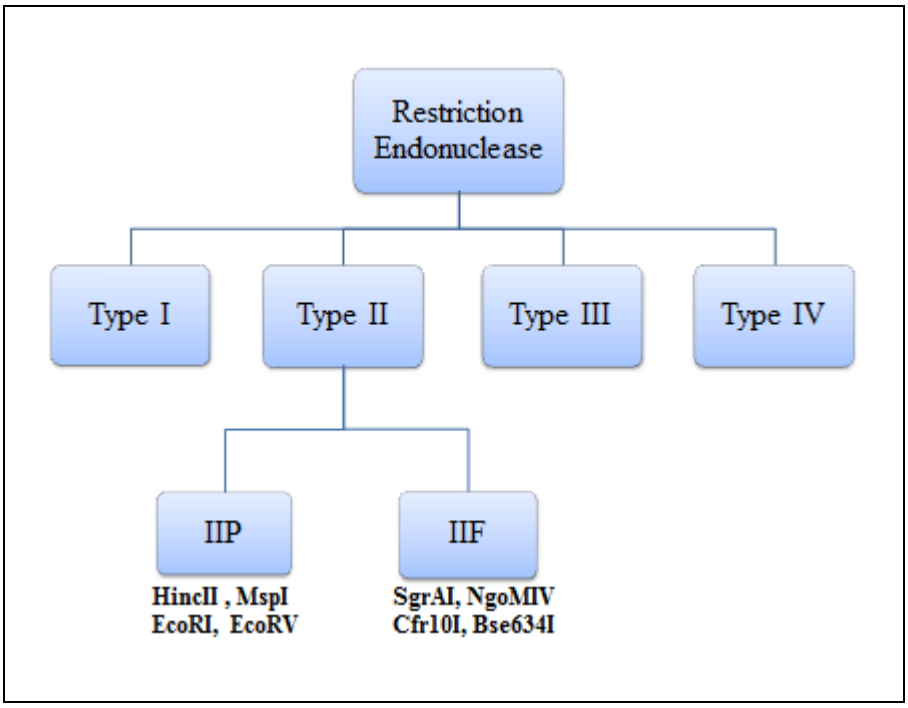


Figure 1.1: Schematic diagram showing types of restriction endonucleases. Only subtypes IIP and IIF, among many others, are shown for type II RE.

Type II REs have been a valuable tool for genetic manipulation, and a model system for studying specific DNA-protein interactions and mechanisms of phosphodiester hydrolysis (Kovall *et al.* 1999). Almost all the enzymes that cleave DNA, i.e. hydrolysis of phosphodiester bond, do so by coordination of divalent cations (Cowan 1998, Horton *et al.* 2001). Phosphodiester bond hydrolysis occurs by an S_N2-type mechanism with inversion of stereochemistry taking place at the phosphorous (Connolly *et al.* 1984). The catalytic center generally contains a PD…D/ExK motif (Anderson 1993). Two negatively charged amino acid side chains (usually aspartate or glutamate) and a lysine residue ligate divalent metal ion cofactors, usually Mg^{2+}, to carry out the catalysis. The number of divalent metal ions involved in cleavage of DNA in REs is still controversial; however, a two metal ion coordination mechanism is believed to be a

common mechanism based on crystal structures (Figure 1.2, Etzkorn *et al.* 2004a, Etzkorn *et al.* 2004b). One of the metal ions is involved in deprotonating a neighboring water molecule by lowering its pK_a, while both the metal ions are involved in stabilizing the doubly charged pentavalent transition state. The water molecule attacks the phosphorous atom thus, generating 3´-OH as the leaving group. The products of the reaction are DNA fragments with a 3´-OH and a 5´-phosphate (Beese *et al.* 1991).

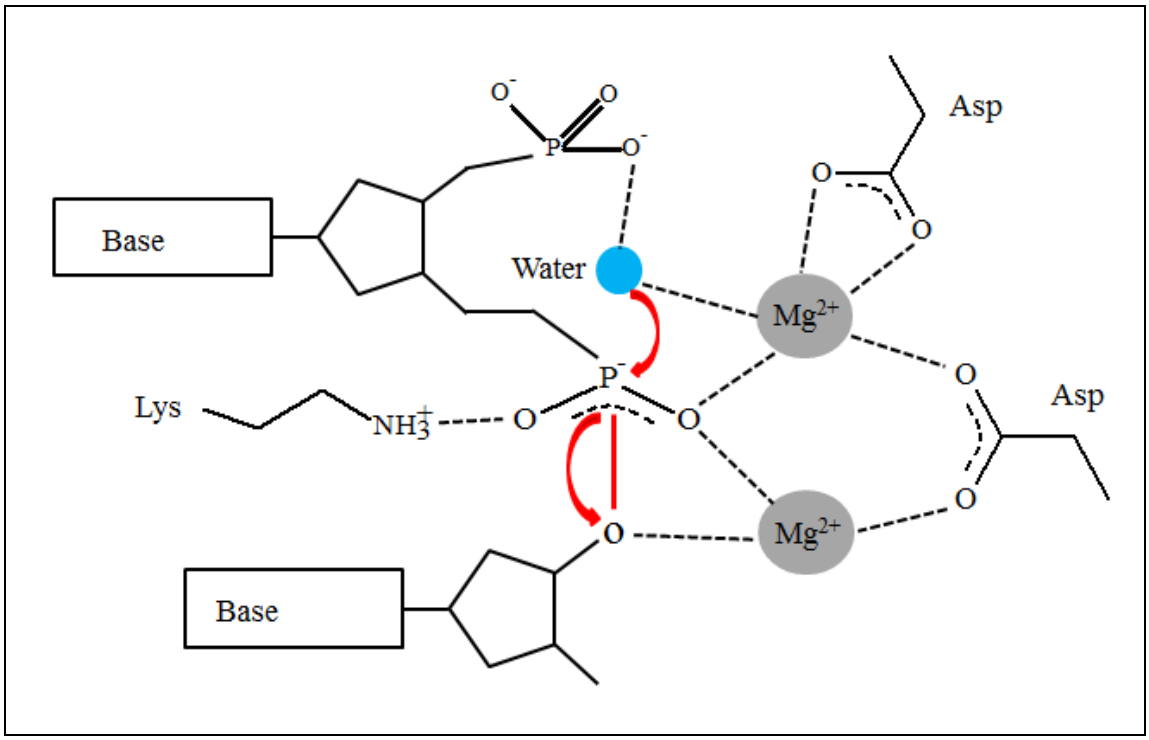


Figure 1.2: Two metal ion coordination mechanism. The divalent metal ions stabilize the transition state, which are ligated by conserved amino acids. Adapted from Etzkorn *et al.* (2004a), Etzkorn *et al.* (2004b).

Type IIF REs are typically homotetrameric which interact with two copies of their recognition site, and cleave both sites in a concerted manner (Welsh *et al.* 2004). Structure of type IIF restriction endonucleases (REnases) SfiI, NgoMIV, and Cfr10I are homotetrameric composed of two dimers in a tail-to-tail orientation (Siksnys *et al.* 1999,

Deibert *et al.* 2000, Vanamee *et al.* 2005). SgrAI is a type IIF restriction endonuclease which recognizes and cleaves an 8 base pair (bp) primary recognition sequence (1°), CR|CCGGYG (| indicates cut site, R denotes A or G, and Y denotes C or T), and is dimeric in solution (Tautz *et al.* 1990a, Tautz *et al.* 1990b, Roberts *et al.* 2001, Roberts *et al.* 2003). SgrAI shows unusual biochemical activity. It cleaves plasmids having two recognition sites faster than the ones with only one recognition site (Bilcock *et al.* 1999, Daniels *et al.* 2003). In addition to its primary cleavage activity it also cleaves secondary sequences CR|CCGGGG or CR|CCGGY(A/C/T) in the presence of the primary sequence (Bitinaite *et al.* 2002). Primary site DNA, either cleaved or not, with sufficient flanking DNA has been shown to be an allosteric effector. It stimulates cleavage of both primary and secondary site DNA sequences (Park *et al.* 2010b, Lyumkis *et al.* 2013). Upon activation, SgrAI forms a run-on oligomer by the addition of DNA-bound dimeric SgrAI subunits (Lyumkis *et al.* 2013).

Run-on oligomerization

Diseases such as Alzheimer's, type II diabetes and Huntington's are known as deposition diseases since they are linked to protein aggregation. In general, proteins can oligomerize by cross-β spine, end-to-end stacking, and/or 3D domain swapping mechanisms. 3D domain swapping could lead to formation of homodimers and higher order oligomers due to exchange of all or part of protein domains (Bennett *et al.* 2006). As shown in Figure 1.3, homodimerization occurs when domains of a pair of "closed monomers" swap to form a pair of "open monomers" by rotating about the hinge loop. The self-complementing domain-swapped homodimer is stabilized by the same

interdomain interactions that are found in the closed monomer. Usually, the major conformational changes occur only in the hinge loop during domain swapping. The exchange of domains between protomers allows formation of two "functional units" (FU). Each FU consists of residues from two protein chains. If the swapping occurs within two monomers leaving no exposed domain it forms a closed domain swapped homodimer. Alternatively, the domains can propagate leaving one unsatisfied domain at each terminus to form an open-ended runaway domain swapped oligomer (Bennett *et al.* 2006). Domain swapping generates a strong, flexible tether. The tight bonding of proteins makes these domain-swapped oligomers resistant to proteolysis. This might explain how these oligomers might gain resistance from cell's protein degradation machinery (Bennett *et al.* 2006). Domain-swapped filaments observed in crystal structures can exist *in vivo* as well. Fibrils of HbS, the filament observed in crystal structure, is believed to be a fundamental component of the fibers associated with sickle cell anemia (Wishner *et al.* 1975, Bennett *et al.* 2006). Similarly, the P6-P7-cleaved Pittsburgh variant of α_1-antitrypsin undergoes irreversible domain swapping within the crystals. Polymers of cleaved α_1-antitrypsin have also been characterized in solution by biochemical and ultrastructural techniques, demonstrating that similar filaments occur in solution (Mast *et al.* 1992, Bennett *et al.* 2006).

The domain swapping phenomenon was first proposed in 1960s to explain the behavior of RNase A dimer (Crestfield *et al.* 1962), tryptophan synthase (Jackson *et al.* 1969), and tryptophanase (London *et al.* 1974). The first reported crystal structure of domain swapped proteins were that of cro repressor (Anderson *et al.* 1981), beef liver

catalase (Fita *et al.* 1985), human CksHs2 (Parge *et al.* 1993), citrate synthase (Remington *et al.* 1982), and *E. coli* recA protein (Story *et al.* 1992). Unlike SgrAI, the oligomerization of these enzymes do not relate to activation of enzyme activity or changes in specificity. Acetyl-CoA carboxylase also forms a run-on oligomer but it does not possess other unusual phenomena such as product stabilization of the activated form, specificity expansion and domain swapping as SgrAI (Brownsey *et al.* 2006). Thus, activation of SgrAI through run-on oligomerization is a new paradigm in enzyme activity.

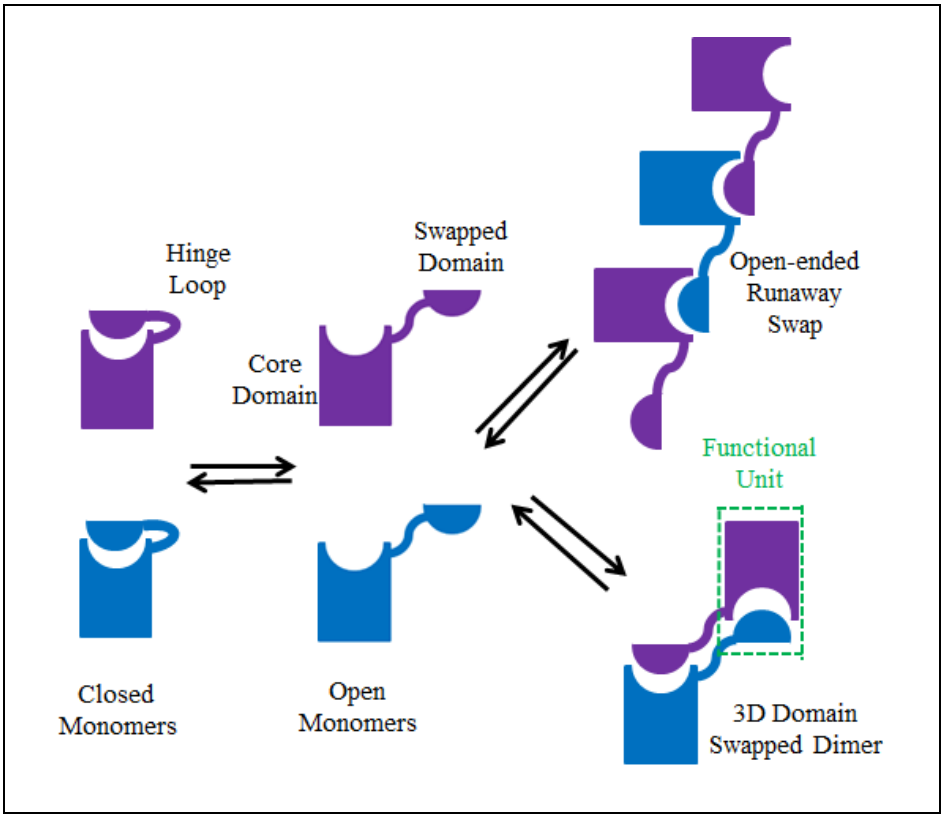


Figure 1.3: Schematic diagram showing 3D domain swapping that could lead to either closed ended domain swapped structure or open-ended runaway swapped structure. Adapted from Bennett *et al.* (2006).

In addition to domain swapping, run-on oligomerization can also be formed by enzyme binding to DNA. The oligomeric form can then be stabilized by the flanking DNA that might extend from one monomer to another. Park *et al.* 2010b and Lyumkis *et al.* 2013 show that the longer flanking DNA has a direct correlation with activation of the enzyme. SgrAI gets activated upon binding to primary site sequence with 7 or more bp flanking DNA on either side of the recognition sequence. The 8.6 Å cryo-EM structure of SgrAI bound to PC DNA (Lyumkis *et al.* 2013) shows that the flanking DNA plays a vital role in stabilizing the run-on oligomer structure. The negatively stained EM images showed that SgrAI formed a rod shaped structure of varying lengths with a regular repeating structure. The run-on oligomer was found to consist of 1 through 18 DNA bound SgrAI dimers (DBD) which is consistent with the analytical ultracentrifugation data (12 DBDs, Park *et al.* 2010b). The DBDs were observed to be arranged in a left handed helix with roughly 4 DBD per turn of the helix. The flanking DNA was seen to closely approach neighboring SgrAI especially with loops 56-60 and 122-140. In addition to these contacts, the SgrAI proteins contacted each other using their N-terminal domains near the central helical axis. At this resolution, the existence of domain swapping as seen in the previous crystal structure cannot be discerned, but the adjacent locations of the N-termini make it a possibility. Upon fitting the three dimensional structure of SgrAI (low activity conformation) into the cryo-EM map it showed that a 10º rotation of one subunit to the other occurred in the cryo-EM structure which we regard as high activity conformation. It is possible that the rotation of subunits might align the active site better to achieve the activation seen by the run-on oligomer (Lyumkis *et al.* 2013). Because of

the importance of the flanking DNA to the activated form, DNA cleavage measurements were done and presented in Chapter 2 using longer DNA, i.e. with 40-mer DNA which has longer flanking DNA than 18-mer (16 vs 5 bp flanks). A much greater degree of activated DNA cleavage than reported previously was found when secondary sites were embedded in these longer DNAs.

Streptomyces griseus and significance of run-on oligomerization

Streptomycetes are soil-dwelling, Gram-positive, filamentous bacteria which produce a wide variety of secondary metabolites, including antibiotics, parasiticides, herbicides, and pharmacologically active substances (such as antitumor agents and immunosuppressants) (Bentley *et al.* 2002, Ohnishi *et al.* 2008). The characteristic earthy smell of freshly plowed soil is thought to be that of the aromatic terpenoid geosmin produced by streptomycetes. There are currently 364 known species of streptomycetes. They produce a wide variety of secondary metabolites, such as antibiotics, parasiticides, and herbicides. They supply over two-thirds of all the naturally derived antibiotics used today e.g. streptomycin, tetracycline and chloramphenicol (Hodgson 2000, Kong *et al.* 2000, Bentley *et al.* 2002, Ohnishi *et al.* 2008). They also exhibit other useful biological properties such as modulation of the immune system, anti-tumor suppression and enzyme inhibition. They are also responsible for degrading insoluble, recalcitrant compounds of other organisms, such as lignocellulose and chitin, making them one of the key organisms in carbon recycling (Bentley *et al.* 2002). Streptomycetes are also known to oxidize sulfur, thiosulfate and tetrathionate to sulfate (Wainwright *et al.* 1984). Streptomycete glucose isomerase is used widely in the food industry for converting glucose to 2.5-fold

sweeter fructose (Hodgson *et al.* 2000). Unlike most bacteria, streptomycetes have complex multicellular development. (Bentley *et al.* 2002).

Streptomyces griseus produces streptomycin, an antituberculosis agent which is the first aminoglycoside antibiotics discovered in Waksman's laboratory (Waksman 1953). *S. griseus* has a complex multicellular development, and unlike other eubacterial chromosomes, its chromosome is linear consisting of 8,545,929 base pairs, one of the largest bacterial genomes discovered so far and almost twice the size of most bacterial genomes. It also has a very high G+C content i.e. 72.2% (Ohnishi *et al.* 2008). The genome contains a large proportion of regulatory genes mainly that are involved in responding to external stimuli and stresses. It has about 7,138 genes that could, in theory, be translated into proteins (Bentley *et al.* 2002, Ohnishi *et al.* 2008). The large genome might be essential for them to survive in the highly competitive soil environment, allowing them to have a more complex life cycle, exploit a greater variety of nutrient sources, adapt to multitude of stress conditions and transport variety of substrates such as sugars, amino acids, peptides, metals and ions in the cellular level (Bentley *et al.* 2002). The linear chromosomal DNA of *S. griseus* is highly unstable and frequently undergoes large rearrangements at the extremities. These rearrangements might play significant role in amplifying certain adjacent sequence and leading to metabolic diversity (Kong *et al.* 2000). *S. griseus* genome has a unique telomere sequence compared to other streptomycetes. The conserved terminal proteins (Tpgs) and telomere-associated proteins (Taps) are absent in the terminal inverted repeats (TIRs) and have been replaced by a novel pair of Tpg and Tap proteins. It is believed that the novel pair of Tpg and Tap

proteins might have been acquired from other linear plasmid over the course of evolution and the original *tpg* and *tap* genes became obsolete and decayed away eventually (Bentley *et al.* 2002). The large genome might hold clues to why SgrAI has evolved into such unusual enzyme. The large genome results in more recognition sites than in other bacterial species, which puts pressure on the methyltransferase activity to methylate the greater number of sites. However, since the methyltransferase activity may be limited by its cofactor SAM concentration, *S. griseus* may have evolved SgrAI to recognize a longer sequence (8 bp) and a slow basal DNA cleavage rate to allow the methyltransferase ample time to methylate host recognition sites before cleavage by SgrAI. However, the slow basal cleavage rate and rare recognition sequence may also limit effective phage protection. Thus, SgrAI has evolved activation via run-on oligomerization (providing 200-1000 fold rate enhancement) and sequence-specificity expansion from 3 primary sites to total 17 different primary and secondary sites. However, to prevent host DNA from being cleaved due to specificity expansion, the run-on oligomer sequesters the activated SgrAI away from the host DNA. Chapter 2 investigates the structure and biological role of the run-on oligomer formed by SgrAI.

Impact of His tag on Crystal Structure

His tags (Smith *et al.* 1988) have been really useful as a purification tool for recombinant proteins. The six consecutive histidines can be cloned into the protein of interest either on the N- or C-terminal. These His tags allow easier purification of the protein of interest by using a nickel-affinity column. His tag can be cleaved off after purification if a protease-cleavage site is introduced between the His tag and the protein

(Carson *et al.* 2007). About 90% of the crystal structures are based on recombinant methods and nearly 60% of them use some sort of His tag (Derewenda 2004). Many crystallographers believe these tags hinder crystallization, and thus, cleave them off before crystallizing their protein (Derewenda 2004, Waugh *et al.* 2005). However, some believe the tags have no effect on the structure or function of the protein (Chant *et al.* 2005), and yet some believe the tags might actually be helpful (Carson *et al.* 2007). There are very few cases of adverse effects of His tag on structure-function relationships. One case, (Klose *et al.* 2004), indicated that His tag influenced folding and actually changed the disulfide-bonding pattern. However, in another case (Tajika *et al.* 2004), it was reported that the His tag was required for crystallization to take place. The His tag that was on the C-terminal helix protruded outside the molecule and packed with another molecule about the crystallographic two-fold axis. Many His tagged crystal structures have been solved to date. Over 90% of the His tags are disordered, and are thus not seen in the crystal structures. Since His tags have little effect on protein structure, it may be possible to design crystallization tags to enhance structure determination (Carson *et al.* 2007).

Indirect readout of DNA sequence

DNA binding proteins carry out many essential functions in living cells. Many of these proteins that cleave duplex DNA are divalent cation-dependent and are involved in DNA repair, transposition, recombination, modification of DNA supercoiling, and bacterial immunity (Cowan 1998, Horton *et al.* 2002, Etzkorn *et al.* 2004a). Most of these enzymes recognize specific sequences in DNA using direct and indirect readout (Garvie

et al. 2001). Direct readout involves direct contacts between the protein side chain or main chain atoms and the DNA bases via hydrogen bonding and/or van der Waals interactions (Figure 1.4). These interactions usually occur in the major groove of the DNA and provide a high degree of sequence specificity (Steffen *et al.* 2002). Sequence specificity occurring in the absence of direct readout is considered to be evidence of indirect readout. Indirect readout is the ability of a protein to distinguish certain sequences over other, based solely on interactions other than those considered direct readout (Koudelka *et al.* 2006). Indirect readout involves contacts mediated by water (Otwinowski *et al.* 1998), other small molecules (Segal *et al.* 2006), to the sugar-phosphate backbone, and/or using distortion of the DNA to distinguish different DNA sequences (Rice *et al.* 1996, Koudelka *et al.* 2006, Blakaj *et al.* 2006, Little *et al.* 2008). Both direct and indirect readout can contribute to specificity, albeit with varying relative contributions (Kono *et al.* 1999, Sarai *et al.* 2001). The first proteins that were known to use an indirect readout mechanism were TATA binding protein (Kim *et al.* 1993), 434 repressor (Rodgers *et al.* 1993), and trp repressor (Scheviz *et al.* 1985). Since then, several other proteins have been characterized which rely on indirect readout: met repressor (Somers *et al.* 1992), integration host factor (Yang *et al.* 1989), P22 repressor (Wu *et al.* 1992), c-myb (Gabrielsen *et al.* 1991), papillomavirus E2 proteins (Hegde *et al.* 1992), and estrogen receptors (Schwabe *et al.* 1993). Although these proteins are found ubiquitously in organisms in all domains of life, they do not show much structural similarity to one another. For example, the papillomavirus E2 proteins utilize the α1-helix of a β1-α1-β2-β3-α2-β4 DNA-binding domain to insert into the major groove and

participate in direct readout as well (Hegde *et al.* 1992). The DNA-binding domain of 434 repressor, composed entirely of α-helices, utilizes a helix-turn-helix motif to contact base functional groups (Rodgers *et al.* 1993). Similarly, in *E. coli*, IHF utilizes antiparallel β-turns to make contacts with major groove bases (Yang *et al.* 1989). This shows that indirect readout is a universal mechanism used by many proteins to modulate sequence-specific recognition of DNA, but since it has not been well studied it has been called the "missing half" of binding site recognition (Steffen *et al.* 2002).

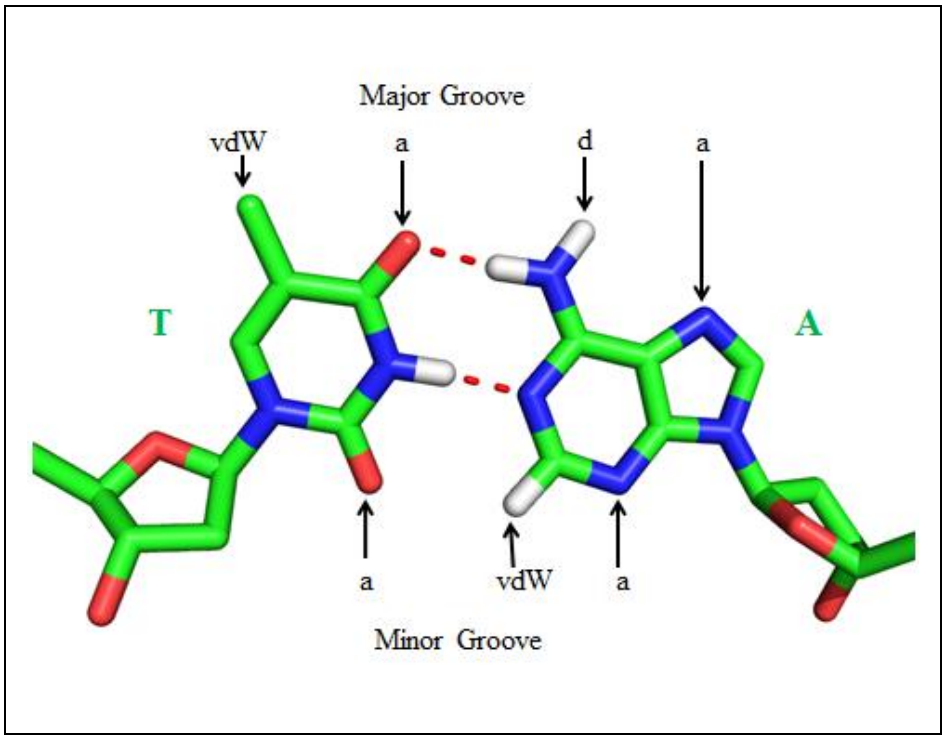


Figure 1.4: DNA base pairs showing possible discrimination sites. The arrows indicate sites that could be involved in direct readout. a = hydrogen bond acceptor; d = hydrogen bond donor; and vdW = van der Waals. Adapted from Kielkopf *et al.* 1998.

Zinc finger proteins

Zinc finger proteins are good examples where direct and indirect readout are known to occur. Figure 1.5 (left) shows a 1.6 Å structure of zinc finger protein, Zif268, bound to DNA and zinc ions (Elrod-Erickson *et al.* 1996, Elrod-Erickson *et al.* 1998). The major groove tends to provide direct, sequence-specific interactions since it can distinguish between AT and TA base-pairs based on the unique hydrogen bond donors and acceptors **(**Rhodes *et al.* 1996**)**. Zinc finger proteins have high affinity and specificity for their DNA target sites **(**Wolfe *et al.* 1999**)**. The schematic diagram (Figure 1.5 right) shows that most of the contacts made by three fingers occur with one strand of the DNA known as the primary strand. Arg at position -1 and 6, and His at position 3 of finger 2 mainly form hydrogen bonds with guanine on the primary strand of the DNA. Similarly, Asp2 tends to hydrogen bond to adenine on the complementary strand of the DNA resulting in the overlapping of the interactions. Arg-1 and Asp2 form a hydrogen bond which is believed to orient the Arg side chains and help increase specificity of the Arg-guanine interactions **(**Elrod-Erickson *et al.* 1996, Wolfe *et al.* 1999**)**. The structure also shows various water-mediated contacts between zinc fingers and DNA which also add to the specificity. Asp2 in all three fingers forms water-mediated contacts with cytosine, and also Arg6 of the helices with respective guanines. In finger one and three Arg-1 forms water-mediated contacts to the phosphate backbone. More water molecules were seen clearly with high resolution crystal structures. It appears that these water molecules are well ordered and might play a bigger role in aiding specificity (Wolfe *et al.* 1999). In other studies, for example, in the structure of *trp* repressor bound to DNA it has been

shown that there are few direct amino acids to base contacts but many water-mediated contacts. The water molecules were found to be present in the similar positions on the DNA even in absence of the protein, thus implying that the protein was recognizing the water associated structures as well in an indirect readout manner **(**Rhodes *et al.* 1996, Bareket-Samis *et al.* 1998**)**.

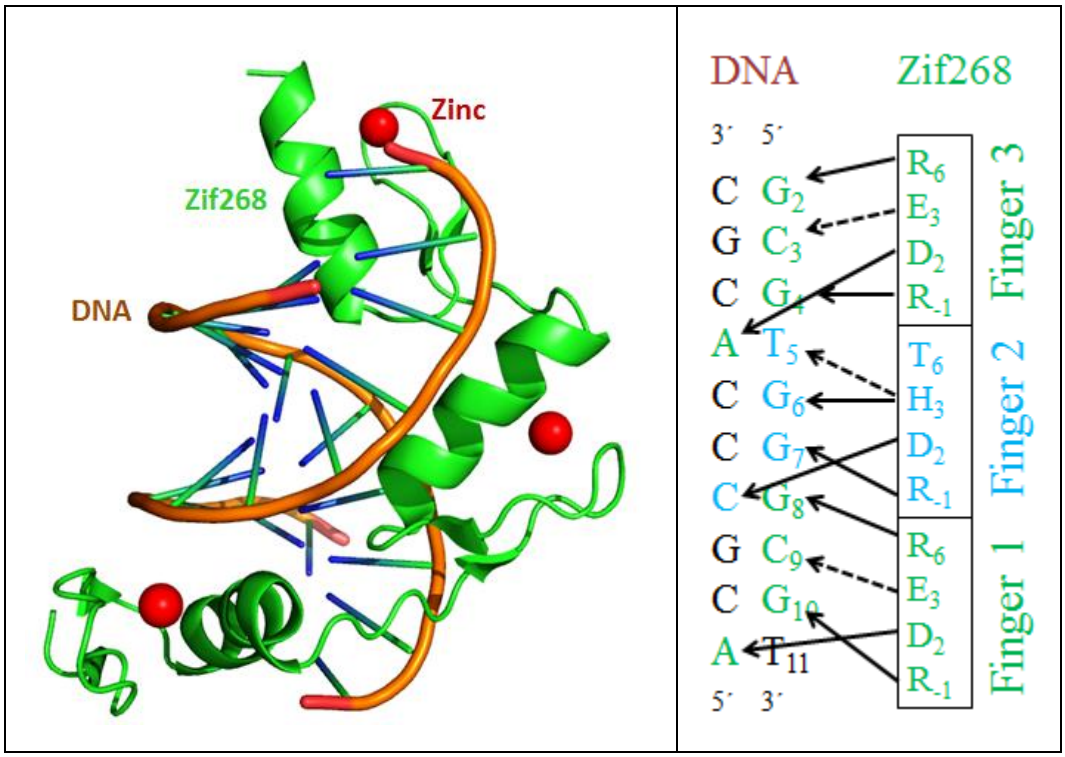


Figure 1.5: (Left) Ribbon diagram of zinc finger protein Zif268 bound to DNA along with the zinc ions. (Right) Schematic diagram showing three fingers of Zif268 bound to DNA. Solid arrow represents hydrogen bond, and the dashed arrow represents hydrophobic interaction. Adapted from Elrod-Erickson *et al.* 1998.

Integration host factor proteins

Another great example of indirect readout is seen in the integration host factor (IHF)-DNA complex (Figure 1.6). IHF, a ubiquitous protein in prokaryotes, employs both direct and indirect readout in maintaining DNA structure and gene regulation (Rice

1997). It binds to DNA as a histone-like protein to organize chromosome structure and bends DNA to form recombinogenic DNA complexes and to serve as a transcription factor for global gene regulation (Hatfield *et al.* 2002). IHF prefers specific binding sites even though the binding sites have a highly degenerate consensus sequence. The short consensus sequence is degenerate and has only 9 conserved bases, out of which only 3 are involved in direct readout recognition. This indicates the significance of indirect readout in specific binding of IHF to DNA (Rice *et al.* 2006). The binding of enzyme to the DNA causes the DNA to bend more than 160°. IHF primarily makes contacts using the phosphodiester backbone and the minor groove and relies mainly on indirect readout to recognize its binding sequence (Rice *et al.* 1996). Sequence specificity is provided most likely by the sum of a large number of small interactions, since there are only a few conserved bases in the region. The interaction of IHF with the six-base A-tract and well-ordered string of water molecules, also known as spine of hydration, indicates that the DNA sequence recognition takes place through structure rather than by base-specific contacts (Rice *et al.* 1996).

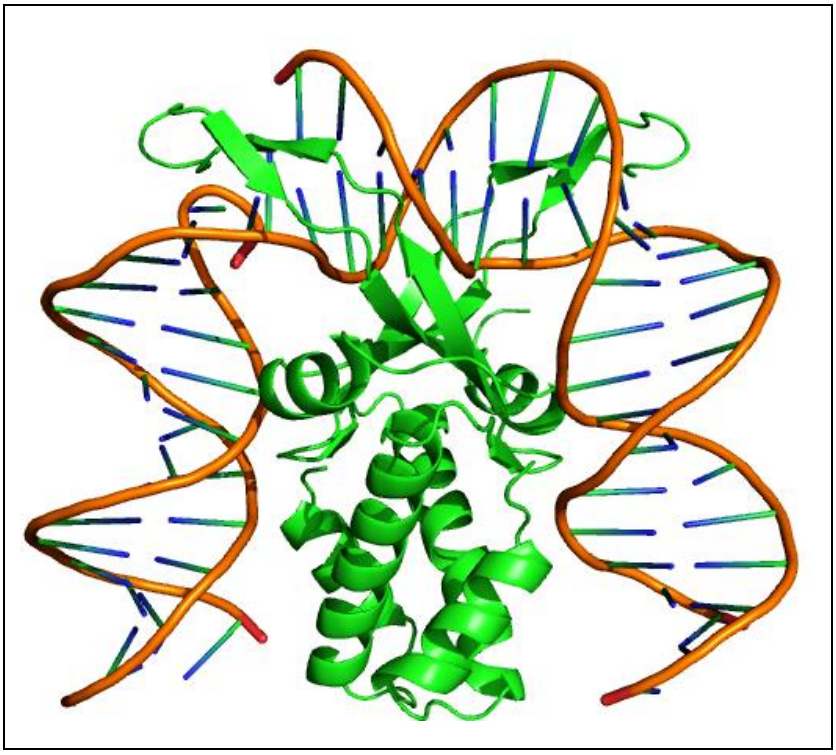


Figure 1.6: Crystal structure of an IHF-DNA complex showing a protein-induced DNA bending. Adapted from Rice *et al.* 1996.

Purine Purine > **Purine Pyrimidine** > **Pyrimidine Purine**	
Staked Bases	**Stacking Free Energy (kcal/mol)**
GpG	-7.79
GpA	-7.41
GpT	-5.93
GpC	-5.76
TpG	-5.25
CpG	-5.07

Table 1.1: Stacking energy of DNA bases showing Purine-Purine stacking being energetically more stable than Purine-Pyrimidine or Pyrimidine-Purine stacking. Adapted from Friedman *et al.*1995.

DNA sequence recognition has been studied extensively but there are still many unanswered questions regarding how enzymes carry out sequence-specific DNA binding. Some enzymes confer specificity by indirect readout of distorted DNA. How does the

enzyme bind to a specific sequence only? Does the enzyme recognize the distorted DNA or does the enzyme induce DNA to distort upon binding? It is also interesting to see how enzymes get activated upon binding to a specific sequence only and do not cleave other sequences. Some studies have shown that such enzymes bind the undistorted DNA initially, and eventually distort the DNA to form an enzyme-substrate complex (Jeltsch *et al.* 1996, Viadiu *et al.* 2000). In some cases the energetics of DNA bending may play a major role (Bareket-Samis *et al.* 1998, Chen *et al.* 2001, Koudelka *et al.* 1987). Enzymes may prefer to bind to pyrimidine-purine stacked bases as opposed to purine-purine stacked bases since the low stacking energy associated with such bases will allow the DNA to be bent more readily thus allowing proper binding (Table 1.1). There is also a possibility that once the enzyme and DNA come in proximity the binding and distortion takes place simultaneously (Parkhurst *et al.* 1996, Hiller *et al.* 2003). Alternatively, the DNA might be bent intrinsically which is then recognized by the enzyme (Jeltsch *et al.* 1996, Winter *et al.* 1981). It is also known that some disordered endonucleases become more ordered upon binding to the specific DNA (Spolar *et al.* 1994). In fact, structures of type II restriction endonucleases, PvuII (Athanasiadis *et al.* 1994), BamHI (Newman *et.al* 1995), BglII (Lukacs *et al.* 2001), and EcoRV (Perona *et al.* 1997) in the presence and absence of substrate show different conformations.

HincII

Indirect readout of DNA has been studied in another Type II RE, HincII. HincII, originally from *Haemophilus influenzae,* has a common core fold shared with SgrAI, and

both of these enzymes recognize palindromic sequences with degenerate bases. HincII recognizes and cleaves duplex DNA at sequences GTY|RAC (numbering is $G_1T_2Y_3R_4A_5C_6$, Horton *et al.* 2002). It makes many direct contacts with the bases and sugar-phosphate backbone in the major groove throughout the 6-bp recognition site. The structure of wild type HincII shows that direct readout contacts occur at all bases except at the Y_3 (Figure 1.7). In addition, the Y_3R_4 base pairs exhibit an unusual conformation, a distortion from B form likely stabilized by binding to the HincII enzyme. Q138 intercalates and hydrogen bonds with the exocyclic guanine oxygen of the outer G_1C_6 bp. Modeling suggested that the insertion of Q138 into the DNA duplex just outside of the 6-bp recognition site lead to distortions in the DNA at the Y_3R_4 base pairs that could be important for indirect readout of this part of the recognition sequence (Horton *et al.* 2002). Joshi *et al.* 2006 made and analyzed the mutant enzyme Q138F HincII, and found that it possessed altered sequence specificity at the center two base pairs of the cognate recognition site. The mutant enzyme has a 36 times greater cleavage rate for the sequence GTTAAC over GTCGAC in multiple turnover assays, compared to only 6-fold in the case of the wild type enzyme. The single turnover rate of DNA cleavage by Q138F HincII is reduced relative to that of the wild type enzyme, but is reduced much more for the sequence GTCGAC than for GTTAAC, resulting in a greater relative cleavage rate for the GTTAAC sequence. Interestingly, the binding affinities of the wild type and the mutant enzyme for the two sequences remain very similar, both having a two-fold preference for GTTAAC over GTCGAC (Joshi *et al.* 2006, Babic *et al.* 2008).

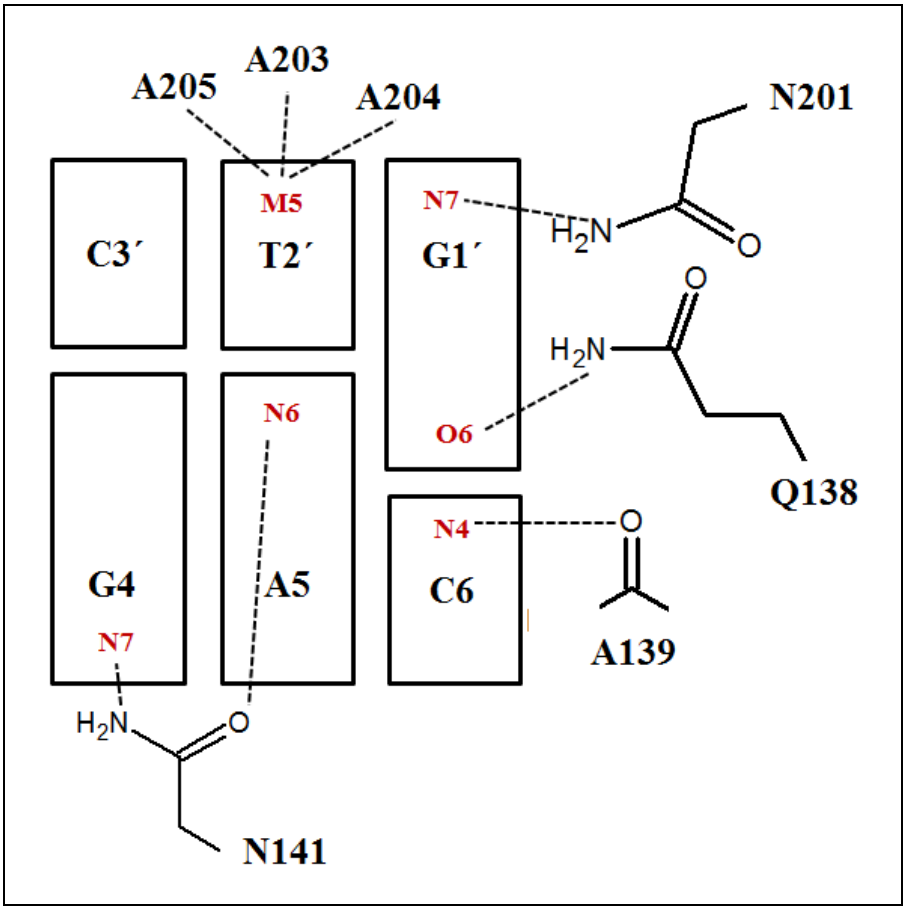


Figure 1.7: Cartoon showing protein-DNA contacts to bases in the major groove in the wild type HincII and DNA structure. Note: Only half of the palindromic recognition sequence is shown. Adapted from Little *et al.* 2008.

Structural studies of WT and Q138F HincII show that both the DNA and the protein conformations are altered when the enzyme binds to different DNA sequences. The wild type HincII distorts the DNA through both intercalation and bending. HincII bound DNA has slightly differently stacked bases than in B form DNA. Figure 1.8 illustrates how mutation of Gln 138 to Phe alters the interaction between the DNA and the protein. Upon mutation, the position of the side chain of Phe 138 is shifted in order to

stack properly in the intercalated DNA (Step 1). This, in turn, causes the main chain of HincII to shift, causing the neighboring Ala 137 side chain to move away from the DNA (Step 2); resulting in loss of contact between the Ala 137 methyl group and the deoxyribose ring of the DNA (Step 3). The altered puckering of the deoxyribose at Cyt 10 causes a twisting of the sugar-phosphate backbone of the DNA strand (Step 4). The change in conformation of the DNA causes adenine O5' to stick out towards the protein and results in pushing away the side chain of Thr 130 (Step 5). All of these changes cause shifts in active site residues Lys 129 to Asp 127. The phosphate of Ade 9 is rotated in such a way that it blocks the position of the putative nucleophilic water (Step 6). The altered conformation is unable to cleave efficiently in the active site due to the misalignment of the attacking water, Lys 129, the neighboring phosphate, the scissile phosphate, and the metal ions. This misalignment results in the reduced DNA cleavage rate. The cleavage rate is further reduced when the DNA has C_3G_4 at the center step instead of T_3A_4, because of the greater misalignment around the active site, and greater disruption of the cross-strand purine stacking (CSPS). When T_3A_4 is in the center, the weaker stacking of two adenines allows for adjustments in the DNA structure that result in reduction of the misalignments at the protein-DNA interface (Joshi *et al.* 2006, Babic *et al.* 2008).

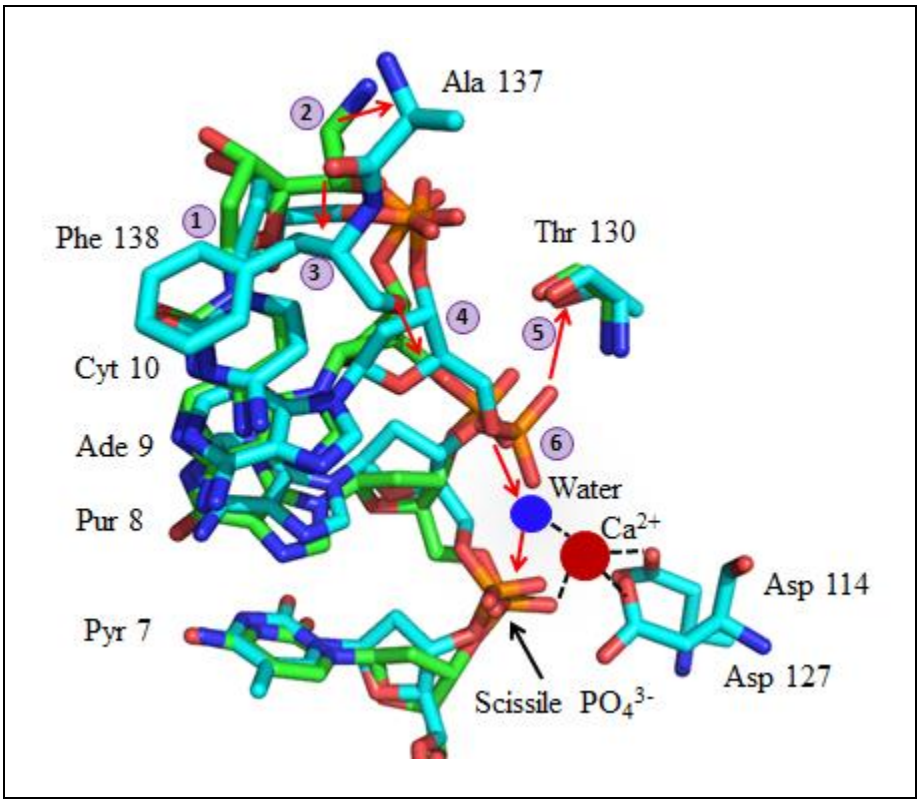


Figure 1.8: Superposition of WT HincII (white) and Q138F (colored) bound to DNA recognition sequence. The changes occurring due to Q138F mutation are described in steps above. Adapted from Babic *et al.* 2008.

Chapter 3 investigates the indirect readout of DNA by the type II restriction endonuclease SgrAI. A mutation, K96A, was made to investigate the role of an unstacking of the Y_7 and G_8 in the primary recognition sequence $C_1R_2C_3C_4G_5G_6Y_7G_8$. The structure of SgrAI bound to this recognition sequence shows the side chain of K96 in the minor groove in contact with the unstacked bases. The DNA sequence binding and cleavage preference of K96A SgrAI, as well as the structure of K96A SgrAI bound to primary site DNA, were determined. Unexpectedly, this mutation did not alter the structure of the enzyme, nor did it result in an enzyme without sequence preference at the 7^{th} position. Instead, the largest effect of the mutation appears to be in making the

enzyme more specific such that it fails to cleave either type of secondary site including CRCCGGGG containing a G at the 7th position, and CRCCGGYC which contains a C at the 8th position. It may be that the K96 side chain is required to distort the non YG sequences (specifically GG and TC) of secondary site DNA for proper positioning in the enzyme active site upon activation and specificity expansion.

Nucleotide Excision repair

DNA nucleases perform various biological functions such as replication, recombination, and repair. The DNA repair mechanism was first discovered by Albert Kelner at the Cold Spring Harbor Laboratory (Kelner 1949), and Renato Dulbecco at The University of Indiana (Dulbecco 1949) inadvertently. This photoreactivation phenomenon, whereby the UV light damaged DNA is repaired by a light-dependent enzyme reaction, was later explained by Friedberg *et al.* (Friedberg *et al.* 1995). DNA nucleases have high specificity and efficiency in performing activities such as endonucleolytic cleavage of double and single stranded DNA, and 5´ or 3´ exonucleolytic cleavage. Failure to perform these critical functions properly leads to diseases such as premature aging and premature cancer development (Thoms *et al.* 2007). It could also lead to autoimmune diseases (Nagata 2007) and repeat expansion diseases, a group of human genetic disorders caused by long and highly polymorphic tandem repeats (Fry *et al.* 1999, Parrish *et al.* 2006). One of the major DNA repair pathways carried out by DNA nucleases is called nucleotide excision repair (NER).

In general, NER takes place when the bulky DNA damage leads to distorted DNA helix. The NER mechanism can remove a wide range of DNA damage by dual incisions on both sides of the lesion (Petit *et al.* 1999). It can also be triggered by UV-induced DNA damage, such as cyclobutane-pyrimidine dimers or pyrimidine-(6-4)-pyrimidone photoproducts as well as polycyclic aromatic carbohydrates found in tobacco smoke and crosslinking agents. Cisplatin is one of such crosslinking agents (Buschta-Hedayat *et al.* 1999, Wood 1999, Mitchell *et al.* 1989). NER is a versatile and universal mechanism found in all living organisms tested so far, from the smallest free-living life form *Mycoplasma genitalium* to humans (Sancar 1996, Petit *et al.* 1999). The NER mechanism was first known in 1960s (Setlow *et al.* 1964, Boyce *et al.* 1964, Pettijohn *et al.* 1964) when the UV-induced damage of thymine dimers were found to be repaired in *E. coli*.

In *E. coli* bacteria, only three proteins, UvrA, UvrB and UvrC, are needed to perform NER (Petit *et al.* 1999). $UvrA_2B$ complex locates the lesion by tracking along the DNA. Formation of transient $UvrA_2B$-DNA complex kinks and unwinds the DNA in an ATP-dependent reaction. UvrA, the molecular matchmaker, dissociates allowing formation of stable UvrB-DNA complex. Once UvrC binds to the UvrB-DNA complex, UvrB makes the 3´ incision and then, UvrC makes the 5´ incision shortly (Sancar 1996, Petit *et al.* 1999, de Laat *et al.* 1999). In contrast, Human NER (Figure 1.9) require 18 different proteins, including 7 xeroderma pigmentosum (XPA to XPG) proteins, the trimeric replication protein A (RPA), and the multisubunit (7-10) general transcription factor (TFIIH), to fix the damage (Sancar 1996). In human NER, the DNA damage (black star, Figure 1.9) is initially recognized by XPC/HR23B (red). Helicases XPB and XPD

(light blue), part of the general transcription factor (TFIIH), are recruited to unwind the DNA and generate the bubble. XPA (green) and RPA (purple), the single stranded DNA binding proteins, bind the damaged and undamaged DNA strands, respectively. They are believed to guide XPG (dark blue) and XPF (orange) to cleave the damaged strand only. XPG cleaves the 5′ end of the damage, and the heterodimer XPF/ERCC1 cleaves on the 3′ end (Gillet *et al.* 2006, Thoms *et al.* 2007). In prokaryotes, a 12-13 base-long oligomer is excised out from the damaged region, whereas, in eukaryotes, a 24-32 base-long oligomer is excised out (Sancar 1996). Although the basic NER steps are very similar in both prokaryotes and eukaryotes, the subunits of prokaryotic excinuclease do not share significant homology with any of the eukaryotic excinuclease subunits (Sancar 1996).

Defects in NER in humans may lead to diseases such as xeroderma pigmentosum (XP), Cockayne's syndrome (CS), and trichothiodystrophy (TTD). Symptoms of XP are photodermatoses and neurological abnormalities. Photodermatoses include extreme photosensitivity to sunlight with manifestations ranging from erythema to xerosis and skin atrophy with a high probability of acquiring skin cancer. Neurological symptoms include mental retardation, progressive ataxia, and deafness (Cleaver *et al.* 1989). CS patients can suffer from cachectic dwarfism, photosensitivity, mental retardation, and progressive neurological symptoms caused by demyelination (Nance *et al.* 1992). TTD patients have sulfur-deficient brittle hair and suffer from dental caries, ichthyosis, skeletal abnormalities, and progressive mental retardation (Stefanini *et al.* 1993, de Laat *et al.* 1999).

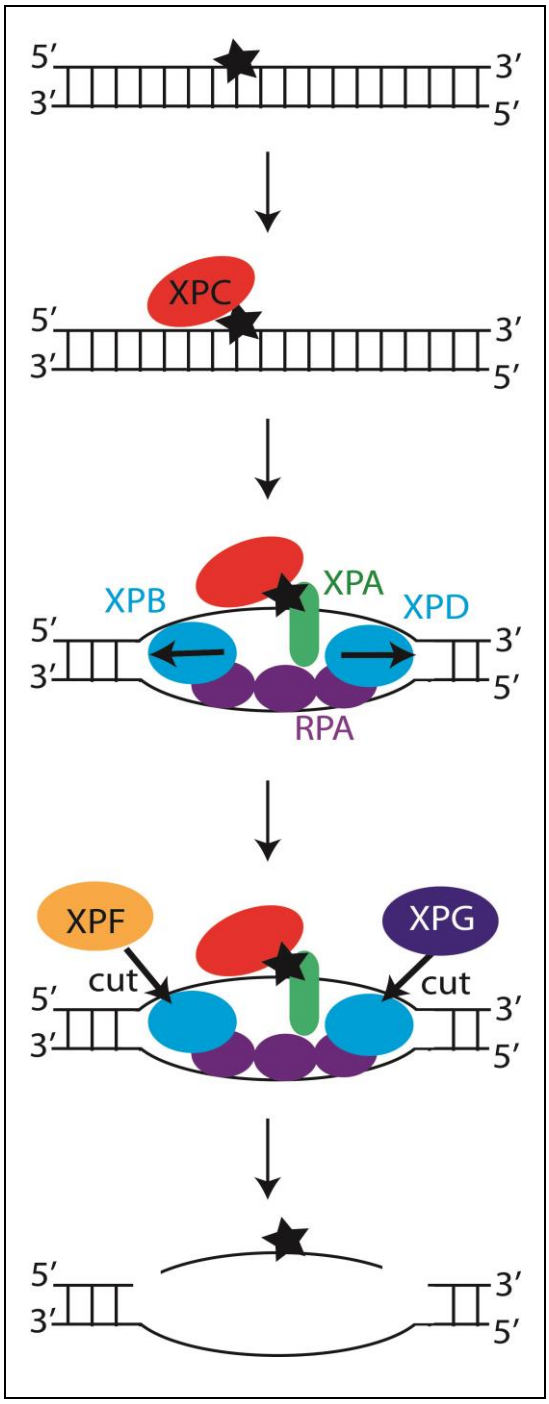


Figure 1.9: A model showing only the key proteins involved in human NER pathway. Star represents the damaged site in the nucleotide. Figure courtesy of Dr. Nancy C. Horton.

The now widely accepted phylogenetic tree shows that the archaeal species branch out from eukaryotes after the branching from bacteria. This suggests a closer ancestral relationship between archaea and eukaryotes as opposed to bacteria (Woese *et al.* 1990). Sequencing of first the archaeal genomes showed strong similarities between archaeal and eukaryotic DNA replication machineries. However, the archaeal DNA repair pathways show similarities to both bacterial and eukaryotic processes. Bacterial UvrABC excision repair machinery has been found in certain mesophilic archaea. However, most archaea, including all hyperthermophiles, have several homologues of the eukaryotic NER nucleases XPF-ERCC1 and FEN-1 (a homologue of XPG) and helicases XPB and XPD. This indicates that the eukaryotic-like NER pathway could be working in archaea. Interestingly, some archaea, such as *Methanothermobacter thermautotrophicus*, and *Methanosarcina mazei*, have both UvrABC and homologues of the eukaryotic NER. In such cases, the UvrABC proteins are thought to perform NER, and the XP homologues are involved in other functions such as replication and/or replication related repair (Kelman *et al.* 2005). But, how does the hyperthermophilic archaea, which do not have UvrABC-like system, carry out the NER function?

The homologues of damage recognition proteins XPA and XPC are not present in any of the archaea. So how does the NER pathway become activated without the presence of proteins that detect such damage? Some higher plants also do not have these proteins, so does that mean these proteins are not essential and they might have been a relatively recent evolutionary addition to the NER system? The structure of Rad4/Rad23, the yeast homolog of XPC/hHR23B, bound to DNA containing a thymidine dimer shows

intercalation of the protein into the DNA duplex at the damage site (Min *et al.* 2007). This indicates that the damage is being detected by the weak base stacking in the duplex. So for hyperthermophilic archaea the damage recognition need not be carried out by some specific proteins, instead, the damaged DNA might already be in a partially melted state which could in turn be recognized by single stranded DNA binding proteins. *M. kandleri* has Mk1441, a human homologue of RPA which is a single stranded DNA binding protein. On the other hand, the archaeal genome does not have other eukaryal-type NER enzymes as well, suggesting that there might not be any NER-like pathway the way we understand. The homologous proteins found in archaea can also perform alternative functions *in vivo*. For example, FEN-1 is involved in Okazaki fragment processing as well as in base excision repair. Similarly, the helicases XPB and XPD play important roles in transcription initiation as well as NER (Kelman *et al.* 2005). It is very likely that in the absence of XPA or XPC, some other archaeal proteins have an additional function of damage recognition. The recently solved crystal structure of *Aeropyrum pernix* XPF bound to DNA indicates a mechanism for DNA substrate recognition, distortion and DNA cleavage (Newman *et al.* 2005). The *M. kandleri* homolog of XPF (Mk1678) contains a helicase domain along with the nuclease domain as is regarded as Hef enzyme for having the dual helicase and endonuclease functions. Alternatively, it is also possible that there are still other proteins to be discovered in the *M. kandleri* NER pathway. The variation in the number of proteins involved in the NER pathway could be the result of the evolutionary need to create new, non-homologous

NER proteins as the hyperthermophilic archaea transitioned to mesophilic temperatures (Kelman *et al.* 2005).

The NER pathway has not been studied well in archaea. The only published NER-like patch repair pathway has been studied in *M. thermautotrophicus*, which has a complete set of UvrABC genes (Ogrunc *et al.* 1998). Therefore, the study of a NER-like repair pathway in the archaeon *M. kandleri* is important to shed light on the functions of the homologous proteins, to understand the repair pathway and how it relates to bacterial and eukaryotic repair pathways evolutionarily.

Methanopyrus kandleri

Methanopyrus kandleri are hyperthermophilic rod-shaped, gram positive archaea. These methanogens can grow at very high temperatures ranging from 84˚C to 110˚C using H_2 and CO_2 as sole energy and carbon source, respectively. They can also grow at high salt concentrations. Their genomes have much higher G+C-content, 60%, than other species of *Methanopyrus*, which might explain their ability to survive in such high temperatures. They were first discovered on the sediment cores and "smoker" walls from the Guaymas Basin, Gulf of California hot vent area at the depth of 2000 meters by the research submersible ALVIN (Huber *et al.* 1989, Kurr *et al.* 1991). Their genus name, *Methanopyrus*, means "methane fire" and their species type, *kandleri*, was named in honor of Otto Kandler, a microbiologist and botanist who was one of the first scientists to propose a new domain archaea along with bacteria and eucarya (Huber *et al.* 1989, Kurr *et al.* 1991, Woese *et al.* 1990).

Studying NER in archaea may provide an opportunity to better understand the more complex version of the eukaryotic NER system. This might also aid in learning the evolutionary aspects of DNA repair. Moreover, for practical purposes, archaeal enzymes are easier to express, purify, and crystallize than the eukaryotic equivalents. Chapter 4 describes studies of the *M. kandleri* XPG homologue, Mk0566. The enzyme was prepared recombinantly, and its crystal structure was solved to 2.48 Å resolution. Although, a DNA substrate was present in the crystallization solution, no electron density for bound DNA was visible. The structure of Mk0566 was found to be very similar to that of human FEN-1 and to other archaeal FEN-1/XPG homologues. Humans contain both FEN-1 and XPG, with FEN-1 performing functions in DNA replication, while XPG performs its function in Nucleotide Excision Repair (NER). These results, along with the previous studies substrate specificity of Mk0566, suggest that the main biological role of Mk0566 is in DNA replication; however, they do not preclude involvement in a modified form of NER.

CHAPTER 2

ACTIVITY, STRUCTURE, AND BIOLOGICAL ROLE OF ACTIVATED SGRAI OLIGOMER

Abstract:

Restriction endonucleases are enzymes that recognize and cleave specific DNA sequences. SgrAI is a type IIF restriction endonuclease that cuts an 8 base pair primary sequence, CR|CCGGYG (| indicates cut site, R denotes A or G, and Y denotes C or T). In addition to its primary cleavage activity it also cleaves secondary sequences CR|CCGGGG or CR|CCGGY(A/C/T), but only appreciably in the presence of the primary sequence. Primary site DNA with enough flanks has been shown to be an allosteric effector which stimulates cleavage of both primary and secondary site DNA sequences by SgrAI.

A previous study (Park *et al.* 2010b) showed that the activation of SgrAI was dependent on the concentration of SgrAI bound to primary site DNA. Increasing concentrations of primary site bound SgrAI was found to result in faster rates of DNA cleavage (>200 fold) of primary site DNA, but only a modest increase (~2-3 fold) in the rate of cleavage of secondary site DNA. Although the increase in rate was small with secondary site DNA, the percent of the total amount of secondary site DNA cleaved in the assays was found to increase with the concentration of primary site bound SgrAI. It was also found, in the same study, that primary site bound SgrAI forms large, heterogeneous (in size) oligomers with increasing concentrations of primary site bound SgrAI. These assays utilized 1 μM SgrAI, 1 nM 18 bp DNA containing either the primary

or secondary site to measure DNA cleavage and SgrAI/DNA oligomerization (18-1AT or 18-2A), and varied concentrations of a precleaved 40 bp primary site containing DNA as the activator (PC DNA). These assays also only measured DNA cleavage rates at four concentrations of PC DNA (0, 10 nM, 100 nM, 1000 nM). In order to determine the levels of activation of SgrAI at intermediate concentrations of PC DNA, and to better characterize the activated cleavage of secondary site DNA, these assays were repeated with a longer reporter DNA, containing 40 bp total and either the primary site (40-1AT) or two different types of secondary site (40-2A:CACCGGGG and 40-2B:CACCGGTC). These assays were also performed with a total of 7 different concentrations of PC DNA (0, 10 nM, 100 nM, 250 nM, 500 nM, 1 μM, 2 μM). The intermediate levels of activation with intermediate concentrations of PC DNA were found, as predicted by the prior studies, and significantly a much greater acceleration of secondary site DNA cleavage than seen with the shorter DNA constructs (over 1000 fold) was observed.

Activated cleavage of secondary site DNA was also revisited, however, using DNA constructs with longer flanking DNA, the 7 concentrations of PC DNA listed above, and using sequences representing the two types of secondary sites. The base pairs of the recognition site CRCCGGYG substituted in the secondary site DNA constructs (position 2 and 7 in the case of 40-2A, and 1 and 8 in 40-2B) are recognized using fundamentally different means (Dunten *et al.* 2008), using either indirect, or direct readout, respectively. Without activation, SgrAI cleaves secondary site DNA very slowly, but cleavage of the two types of sites differs; the two strands of the secondary site are cleaved with similar, independent rate constants in the case of 40-2A, but with markedly

different rate constants in the case of 40-2B, in the absence of activation. The strand of 40-2B containing the substitution closest to the cleavage site is cleaved much more slowly than the opposite strand. Interestingly, along with acceleration of secondary site cleavage, this asymmetry is also lost upon activation of SgrAI, suggesting a loss of the communication between outer base pair recognition and the active site in the activated conformation of SgrAI. Finally, the cleavage rate of a noncognate site (with two base pair substitutions in the recognition sequence) 40-NCTA by SgrAI was investigated and found to be similar to the cleavage of secondary sites under conditions without activation, but unlike secondary site DNA, its cleavage was not accelerated upon activation of SgrAI.

One previously reported x-ray crystal structure of primary site DNA bound SgrAI shows two DBD associated in a head-to-head fashion via domain swapping of the N terminal 26 amino acids (Park *et al.* 2010a). Theories of domain swapping predict the formation of closed ended "dimers" as well as "run-away" domain swapped oligomers. This crystal structure may be an example of such a closed-ended "dimer" (a dimer of DNA bound SgrAI dimers), and the run-on oligomer may be an example of the predicted "run-away" domain swapped oligomer. Although the resolution of the cryo-EM structure (8.6 Å) cannot distinguish between swapped and unswapped configurations, domain swapping is a possibility given the proximity of neighboring N-termini of adjacent SgrAI within the oligomer. To investigate the possibility of domain swapping in the activated form of SgrAI, residue R29, which lies in the linker that connects the swapped domain (residues 1-22) to the rest of the protein, was mutated to alanine. R29 makes a hydrogen

bond to the swapped linker of a neighboring SgrAI subunit, but only in the swapped conformation. The R29A mutant was designed to disrupt this interaction and potentially destabilize the domain swapped conformation. The previously described single turnover DNA cleavage assay was utilized to measure any effects on activation of SgrAI mediated DNA cleavage by this mutation. Although the R29A mutation decreased the solubility of SgrAI, no large effects on the rate of DNA cleavage at different PC DNA concentrations was seen. Consistent with these results, the x-ray crystal structure of SgrAI R29A bound to DNA (resolution 1.65 Å) in its low activity form showed same conformation as WT SgrAI. Thus, the R29 side chain does not appear to be important for the oligomerization of SgrAI or the activation of DNA cleavage by SgrAI, whether in domain swapping or via other interactions in the run-on oligomer.

Prior models of the activated form of SgrAI suggested that it may oligomerize utilizing the same tail-to-tail interaction found in the homolog NgoMIV. To test this mode of interaction, residue E301, found in the postulated interface, was mutated to tryptophan and DNA cleavage and SgrAI oligomerization measured. The results again show similar activity to that of WT, predicting that SgrAI does not oligomerize in the tail-to-tail fashion as in the case of NgoMIV, which is consistent with the cryo-EM model.

The structure of the low activity, dimeric form of SgrAI bound to DNA has been solved to high resolution by x-ray crystallography (Dunten *et al.* 2008), and the structure of activated, oligomeric SgrAI has now been solved by cryoelectron microscopy (cryo-

EM) to 8.6 Å (Lyumkis *et al.* 2013). The cryo-EM structure shows loops 56-60 and 122-140 of SgrAI make contacts with the flanking DNA of a neighboring SgrAI, and appear to stabilize the run-on oligomer. To obtain a higher resolution structure of the activated form of SgrAI, the x-ray crystal structure of WT SgrAI, bound to the shortest DNA found to activate SgrAI, 22-1HT DNA, was determined at a resolution of 3.5 Å. Although the conformation of SgrAI did not show the expected changes (based on the cryo-EM structure) and was found to be identical to that of the low activity form (within coordinate error), the crystal structure did show contacts between neighboring SgrAI reminiscent of the contacts formed in the cryo-EM structure. For example, interactions are made between the N-termini of the neighboring SgrAI DNA-bound-dimers (DBDs), which had not been seen in any previous crystal structure. In addition, contacts are made between the flanking DNA of one DBD and the previously identified loops from neighboring SgrAI as seen in the cryo-EM model. However, rather than a left handed helix, the DBD interact via their N-termini forming a complex network in all three dimensions within the crystal.

Finally, to test the postulated biological role of run-on oligomer formation by activated SgrAI, namely to sequester activated SgrAI away from the host genome, DNA cleavage of secondary sites on mixtures of plasmids with or without primary site DNA was measured. At low concentrations of SgrAI and DNA, simulating expected physiological concentrations, plasmid assays showed that the secondary sites of the host-like DNA were protected, and only the sites on the phage-like DNA (containing both

primary and secondary sites) were cleaved, consistent with sequestration of SgrAI. As expected, all sites on both plasmids are cleaved at higher concentrations.

Part A: Activity of SgrAI in the presence of longer flanking DNA

Introduction

Various cellular processes such as repair, gene regulation, transposition, and site-specific recombination require specific interactions between protein and DNA. Type II restriction endonucleases are bacterial enzymes that are thought to cleave phage DNA injected into the cell prior to replication, and thus, protect their host from phage infection (Pingoud *et al.* 2005). Sequence-specific endonucleases capable of cleaving longer unique recognition sequences are highly sought for genomic work, since longer sequences occur less frequently and allow the manipulation of larger DNA fragments. Most of the type II restriction endonucleases characterized to date cleave 4-6 bp recognition sites in DNA (Roberts *et al.* 2001, Roberts *et al.* 2003); however, SgrAI, a type IIF restriction endonuclease from *Streptomyces griseus*, cleaves an 8 bp primary recognition sequence (1°), CR|CCGGYG (Tautz *et al.* 1990) (| indicates cut site, R denotes A or G, and Y denotes C or T), and exhibits several unusual biochemical behaviors. First, SgrAI cleaves plasmids bearing two copies of its recognition sequence faster than those bearing only a single site (Bilcock *et al.* 1999, Daniels *et al.* 2003). In addition, plasmid assays revealed that SgrAI will also cleave the secondary sites (2°) CR|CCGGGG or CR|CCGGY(A/C/T) in the presence of the primary site sequence (Bitinaite *et al.* 2002). Secondary sites are distinct from star sites, in that the secondary sites are cleaved appreciably under solution conditions optimal for primary sequence cleavage. In contrast, star site sequences are sequences that are cleaved appreciably only

under special reaction conditions, such as high enzyme concentrations or the presence of organic solvents or Mn^{2+}, and are discriminated against under optimal enzyme conditions by 2-4 orders of magnitude (Pingoud *et al.* 2001). Overall, activated SgrAI can cleave 17 different sequences (3 primary and 14 secondary) through expansion of DNA sequence specificity.

The mechanism of activation of SgrAI has been investigated using both structural and kinetic methods. Assays containing low concentrations of radiolabeled reporter DNA (1 nM) containing the SgrAI recognition sequence, excess SgrAI (1 μM dimer), and varied levels of unlabeled activating DNA (0-1 μM) show that the single turnover rate of DNA cleavage of the reporter DNA is increased with increasing concentrations of SgrAI bound to the activating DNA (Park *et al.* 2010b). This activating DNA is a duplex oligonucleotide containing sequences simulating a precleaved primary recognition site with 16 bp of flanking DNA (PC DNA, Table 2.8). Native gels have been used to investigate changes in SgrAI/DNA complexes with increasing concentrations of SgrAI bound to activating DNA, and show that these complexes migrate more slowly with increasing concentrations of SgrAI/activating DNA, indicating the formation of complexes of increasing size. Analytical ultracentrifugation was used to measure the sizes of the SgrAI/DNA complexes, determining that they are heterogeneous in size, and contain in the range of 4-12 copies of the DNA bound SgrAI (Park *et al.* 2010b). The size of the complexes has been confirmed by mass spectrometry (Ma *et al.* 2013) and electron microscopy (Lyumkis *et al.* 2013). The current model of activation of SgrAI includes the

formation of run-on oligomers which are stabilized by SgrAI bound to primary site DNA, but may bind and thereby activate SgrAI bound to either primary or secondary site DNA.

The high resolution of the structure of SgrAI bound to primary (Dunten *et al.* 2008) and secondary (Little *et al.* 2009) site DNA show SgrAI in the presumably low activity dimeric conformation. The activated, oligomeric form of SgrAI has been structurally characterized using cryoelectron microscopy (Lyumkis *et al.* 2013), and shows a left handed helical arrangement of the DNA bound SgrAI dimers in a run-on oligomer. Fitting of the crystal structure of the low activity dimeric SgrAI into the cryo-EM envelope suggests a conformational change involving rotation of one subunit of the dimer by 10º relative to the other (Lyumkis *et al.* 2013) (Figure 2.1). This conformation is predicted to exhibit the activated DNA cleavage activity.

Previous assays utilized primary and secondary site DNA embedded in 18 bp DNA as the reporters for DNA cleavage by SgrAI (Park *et al.* 2010b). In these assays, the single turnover DNA cleavage rate constant for cleavage of primary site DNA is accelerated by over 200 fold when the concentration of SgrAI bound to activating DNA is increased, however that for secondary site is accelerated by only 2-3 fold. In contrast, activation of secondary site cleavage in plasmid DNA by neighboring primary DNA sites appears more robust than the 2-3 fold seen in assays (Bitinaite *et al.* 2003). A major difference between the assays with the 18 bp oligonucleotides and those with plasmid DNA is the amount of DNA flanking the recognition site. Secondary site DNA in the 18mer contains only 5 bp flanking the 8 bp recognition site, while that in plasmid DNA

contains many more flanking bp. Further supporting the role of the flanking DNA in activated DNA cleavage by SgrAI, it has been shown that levels of activation increase with increasing lengths of flanking DNA, being very limited with 5 flanking bp but robust with 16 (Park *et al.* 2010b, Lyumkis *et al.* 2013). The DNA flanking the recognition site appears to interact with neighboring SgrAI in the cryo-EM model of the activated, oligomeric SgrAI (Lyumkis *et al.* 2013) (Figure 2.1). The dependence of SgrAI activation on the flanking DNA and its positioning in the cryo-EM structure of oligomeric SgrAI, both suggest that the activated, oligomeric form of SgrAI is stabilized by interactions involving the flanking DNA, and therefore the assays with 18 bp DNA contained flanking DNA of insufficient length to result in the tightest association with the run-on oligomer. Therefore DNA cleavage measurements were repeated using longer DNA, i.e. with 40 bp DNA, which has 16 bp flanking DNA.

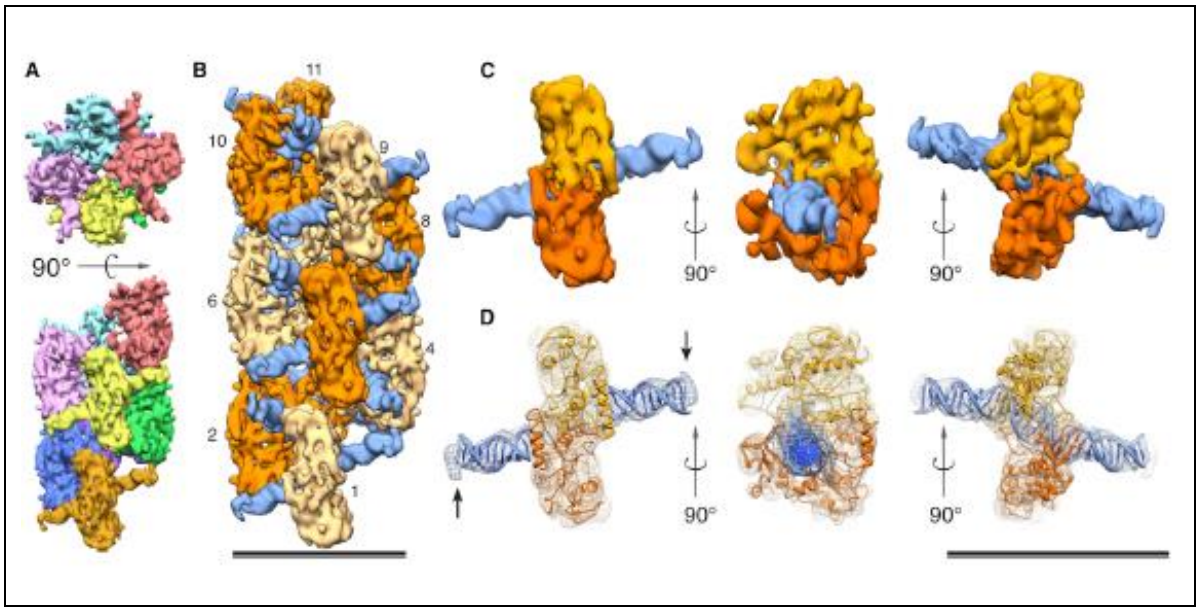


Figure 2.1. Cryo-EM model of run-on activated, oligomeric SgrAI. (A) Top and side views showing the organization of helical asymmetric units (DBDs) within the SgrAI oligomers. Eight distinct DBDs have been segmented out of the cryo-EM map and are each colored differently. (B) Helical reconstruction of oligomeric SgrAI at 8.6 Å resolution, segmented into 11 individual DBDs and labeled by helical asymmetric unit. Protein components of units 1, 4, 6, 7, 9 and 2, 3, 5, 8, 10, 11 are shaded light and dark, respectively. (C) Segmentation and different views of an individual helical asymmetric unit. Each helical asymmetric unit contains two monomeric SgrAI protein subunits (colored light and dark orange), and two copies of precleaved DNA (both colored blue). (D) Flexibly fit coordinates of SgrAI (Dunten *et al*., 2008) into the EM density of a segmented helical asymmetric unit. Arrows mark DNA disorder in the terminal regions. Scale bar is 150 Å. Figure from Lyumkis *et al*. 2013. Copyright permission attached in Appendix F.

SgrAI exhibits secondary site activity in addition to its primary site cleavage activity, cleaving two different types of secondary site sequences $C_1R_2|C_3C_4G_5G_6\underline{G}_7G_8$ or $C_1R_2|C_3C_4G_5G_6Y_7(\underline{A/C/T})_8$ in the presence of the primary sequence, $C_1R_2|C_3C_4G_5G_6Y_7G_8$ (Bitinaite *et al.* 2002). Since the sequence substitutions in the secondary sites are at varying locations, and are recognized by distinct mechanisms (indirect readout for Y_7, direct readout for G_8, Dunten *et al*. 2008), single turnover DNA cleavage measurements

were made with sequences representing the two types of secondary sequences: one with Y_7 in the primary sequence being substituted by G, and another one with G_8 being substituted by a C (with the complementary strand being substituted at R_2 with C and C_1 with G, respectively). Previous studies of secondary site cleavage by SgrAI (Park *et al.* 2010b) utilized only the one type of secondary site DNA, as well as DNA constructs with short flanking DNA (18 bp DNA with 5 bp flanking bp). As mentioned above, the flanking DNA has been shown to be critical in activation of SgrAI (Lyumkis *et al*. 2013), and appears to form stabilizing contacts within the activated oligomer (Lyumkis *et al*. 2013), hence these studies were also carried out with 40 bp constructs, containing 16 bp flanking either side of the recognition sequence. As a control, assays were also performed with a noncognate sequence, embedded in a 40 bp DNA, which contained two substitutions in the 8 bp recognition sequence.

Results

Single turnover DNA cleavage assays of primary, secondary, and noncognate site containing DNA embedded in a 40 bp DNA were performed with radiolabeled DNA at a concentration (1 nM), which is below that capable of activation, and with 1 μM SgrAI and varied concentrations of added unlabeled stimulatory PC DNA (the rate constants are shown in Table 2.1, and the DNA sequences for constructs are shown in Table 2.8). Figure 2.2 shows an example of a denaturing PAGE gel of the reaction products from a single-turnover DNA assay. Image analysis was performed using ImageQuant (GE Healthcare Life Sciences). The ratio of cleaved product to total DNA was plotted against time, and fitted to a single exponential equation. The resulting rate constants are

compared in Figure 2.3. In the case of the primary site DNA, 40-1AT, and without added PC DNA, the cleavage data fit well to a single exponential function with a rate constant of 0.236±0.006 min^{-1}. With the addition of PC DNA, fitting of the cleavage data required two single exponential functions; suggestive of two independent processes (the percent cleaved with each rate constant is shown in Table 2.1). The majority of the 40-1AT DNA was cleaved with the faster of the two rate constants, which ranged from 0.37±0.01 min^{-1} at 10 nM PC DNA to 23.1±2.1 min^{-1} with 2 μM PC DNA, a ~100 fold increase over cleavage without the added PC DNA. To test the stimulatory capacity of the 40-1AT DNA itself, unlabeled 40-1AT was added to the reactions at 1 μM. In this case, a large proportion of the cleaved DNA was cleaved with a single turnover rate constant of 12±1 min^{-1}, a ~40 fold increase.

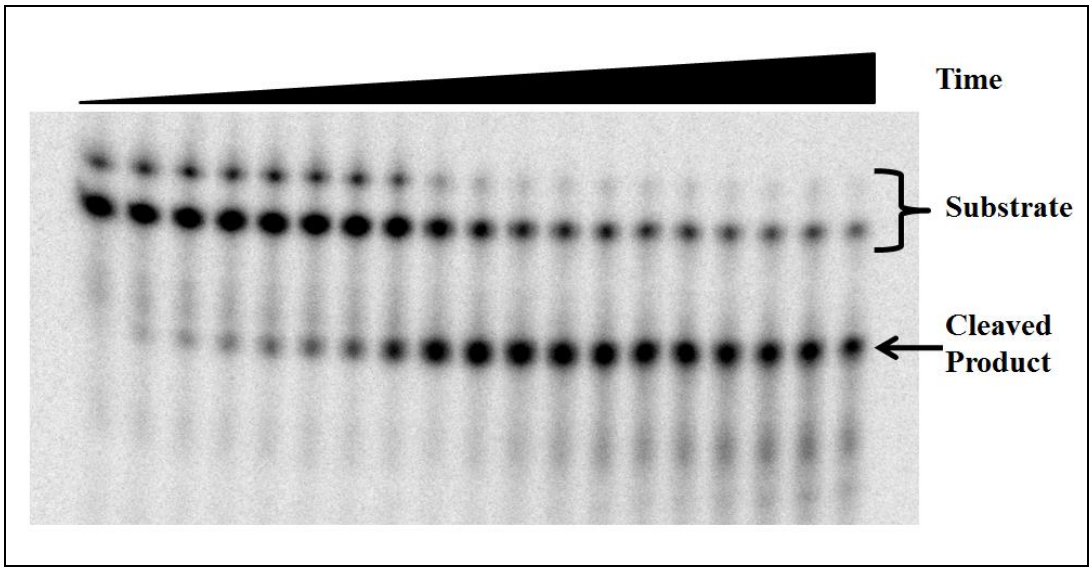


Figure 2.2: A sample denaturing polyacrylamide gel for a single turnover DNA cleavage rate constant determination. The percent of product formed as a function of time was determined by integrating the density of both cleaved and uncleaved DNA bands, and normalizing to the total amount cleaved.

Table 2.1: Single-Turnover DNA Cleavage Rate Constants using 1 μM SgrAI at 37°C

^{32}P-labeled DNA (1 nM)	Added unlabeled DNA	Accelerated Rate constant (min^{-1})	% Cleaved, Accelerated	Nonaccelerated Rate constant (min^{-1})	% Cleaved, Non accelerated
40-1AT	0 nM PC			0.236±0.006	68±1
	10 nM PC	0.37±0.01	71±1	0.0149±0.0007	8.6±0.6
	100 nM PC	4.4±0.4	69±2	0.0337±0.0001	14±3
	250 nM PC	11.0±0.3	71.6±0.4	0.0353±0.0008	14.3±0.3
	500 nM PC	18.0±0.3	75.8±0.2	0.039±0.005	11.5±0.2
	1 μM PC	16.3±1.7	76.3±0.4	0.053±0.005	10.3±0.3
	2 μM PC	23.1±2.1	73.0±0.8	0.23±0.12	4±1
40-1AT	1 nM 40-1AT	0.28±0.02	63±2	0.026±0.007	8±3
	1 μM 40-1AT	12±1	29±3	0.069±0.007	42±2
40-2A	0 nM PC			0.013±0.004	17.7±0.5
	10 nM PC			0.018±0.004	70±7
	100 nM PC	0.039±0.005	63±2		
	250 nM PC	0.82±0.09	28±2	0.036±0.003	39.3±0.2
	500 nM PC	6.1±0.6	53±1	0.038±0.005	24±5
	1 μM PC	9.0±2.5	53±7	0.036±0.011	19±5
	2 μM PC	10.3±2.7	38±3		
40-2A	1 nM 40-2A			0.004±0.003	34±12
	1 μM 40-2A	0.25±0.06	78±5		
40-2A-Top	0 nM PC			0.0010±0.0001	50±1
	1 μM PC	9.1±1.0	50±4	0.31±0.07	29±4
40-2A-Bottom	0 nM PC			0.0024±0.0018	35±26
	1 μM PC	6.3±1.1	39±6	0.030±0.002	32±6
40-2B	0 nM PC			0.002±0.001	58±31
	10 nM PC			0.011±0.004	40±6
	100 nM PC			0.0221±0.0009	54±3
	250 nM PC	1.23±0.07	20±4	0.031±0.003	37±3
	500 nM PC	2.2±0.4	29±2	0.03±0.01	23±3
	1 μM PC	8.1±1.4	33±2	0.24±0.03	20±1
	2 μM PC	11.8±1.4	45±2	0.45±0.14	14±1

40-2B-Top	0 nM PC			0.0129±0.0004	69.7±0.2
	1 μM PC	9.3±1.6	55.2±0.3	0.24±0.01	28.8±0.5
40-2B-Bottom	0 nM PC			0.0023±0.0009	46±17
	1 μM PC	16.9±4.3	50±1	0.40±0.06	29±1
18-1AT*****	0 nM PC			0.094±0.015	
	1 μM PC	22±7			
18-NCTA[§]	0 nM PC			0.010±0.001	21.8±0.8
18-NCCG[§]	0 nM PC			0.031±0.005	22±1
	1 μM PC			0.021±0.004	22±1
40-NCTA	0 nM PC	0.024±0.004	11±1	0.0023±0.0002	7.5±0.2
	1 μM PC	0.052±0.006	5.2±0.6	0.0048±0.0008	11.7±0.6

*****From Park, *et al.* 2010b.
[§] Based on single trial.

Table 2.1 also shows the single turnover rate constants for cleavage of secondary site DNA (40-2A) by SgrAI. 40-2A has the sequence CR|CCGGGG, where the substitution of the primary site (i.e. CRCCGGYG) occurs in the degenerate pyrimidine position. The cleavage rate constant of 40-2A by SgrAI in the absence of added PC DNA is very slow (0.013±0.004 min^{-1}), and the percentage of DNA cleaved is only 17.7±0.5%. However, after adding 500 nM PC DNA to the reaction, the cleavage rate constant increases to 6.1±0.6 min^{-1}, and the total amount of DNA cleaved increases to nearly 80% (although only 50% is cleaved with the faster rate constant). At 2 μM PC DNA, SgrAI cleaves 40-2A DNA with a rate constant of 10.3±2.7 min^{-1}, representing a ~1000 fold increase from its unstimulated cleavage rate with this DNA. Although this is an impressive increase, the rate constant is half of that achieved for the primary site embedded in the same construct (40-1AT) under the same conditions (Figure 2.3). This

shows that the longer flanking bp of 40-2A also aides with activation of cleavage of secondary sites, in contrast to results with 18-2A (Park *et al.* 2010b), and that the cleavage of secondary site sequences can be accelerated to rates approaching that of primary site DNA.

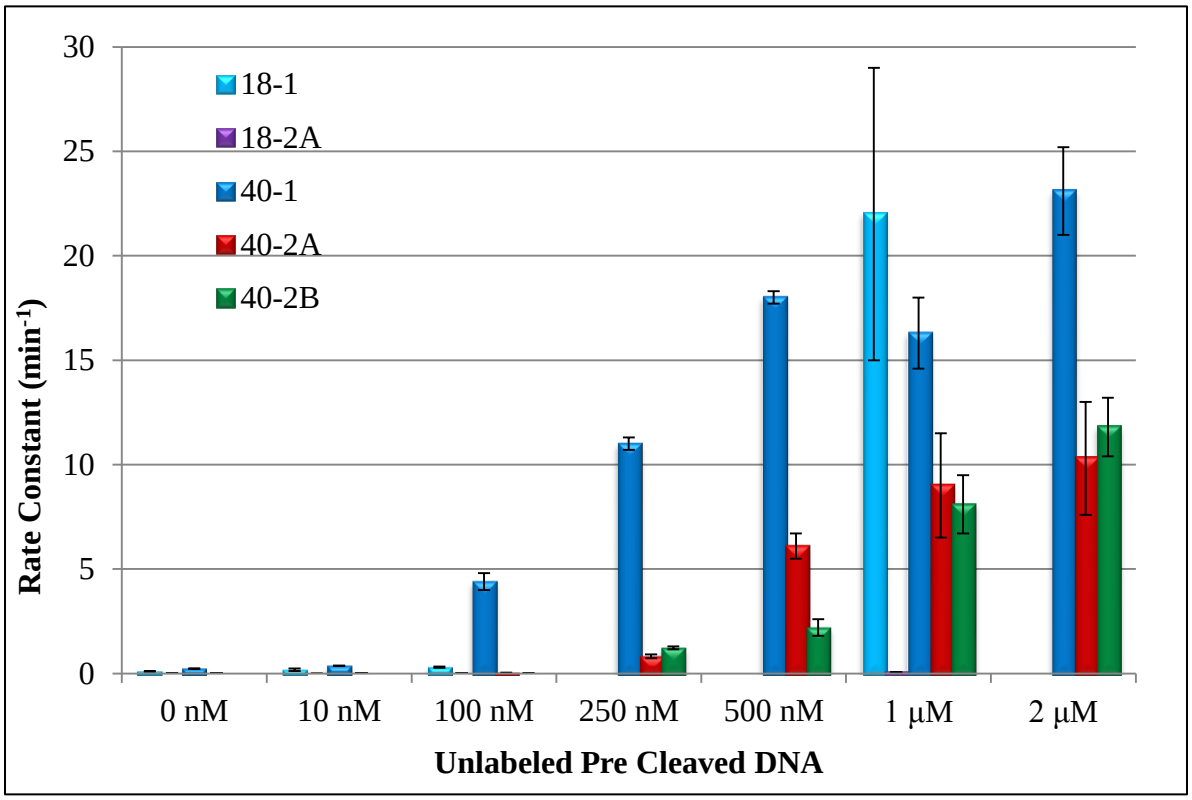


Figure 2.3: Single turnover DNA cleavage rate constants comparison of WT SgrAI with shorter (18-1AT, 18-2A) and longer flanking DNAs (40-1AT, 40-2A, 40-2B). (In the cases where two rate constants are derived, only the accelerated rate constant is plotted).

Single turnover DNA cleavage assays were also carried out using the second type of secondary site DNA, CACCGGTC, which contains a G to C substitution in the last position of the recognition sequence. This construct, 40-2B, contains the same flanking DNA as 40-1AT and 40-2A (Table 2.8). The accelerated cleavage rate constants of 1 nM 40-2B by 1 µM SgrAI in the presence of 1 and 2 µM PC were determined to be 8.1±1.4

min^{-1} and 11.8±1.4 min^{-1} respectively, which are very similar to those of 40-2A under similar conditions (Table 2.1, Figure 2.3). Interestingly, in the absence of unlabeled PC DNA, very little cleavage of the bottom strand of 40-2B is observed. The measured rate constants of 40-2B top or 40-2B bottom in the absence of PC DNA (Table 2.1) indicate that the top strand is cleaved ~5-fold faster than the bottom strand. However, in the presence of 1 μM PC, both strands are cleaved at similar rates, and at a rate constant similar to that of 40-2A (Table 2.1, Figure 2.3). Interestingly, the cleavage rate constants at intermediate concentrations of PC DNA differ, in that those of 40-2A are greater than those of 40-2B (Table 2.1, Figure 2.3). Compared to 40-2A, cleavage of 40-2B in the presence of 500 nM PC is ~3-fold slower (2.2±0.4 min^{-1} for 40-2B, 6.1±0.6 min^{-1} for 40-2A), indicating that 40-2A is stimulated to a greater extent than 40-2B at particular concentration of PC DNA (Figure 2.3).

Similar to that found in previous work (Park *et al.* 2010b), secondary site containing DNA is incapable of stimulating DNA cleavage to the same extent as primary site (Table 1). At 1 μM 40-2A, the cleavage rate constant of 1 nM radiolabeled 40-2A is only 0.25±0.06 min^{-1}, however, this rate constant is greatly increased when compared with 1 nM 40-2A (0.004±0.003 min^{-1}). In addition, the percentage of DNA cleaved is increased from 34±12% to 78±5%. The reactions were also performed with each strand labeled independently, to investigate more closely the possibility that one strand is cleaved faster than the other, however, with 0 nM PC the rate constants were similar (0.0010±0.0001 min^{-1} for top labeled and 0.0024±0.0018 min^{-1} for bottom labeled) and

with 1 μM added PC DNA the rate constants were also similar (9.1±1.0 min^{-1} and 6.2±1.1 min^{-1}, respectively, Table 2.1).

Apart from primary and secondary site sequences, the single turnover DNA cleavage rate constants of a non-cognate sequence, 18-NCCG, 18-NCCG, and 40-NCTA (Table 2.8) containing the sequence CCCCGGGG and CTCCGGAG, respectively, were also determined (Table 2.1). Single-turnover rate constants obtained on cleaving 18-NCTA in absence of PC (0.010±0.001 min^{-1}) or 18-NCCG in absence of PC (0.031±0.005 min^{-1}) and presence of 1 μM PC (0.021±0.004 min^{-1}) are very low (Table 2.1). Similarly in the absence of any added PC DNA, the rate constant for 40-NCTA (0.024±0.004 min^{-1}) was ~10-fold slower than that of primary site containing DNA (40-1AT) under similar conditions. However, unlike the primary and secondary site sequences, the presence of 1 μM PC results in a cleavage rate constant of this noncognate DNA of only 0.052±0.006 min^{-1}, which is ~300-fold slower than that of the primary site DNA, and only 2 fold increase when compared to one with no PC, showing lack of significant activation of SgrAI mediated cleavage. In addition, the amount of the noncognate DNA cleaved is low (<20%) in all reactions with this DNA.

Discussion

Effects of longer flanking DNA on the cleavage of primary and secondary site DNA (40-1 and 40-2)

Previous studies of SgrAI activation made use of a radiolabeled 18 bp DNA containing the primary site sequence (18-1AT, Park *et al.* 2010b) at low concentration (1 nM), excess SgrAI (1 μM), and varied concentrations of unlabeled activating DNA (PC

DNA, 0-1 μM). The activating DNA is a precleaved primary site DNA (DNA with the primary site sequence that simulates the products of cleavage by SgrAI, Table 2.8) with 16 bp flanking the primary sequence. The amount of DNA flanking the recognition sequence has been shown to be important to activation (Park *et al.* 2010b, Daniels *et al.* 2003). The 18 bp DNA, 18-1AT, contains only 5 bp of flanking DNA, which are insufficient to activate SgrAI (Park *et al.* 2010b). In the assay, SgrAI cleaves 18-1AT with a low, unstimulated rate constant (~0.1 min^{-1}). The addition of PC DNA stimulates SgrAI to cleave this DNA up to 200 fold faster (depending on the concentration of PC DNA, 22±7 min^{-1} for added unlabeled 1 μM PC). The assay conditions maintain SgrAI concentration at a high level (1 μM dimer), and therefore concentrations of PC DNA up to 2 μM (since SgrAI binds PC DNA 2:1 PC DNA:SgrAI dimer, Ma *et al.* 2013) result in increasing concentrations of PC DNA bound SgrAI (PC DNA binds SgrAI with nM affinity, Park *et al.* 2010b). The model for activated cleavage of DNA by SgrAI involves formation of a run-on oligomer, stabilized by contacts between SgrAI, as well as contacts between SgrAI and the flanking DNA of DNA bound to adjacent SgrAI dimers (Lyumkis *et al.* 2013) (Figure 2.1). With 18-1AT, it was found that 10 and 100 nM PC DNA stimulated SgrAI only very modestly (2 and 3 fold), and higher concentrations were required for full stimulation. In the case of 40-1AT, DNA with 16 flanking base pairs, significant stimulation occurs at 100 nM PC DNA. The added flanking DNA of 40-1AT allows for stimulation of cleavage by SgrAI at lower concentrations of PC DNA bound SgrAI, consistent with the flanking DNA stabilizing the activated, oligomeric form of SgrAI thus favoring the run-on oligomer formation at lower concentrations.

To test whether uncleaved DNA stimulates SgrAI as well as pre-cleaved DNA, 40-1AT, containing the same number of flanking base pairs as PC DNA, was used as the unlabeled stimulatory DNA. The cleavage rate constant of labeled 40-1AT upon addition of 1 μM unlabeled 40-1AT was found to be 12±1 min^{-1}, a ~43-fold increase over the basal cleavage rate of 1 nM 40-1AT. However, recent mass spectrometry studies (Ma *et al.* 2013) have shown that SgrAI binds PC DNA with 2:1 stoichiometry (duplex DNA: SgrAI dimer), therefore the proper comparison would be to 2 μM PC DNA (23±2 min^{-1}). These results suggest that precleaved DNA does in fact stimulate SgrAI to a greater degree than uncleaved. The origin of this phenomenon awaits high resolution structure determination of the activated form of SgrAI.

The cleavage of secondary site DNA by SgrAI was revisited using oligonucleotides containing long flanking DNA, analogous to the flanking DNA in the activating DNA (i.e. PC DNA or 40-1AT). Two different types of secondary sites were studied containing substitutions in either the second to last base pair (40-2A, CR|CCGG<u>G</u>G) or the outer base pair (40-2B, CR|CCGGY<u>C</u>). For 40-2A, in the absence of PC, the cleavage rate constant was 0.013±0.004 min^{-1}, whereas in the presence of 1 μM PC the rate constant increased to 9.0±2.5 min^{-1}. SgrAI cleaved a very low percentage (17.7±0.5%) of 40-2A, but the percentage increased to 72±7% upon addition of 1 μM PC. This shows that the addition of activating DNA enables faster secondary site cleavage as well as more total secondary site cleavage. A drastic increase in the cleavage rate constant of 40-2A occurs at 500 nM added PC, whereas for 40-1AT it occurs at 100 nM PC. Therefore a higher concentration of PC is needed for accelerated cleavage of 40-2A,

suggestive of a diminished affinity of SgrAI bound to 40-2A, or for the 40-2A bound SgrAI to form the run-on oligomer. In the presence of 2 μM PC, 40-1AT was cleaved more than twice as fast as 40-2A. This shows that even at high levels of activation, SgrAI preferably cleaves primary site faster than the secondary site. The cleavage rate constant determined previously for SgrAI and an 18 bp secondary site containing DNA (Park *et al.* 2010b) in the absence of PC (0.020±0.006 min^{-1}) or presence of 10 nM PC (0.012±0.001 min^{-1}) are comparable to those determined here with 40-2A: 0.013±0.004 min^{-1} (Table 2.1) in the absence of PC and 0.018±0.004 min^{-1} in presence of 10 nM PC. However, the previous study reports only a 2.5-fold increase in stimulation of DNA cleavage in the presence of 1 μM PC (0.05±0.01 min^{-1}, Park *et al.* 2010b). With the secondary site now embedded in a 40 bp DNA, a ~1000-fold increase in stimulation of DNA cleavage was found (16.9±4.3 min^{-1}, Table 2.1). This could be due in part to the longer length of DNA used, since the flanking DNA forms important contacts in the activated oligomer (Lyumkis *et al*. 2013). The flanking DNA of 40-2A might increase the affinity of the SgrAI/DNA complex for the oligomer, shifting the equilibrium towards the activated conformation.

Similar to previously reported results (Park *et al*. 2010b), when unlabeled 40-2A was used as a stimulator, the cleavage rate constant of 40-2A by SgrAI was very low, even when 1 μM 40-2A was added (Table 2.1). The single-turnover cleavage rate constant using top only or bottom only labeled strands of 40-2A (40-2A-Top and 40-2A-Bottom, respectively, Table 2.1) in the presence of either 0 nM or 1 μM PC DNA clearly

shows that there is no significant preference of either top or bottom stands by activated SgrAI.

The present results compare very well with the previously determined cleavage rate constants (Park *et al.* 2010b, Hingorani-Varma *et al.* 2003). The cleavage rate constant of SgrAI in the absence of activating DNA has been measured to be 0.094±0.015 min^{-1} for 18-1AT (Park *et al.* 2010b), 0.14±0.05 min^{-1} (Park et al. 2010b), and 0.236±0.006 min^{-1} (this work) for 40-1AT, and 2.4±0.18 min^{-1} for an 80 bp primary site containing DNA (80-1) (Hingorani-Varma *et al.* 2003). Similarly, in the presence of activating DNA (~1 μM) the cleavage rate constants was measured to be 22±7 min^{-1} for 18-1AT (Park *et al.* 2010b), >20 min^{-1} (Park *et al.* 2010b) and 16.3±1.7 min^{-1} (this work) for 40-1AT, 16.2±1.8 min^{-1} for 80-1 (Hingorani-Varma *et al.* 2003). The cleavage rate constant of SgrAI in the absence of activating DNA for secondary site has been found to be 0.020±0.006 min^{-1} when embedded in an 18 bp construct (18-2A) (Park *et al.* 2010b), 0.013±0.004 min^{-1} for when embedded in a 40 bp construct (40-2A) (this work), and 0.018±0.003 min^{-1} for a secondary site embedded in an 80 bp DNA (80-2) (Hingorani-Varma *et al.* 2003). Similarly, in the presence of activating DNA the cleavage rate constants were measured as 0.05±0.01 min^{-1} for 18-2A (Park *et al.* 2010b), 9.0±2.5 min^{-1} for 40-2A (this work), and 3.48±1.2 min^{-1} for 80-2 (Hingorani-Varma *et al.* 2003). The only major difference found was in comparing the activated cleavage of secondary site DNA when embedded in either an 18 bp or a 40 bp DNA. Cleavage of this site in the 40 bp DNA is up to 1000 fold accelerated in the assays, whereas that in the 18 bp DNA is only 2.5 fold accelerated, relative to cleavage rates without activation. This result

supports the idea that the longer flanking DNA is essential for cleavage of secondary sites to be activated. The flanking DNA may aid in binding of the DNA to SgrAI and/or in binding of the SgrAI/DNA complex to the activated, run-on oligomer of primary site bound SgrAI.

Asymmetric cleavage

The cleavage studies of 40-2B show different cleavage rates for the two strands, but only without significant activation of SgrAI. In the absence or at low concentration of PC DNA, the cleavage of the bottom strand of 40-2B is significantly slower than that of the top (Table 2.1). This was determined using versions of 40-2B labeled at only the top or the bottom strands. Examination of the sequence of the top and bottom strands of 40-2B suggests why the top strand is cleaved faster. The top strand (5′-CA|CCGGTC-3′, underline indicates bp substitution) has the cognate sequence C at the first position, which is closer to the active site (marked by |); however, the bottom strand (5′-GA|CCGGTG-3′, underline indicates bp substitution) has the secondary site substitution of G at this position. The slower cleavage may be a result of misalignments in or near the active site as a result of the substitution. The strain in the protein-DNA complex is apparently relaxed in the activated complex, as top and bottom strands are cleaved with similarly fast cleavage rate constants (in the absence of PC DNA or presence of 1 μM PC DNA, Table 2.1). This may indicate that interactions with the outer base pair are lost in the activated state, and therefore misalignments due to substitutions at this base pair no longer impact the active site conformation.

Single-turnover rate constants obtained on cleaving 18-NCTA in absence of PC (0.010 ± 0.001 min^{-1}) or 18-NCCG in absence of PC (0.031 ± 0.005 min^{-1}) and presence of 1 μM PC (0.021 ± 0.004 min^{-1}) are very low (Table 2.1). Similarly, this is true with longer flanking DNA as well. Single-turnover rate constants obtained on cleaving 40-NCTA by SgrAI either in the presence or absence of PC DNA were very low ($2.3\pm0.2x10^{-3}$ min^{-1} and $4.8\pm0.8x10^{-3}$ min^{-1}, Table 2.1). The percentage of DNA cleaved was also very low (7.5 ± 0.2 and 11.7 ± 0.6% respectively, Table 2.1) indicating the lack of preference of SgrAI for the non-cognate sequence. The rates compare well to that of the 40-2B bottom strand cleavage in the absence of PC DNA, however unlike 40-2B, the cleavage of the noncognate site is not activated with increasing concentrations of PC DNA (Table 2.1). This is interesting, since 40-NCTA contains substitutions in both degenerate positions (i.e. 2 and 7), a symmetric version of the substitutions found in 40-2A. Hence, if specificity expansion is merely the loss of contacts to the outer two base pairs of the recognition sequence, cleavage of this sequence should similarly be accelerated by the addition of PC DNA. However, such is not the case, and therefore some additional factor prevents activated cleavage of 40-NCTA, which may be diminished binding affinity of SgrAI for this sequence, or some other blocking of the activated conformation by those bp substitutions.

Part B: Structural study of run-on oligomer

Introduction

There are several functional and non-functional proteins associated with run-on oligomers. Protein aggregation is normally related to formation of non-functional aggregate which can lead to various deposition diseases such as Alzheimer's, type II diabetes, and Huntington's disease. In general, proteins in amyloid have been shown to oligomerize by cross-β spine, end-to-end stacking, and/or 3D domain swapping. 3D domain swapping could lead to formation of homodimers, as well as higher order run-on oligomers due to exchange of protein domains (Bennett *et al.* 2006). Run-on oligomerization has been seen in RecA (Story *et al.* 1992, Xing *et al.* 2004), T7 helicase (Sawaya *et al.* 1999), α1-antitrypsin (Huntington *et al.* 1999) and "cab"-type carbonic anhydrase (Strop *et al.* 2001). SgrAI also exhibits run-on oligomerization, but which is distinct from that of amyloid in being reversible, as well as in modulating the enzymatic activity of SgrAI (Park *et al.* 2010a, Lyumkis *et al.* 2013). The run-on oligomers of SgrAI show the unusual phenomena of substrate specificity expansion (Bitinaite *et al.* 2006), stimulation by reaction products (Park *et al.* 2010b) and possibly reversible domain swapping (Park *et al.* 2010a). To date, SgrAI is the only enzyme possessing all of these characteristics. Thus, the SgrAI system presents a unique opportunity to study a new paradigm of enzyme regulation. This section focuses on answering some of the questions related to activation of SgrAI by investigation of the structure of its activated, oligomeric form.

Involvement of domain swapping in run-on oligomer formation

One of the previously reported crystal structures of SgrAI bound to primary site DNA shows two DBD interacting via domain swapping (Figure 2.4B). The resolution of this crystal structure is 2.03 Å (Park *et al.* 2010a), and the conformation of SgrAI was found to be the same global conformation as described previously (Dunten *et al.* 2008) (Figure 2.4A), and argued as the low activity conformation. The swapping domain consists of 24 amino acid residues which are swapped with another subunit of a neighboring dimer (Figure 2.4B). Swapping occurs in all eight copies of the SgrAI chain in the asymmetric unit. Residues 25-30 of each subunit comprise a linker or hinge loop which have different conformations in the swapped and unswapped structures. In the swapped conformation, SgrAI arginine 29 makes a hydrogen bond to the main chain oxygen of isoleucine 23 of another swapping linker (Figure 2.5). In order to test if domain swapping is involved in run-on oligomer formation and activation of SgrAI, the R29A substitution was designed to disrupt the hydrogen bonding interaction that appears to stabilize the swapping linkers, and thereby destabilize the domain swapped conformation. Single turn-over DNA cleavage assays, DNA binding affinity measurements, and structure analysis using crystallography were carried out to determine the effect of the mutation on oligomerization, activation of DNA cleavage, and the structure of SgrAI bound to DNA.

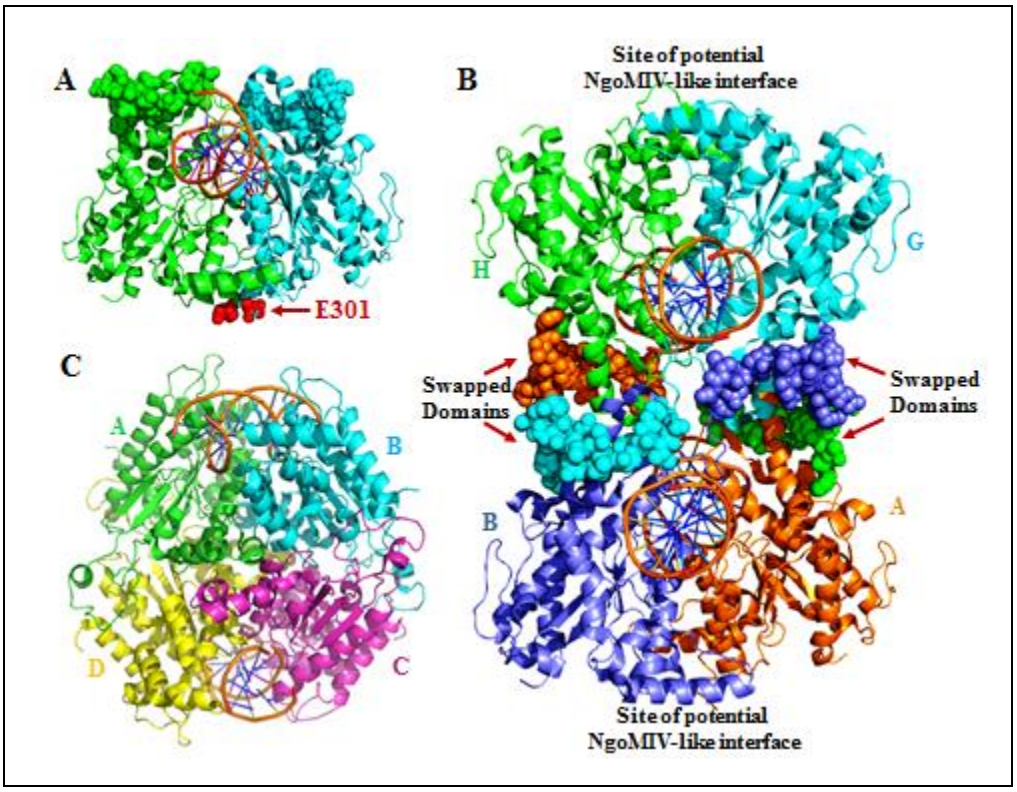


Figure 2.4: Comparison of tetrameric forms of SgrAI and NgoMIV. (A) Dimeric structure of SgrAI bound to primary site DNA (PDB code: 3DVO, adapted from Dunten *et al*. 2008). Residues of the swapping domain (1–24) shown in space filling spheres (green and cyan). Residues 25–339 shown as ribbons, and the bound DNA (brown) shown as cartoon. Red spheres show the location of residue E301. (B) Tetrameric structure of SgrAI (PDB code: 3MQ6, adapted from Park *et al*. 2010a) with subunits A, B, G, and H labeled and colored in orange, slate, cyan and green, respectively. Each subunit swaps the amino-terminal 24 amino acid residues (shown as space filling spheres) with those of a subunit in an opposing dimer. Residues 25–339 shown as ribbons, and the bound DNA (brown) shown as cartoon. (C) Ribbon diagram of NgoMIV (PDB code: 1FIU, adapted from Deibert *et al*. 2000) (subunits in green, cyan, magenta and gold) bound to DNA (brown). Adapted from Park *et al*. 2010a.

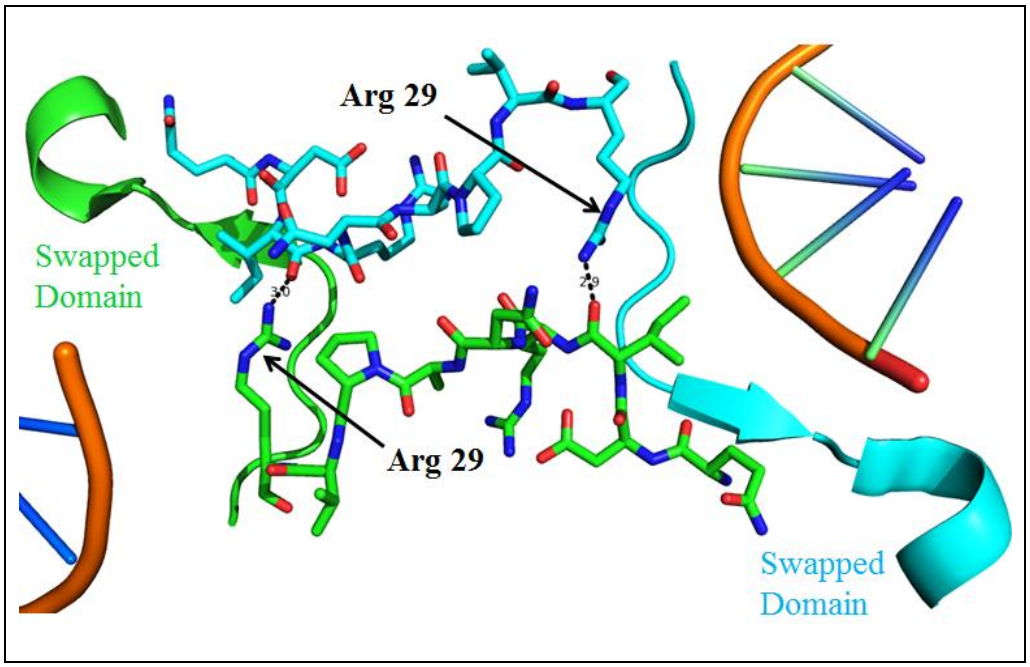


Figure 2.5: Swapped domains of SgrAI showing the location of Arg 29 in the linker region. Different domains are shown in different colors (green or cyan). Adapted from Park *et al.* 2010a.

Exploring the NgoMIV-like oligomeric interface in SgrAI

NgoMIV, a type IIF restriction endonuclease from *Neisseria gonorrhoeae* strain M, has 22 % sequence identity with SgrAI and cleaves a recognition sequence similar to that of SgrAI (NgoMIV cleaves G|CCGGC, SgrAI cleaves CR|CCGGYG). NgoMIV functions as a homotetramer, and can be considered a dimer of dimers. The crystal structure of NgoMIV shows that it has a similar core structure as that of other restriction endonucleases. The central core region consists of a six-stranded β–sheet with α–helices on each side (Deibert, *et al*. 2000, Aggarwal, 1995) (Figure 2.4C). NgoMIV forms a tetramer in a back-to-back or tail-to-tail orientation, resulting in the placement of the DNA binding sites of the primary dimer at opposite ends of the tetramer (Figure 2.4C.

The loop region between β1 and β2 strands protrudes 25 Å out of the core of each monomer, and is mainly involved in dimer and tetramer formation. The tetrameric structure of NgoMIV is fixed by side-by-side contacts (between A/D and B/C subunits) and cross contacts (between A & C and B & D subunits) between primary dimers. The cross contacts are mainly responsible for the dimerization interface while the loops (residues 147-176, including helix α6) between A/D and B/C are primarily responsible for tetramerization (Deibert *et al*. 2000, Pingoud *et al.* 2001).

Based on native gel shift assays and analytical ultracentrifugation (AU) assays it has been shown that SgrAI forms a run on oligomer giving rise to high molecular weight species in the presence of sufficient concentrations of SgrAI and its primary site DNA sequence (Park *et al.* 2010a). AU indicated that the run-on oligomer was heterogeneous and was composed of 4-12 DNA-bound SgrAI dimers. Prior to the determination of the cryo-EM structure, it was not known how the dimers were oriented to form the large oligomers. Since SgrAI shares many similarities with NgoMIV, it was considered possible that SgrAI may utilize a similar interface as that of NgoMIV to bind two dimers together in a tail-to-tail fashion (Figure 2.4C). Structure alignment of SgrAI and NgoMIV (Figure 2.6, RMSD 1.18) shows that glutamate 301 lies at the tail side of SgrAI. If SgrAI forms tetramer in NgoMIV-like fashion, mutation of glutamate 301 to tryptophan should disrupt the interface and thus, prevent oligomerization and activation of SgrAI. Therefore, the E301W substitution was constructed and single turnover cleavage rate constants of the mutant enzyme was compared with that of WT SgrAI.

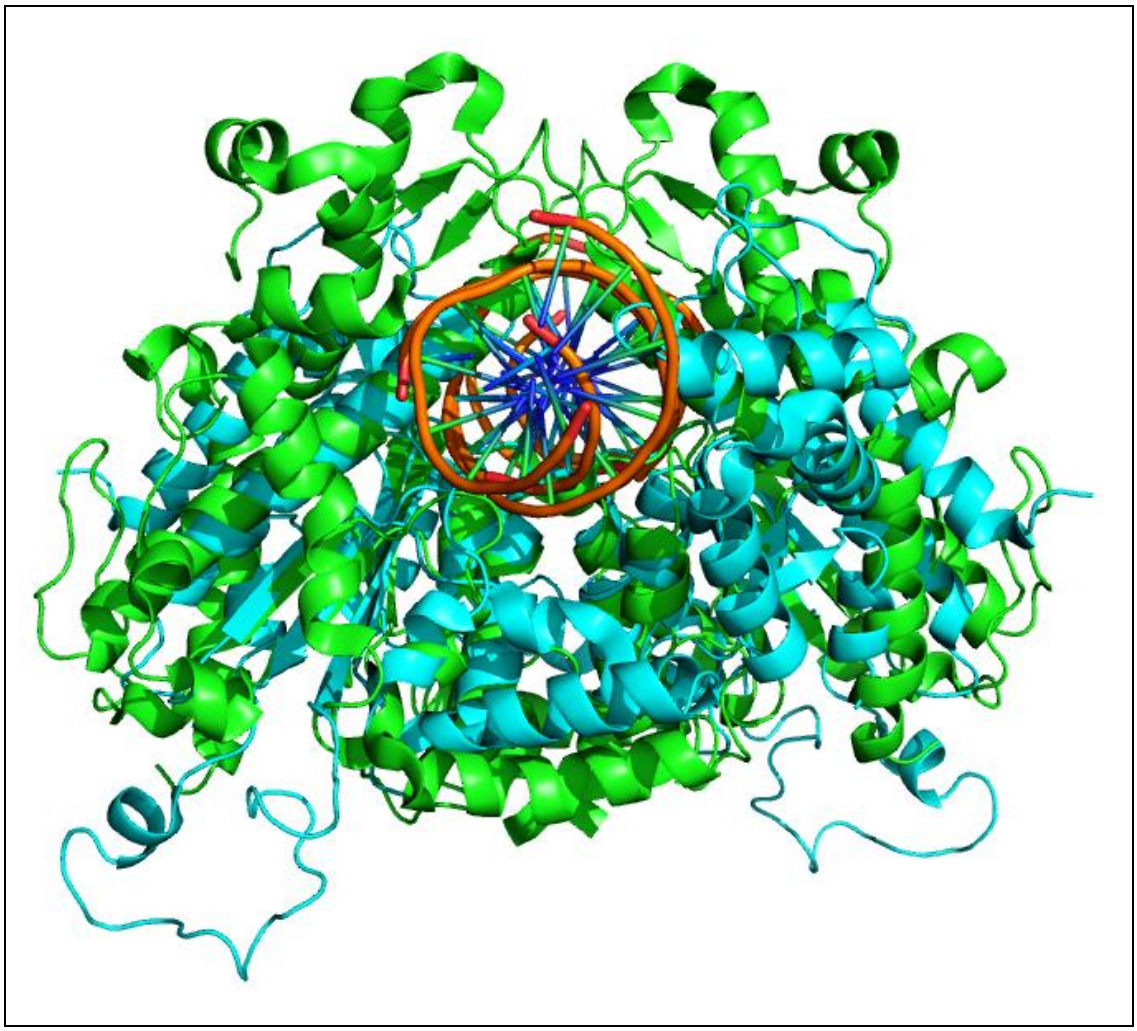


Figure 2.6: Alignment of crystal structure of SgrAI (green) and NgoMIV (cyan). Loops that form tetramer contacts in NgoMIV (cyan loops hanging down in the bottom) are folded upwards in SgrAI. RMSD = 1.18.

Structural clues on the role of flanking DNA in the stability of oligomeric form of SgrAI

The 8.6 Å cryo-EM structure of SgrAI bound to PC DNA (Lyumkis *et al.* 2013) suggests that the flanking DNA may play a vital role in stabilizing the run-on oligomer structure. The negatively stained EM images show that SgrAI forms rod shaped structures of varying lengths with a regular repeating structure. The run-on oligomer was

found to consist of 1 through 18 DNA bound SgrAI dimers (DBD), consistent with the analytical ultracentrifugation data (12 DBDs, Park *et al.* 2010b). The DBDs are arranged in a left handed helix with roughly 4 DBD per turn of the helix (Figure 2.1). The flanking DNA appears to make contacts to neighboring SgrAI at or near the loops 56-60 and 122-140. In addition to these contacts, the SgrAI proteins contact each other using their N-terminal domains near the central helical axis. At this resolution, the existence of domain swapping as seen in the previous crystal structure cannot be discerned, but the adjacent locations of the N-termini make it a possibility. Upon fitting the three dimensional structure of SgrAI (in the low activity conformation) into the cryo-EM map, a 10º rotation of one subunit relative to the other was found, which according to our model is the conformational change which activates its DNA cleavage activity. The rotation of subunits might align the active site better to achieve the activation of DNA cleavage seen by the run-on oligomer (Lyumkis *et al.* 2013), or stabilize a conformation which moves a metal ion binding site closer to the DNA (Dunten *et al*. 2008). However, since the cryo-EM structure is not of high resolution, it is not possible to be certain about individual side chain atoms and nucleotides involved in hydrogen bonding and other contacts.

The unusual enzyme mechanism shown by SgrAI raises several fundamental questions: how is DNA cleavage activated? How is the sequence specificity of SgrAI relaxed in the activated form? Does flanking DNA stabilize the oligomeric form? In order to answer these questions, we carried out crystallography trials with longer flanking DNA known to activate SgrAI, to obtain a higher resolution x-ray crystal structure of activated run-on oligomer of SgrAI. A new crystal structure with the shortest DNA known to

activate SgrAI, 22-1HT, was determined and showed several features similar to the cryo-EM model of activated, oligomeric SgrAI (Lyumkis *et al*. 2013). However, the enzyme conformation was found to be identical, within coordinate error, to that of the low activity, dimeric form.

Results

The effect of R29A on activation of SgrAI

The expression clone for R29A SgrAI was prepared by Dr. Jurate Bitinaite of New England Biolabs. Purification was performed as described (Methods). Initial steady state kinetic methods performed by Dr. Bitinaite suggested that the mutation inhibited secondary site DNA cleavage by SgrAI. To better understand the origin of this effect, the dissociation constant (Table 2.2) of R29A SgrAI and18 bp secondary site containing DNA (18-2A) was measured, and found to be very similar to that of WT SgrAI to this DNA (5.2±2.7 nM for R29A SgrAI compared to 4.4±3.6 nM for WT SgrAI, Table 2.2). Next, the single turnover DNA cleavage rate constants were measured for R29A SgrAI and primary and secondary site DNA (Table 2.3-2.4, Figure 2.7-2.8) and found to be very similar to those of WT SgrAI as well. The only large difference is likely an artifact of the way the assay was conducted with that measurement; in the presence of 1000 nM PC, the single turnover DNA cleavage rate constant is 22±7 min^{-1} with WT SgrAI, whereas with R29A was found to be greater than or equal to 3.14±1.30 min^{-1} (Table 2.3). Because full cleavage of the DNA was found at the initial time points in the assay with R29A SgrAI, the exact rate constant could not be measured and only a lower limit determined. In the case of cleavage of secondary site DNA, 18-2A, the single turnover DNA cleavage rate

constants were also very similar for WT and R29A SgrAI at all PC DNA concentrations tested. The data shown in Tables 2.3 and 2.4 was measured by Amanda Stiteler, who also determined rate constants for the wild type enzyme. These values for the wild type enzyme differ to some degree from those of (Park, *et al*., 2010b), however, they do show activated DNA cleavage by R29A SgrAI, indicating that the substitution did not inhibit activation of SgrAI. One difference that was found as a result of the R29A substitution was in the solubility of the enzyme (*Prof. Nancy Horton, personal communication*). The enzyme was found to precipitate when incubated at 37ºC in the reaction buffer used in the initial steady state measurements. This precipitation likely gave rise to the lack of observed secondary site cleavage, as cleavage of primary sites is fast and occurs before significant precipitation, and therefore loss of the enzyme.

Dissociation constants (measured by Prof. Nancy C. Horton using fluorescence anisotropy)

Table 2.2: Dissociation Constant of WT and R29A SgrAI for 18 bp secondary site containing DNA (18-2A)	
Enzyme	**K_D**
WT	4.4±3.6 nM
R29A	5.2±2.7 nM

Single turn-over rate constants (measured by Amanda Stiteler)

Table 2.3: Single-Turnover DNA Cleavage Rate Constants using 1 µM Enzyme, 1 nM ^{32}P-labeled 18 bp primary site (18-1AT) DNA at 37°C		
	WT SgrAI*	**R29A SgrAI**
Added Unlabeled DNA	**Rate Constant (min^{-1})**	**Rate Constant (min^{-1})**
0 nM PC	0.094±0.015	0.13±0.05
10 nM PC	0.18±0.06	0.22±0.11
100 nM PC	0.30±0.03	0.79±0.43
1000 nM PC	22±7	> 3.14±1.30

***** Data from Park *et al.* 2010b

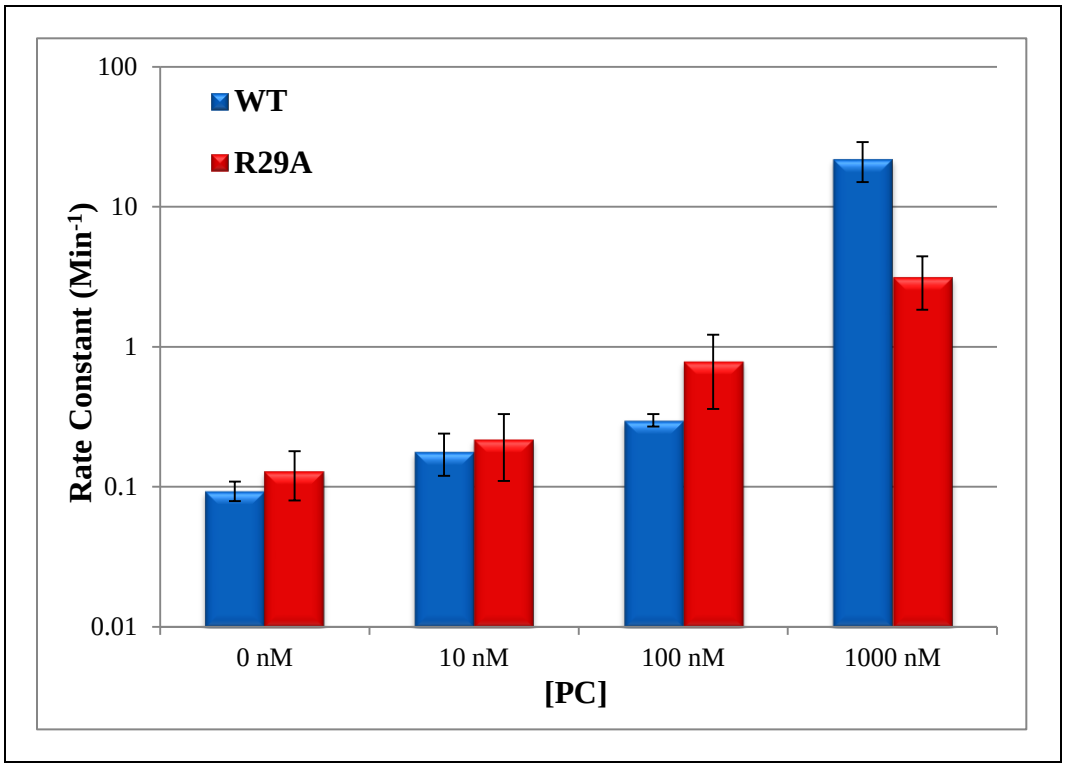


Figure 2.7: Comparison of single-turnover DNA cleavage rate constants of 1 µM WT and R29A SgrAI using 1 nM ^{32}P-labeled 18-1AT DNA at 37°C.

Table 2.4: Single-Turnover DNA Cleavage Rate Constants using 1 µM Enzyme, 1 nM ^{32}P-labeled 18 bp secondary site (18-2A) DNA at 37°C

	WT SgrAI	R29A SgrAI
Added Unlabeled DNA	**Rate Constant (min^{-1})**	**Rate Constant (min^{-1})**
0 nM PC	0.022±0.019	0.025±0.011
10 nM PC	0.015±0.010	0.053±0.066
100 nM PC	0.031±0.007	0.057±0.044
1000 nM PC	0.206±0.187	0.173±0.060

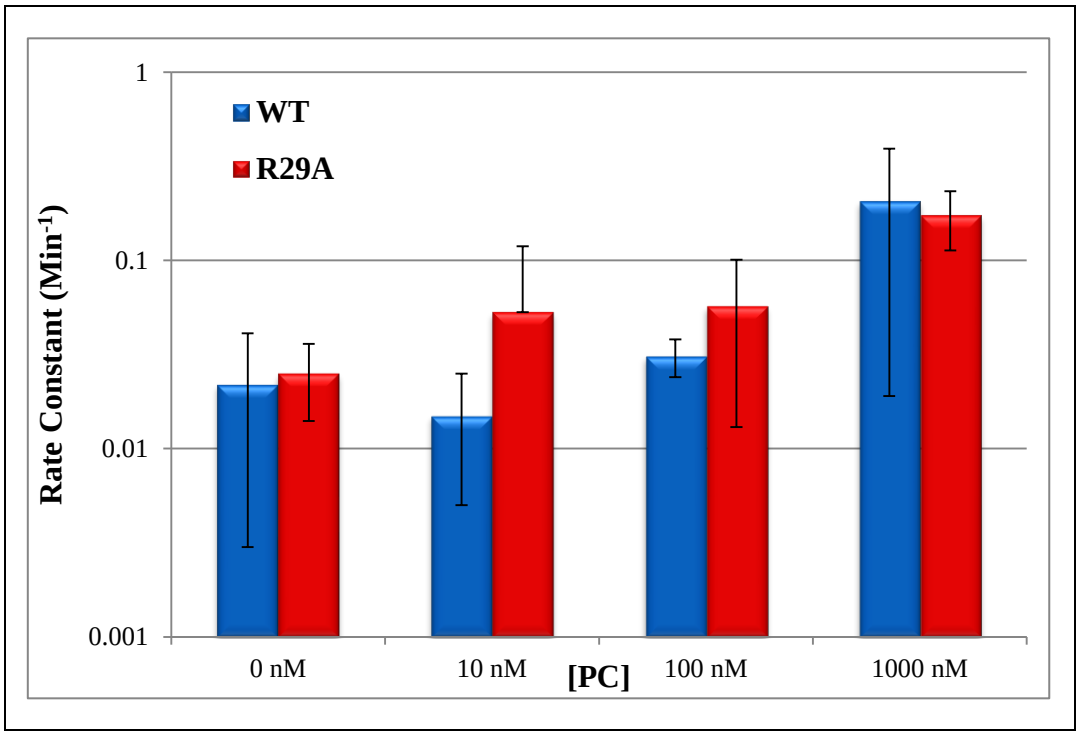


Figure 2.8: Comparison of single-turnover DNA cleavage rate constants of 1 µM WT and R29A SgrAI using 1 nM ^{32}P-labeled 18-2A DNA at 37°C.

Crystal structure (Data collection and refinement by Prof. Nancy C. Horton)

The 1.65 Å crystal structure of R29A SgrAI bound to 18-1AT DNA was determined (Table 2.5, Figure 2.9-2.11). The conformation of R29A SgrAI was found to be the same as that of WT SgrAI (RMSD = 0.351), which has been argued to be the low activity conformation, and without domain swapping. In the R29A SgrAI structure, the C_{beta} atom of alanine 29 is 3.8 Å away from both O_{delta} of asparagine 30 and O2P of Guanine3 in the DNA. The DNA seems to be slightly distorted in the mutant structure compared to the WT (Figure 2.11).

Table 2.5: Diffraction Data and Structure Refinement Statistics for R29A SgrAI bound to primary site DNA (18-1AT)	
Code	S451_R29A
Beamline	SSRL 9-2
Processing Program	MOSFLM/SCALA
DNA	18-1AT
Space Group	$P2_1$
Cell	59.14 Å, 117.38 Å, 64.00 Å, 108.58º
Resolution	1.65 Å
Total Observations	324,269
% Complete	95.4 %
I/Sigma	9.3
Multiplicity	3.4
R_{merge}[1]	6.5 %
R_{cryst}[2]	21.33 %
R_{free}[3]	24.10 %
Overall B factor (Wilson Plot)	15.74 $Å^2$
RMSD-bonds	0.005 Å
RMSD-angles	1.081º
Asymmetric unit	2 SgrAI chains and 2 DNA strands
Number of waters	736
Number of divalent cations	4 Ca^{2+}

[1] $R_{merge}= \sum_{hkl}(|<I_{hkl}>-I_{hkl}|)/(\sum_{hkl} I_{hkl})$ where $<I_{hkl}>$ is the average intensity over symmetry related and equivalent reflections and I_{hkl} is the observed intensity for reflection hkl.

[2] $R_{cryst}=\sum_{hkl}(||F_{obs}|-|F_{calc}||)/(\sum_{hkl}|F_{obs}|)$ where $|F_{obs}|$ and $|F_{calc}|$ are the observed and calculated structure factor amplitude for reflection hkl. The sum is carried out over the 98% of the observed reflections which are used in refinement.

[3] R_{free} refers to the R factor for the test reflection set (2% of the total observed) which was excluded from refinement.

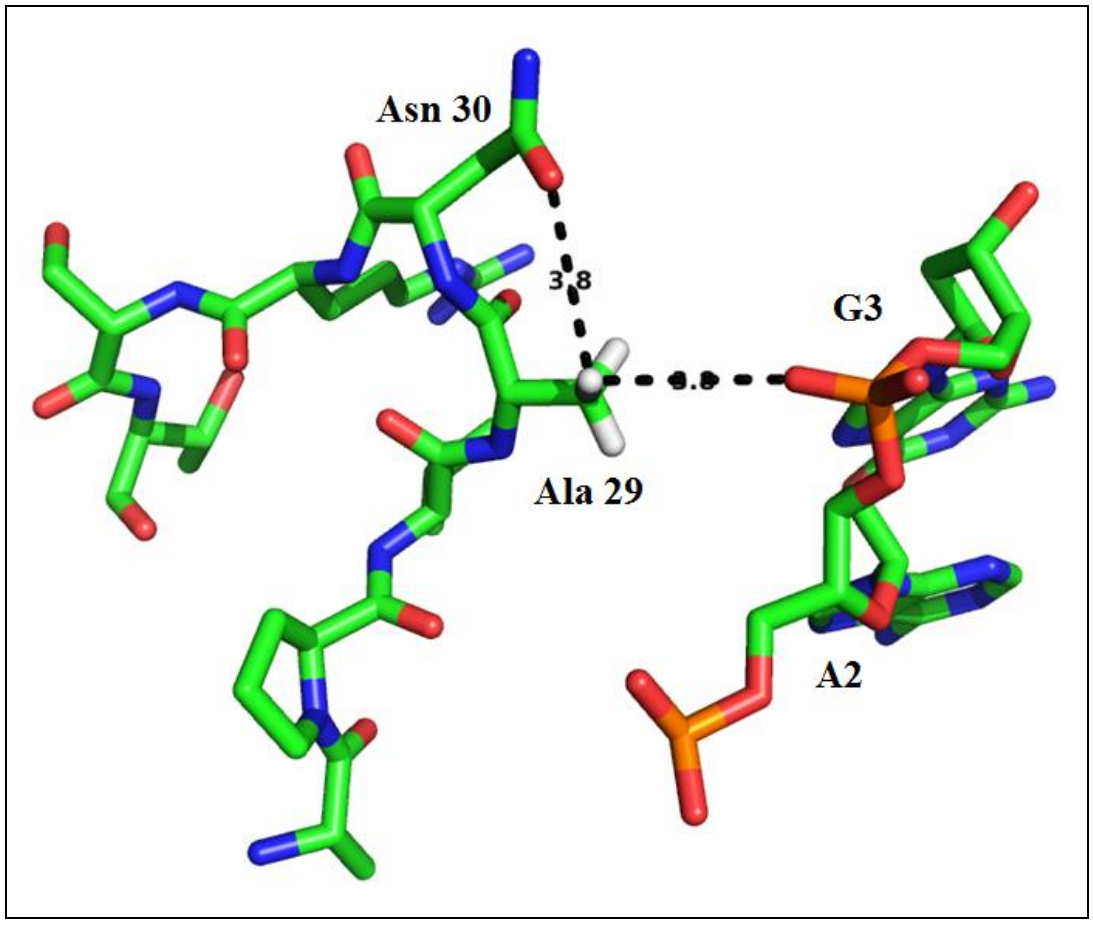


Figure 2.9: Stick diagram showing distances of closest neighbors of alanine 29 in the R29A SgrAI mutant.

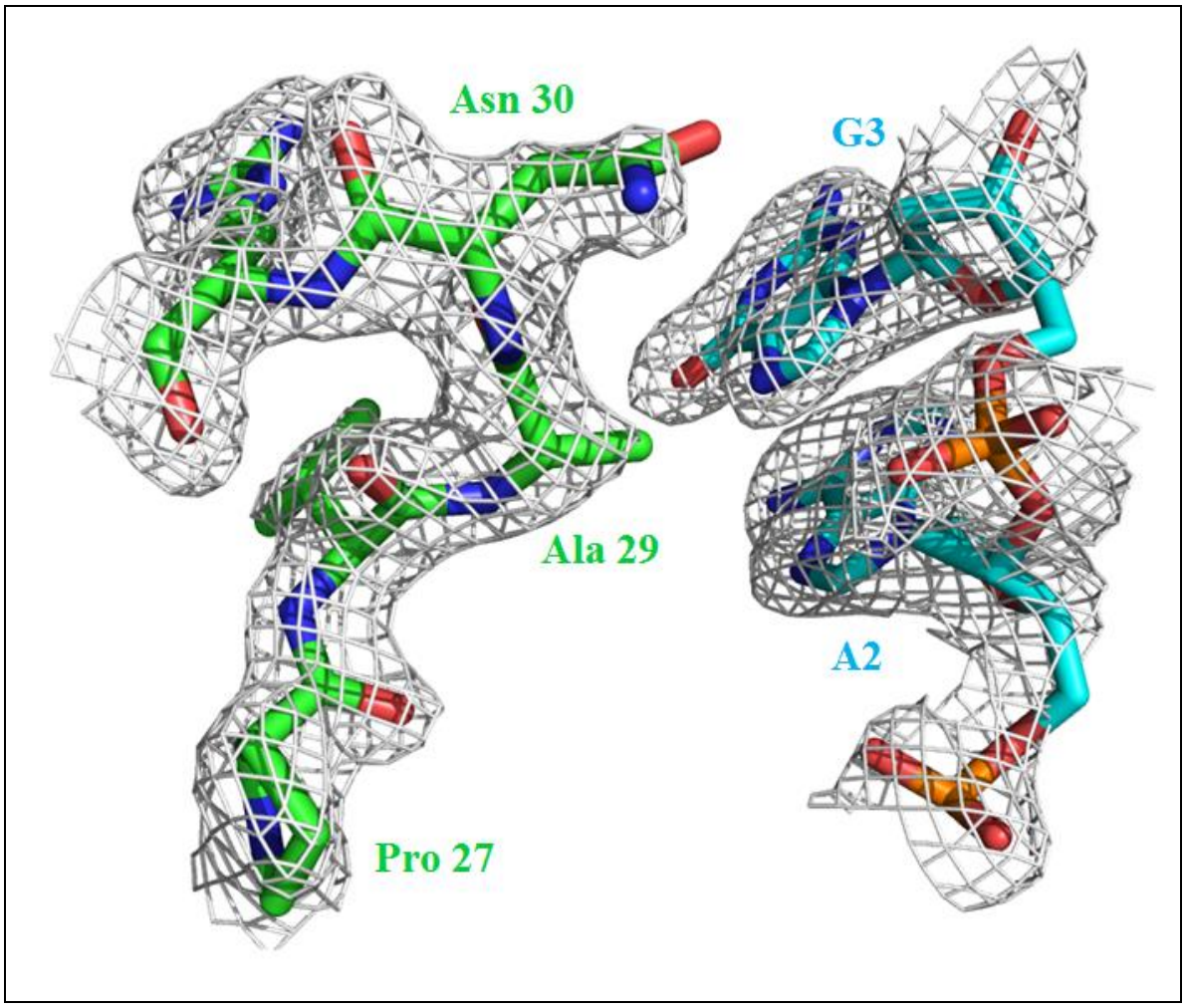


Figure 2.10: Electron density map of R29A SgrAI (green) showing the absence of electron density around the mutated arginine side chain. Electron density contoured at 1.0 sigma. DNA is shown in cyan.

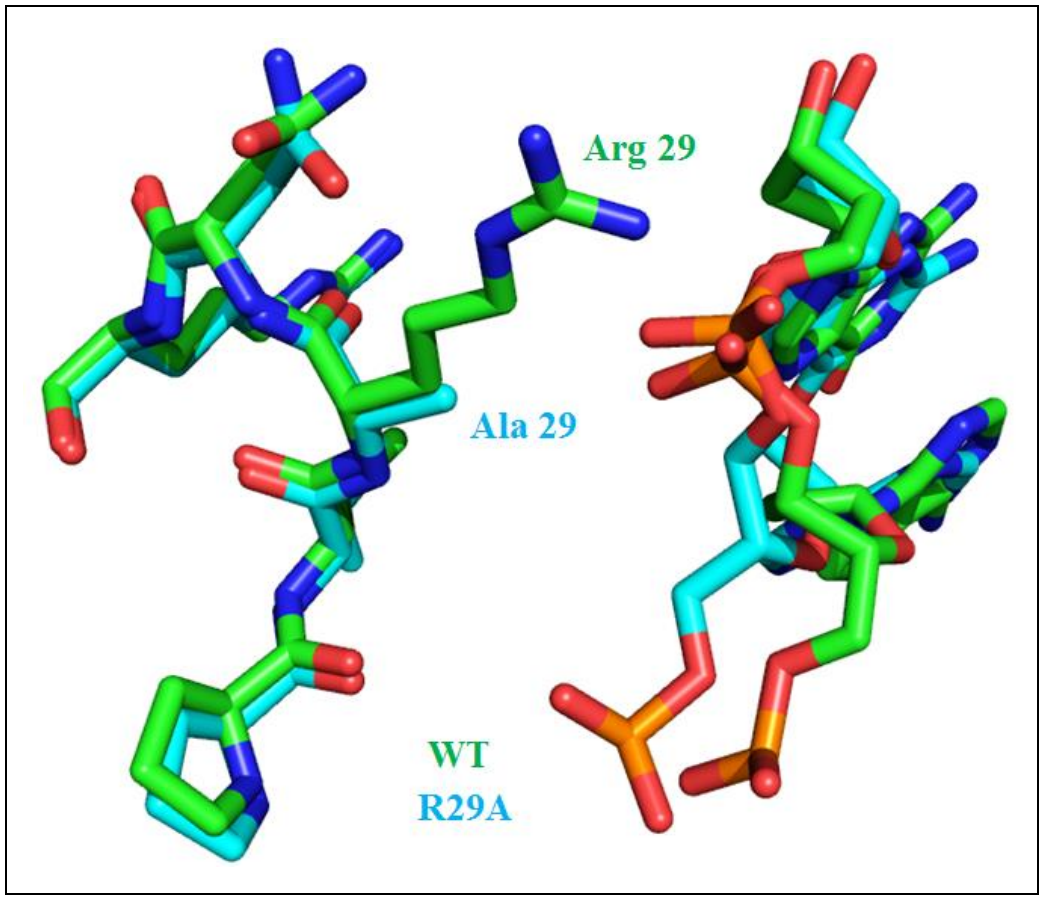


Figure 2.11: Stick diagram showing alignment of WT and R29A SgrAI. RMSD = 0.351.

Effect of E301W on SgrAI oligomerization and DNA cleavage

The E301W mutation was prepared by Dr. Alka Agrawal of New England Biolabs. E301W SgrAI was purified using FPLC (GE Healthcare Biosciences) chromatography and the following chromatographic resins: Heparin FF Sepharose (GE Healthcare Biosciences), SP FF Sepharose (GE Healthcare Biosciences), Q FF Sepharose (GE Healthcare Biosciences), followed by a second Heparin FF Sepharose (GE Healthcare Biosciences) chromatographic step. An SDS-PAGE purity gel of SgrAI E301W mutant is shown in Figure 2.12. The E301W mutation is designed to disrupt any potential NgoMIV-like tetramer interface. The single turn-over DNA cleavage rate

constant of E301W SgrAI and 40 bp primary site containing DNA was determined (Table 2.6) at different concentrations of PC DNA and compared to those of WT SgrAI (Figure 2.13). 1 μM E301W SgrAI cleaved labeled 1 nM 40-1AT DNA with a cleavage rate constant of 0.251±0.008 min^{-1}, which is very similar to that of WT SgrAI (0.251±0.003 min^{-1}) (Figure 2.13). Similarly, upon adding 1 and 2 μM PC the cleavage rate constants were enhanced to 17.2±2.4 and 16.6±0.7 min^{-1} respectively. Again, the rate constants were very similar to WT SgrAI (16.3±1.7 and 23.1±2.1 min^{-1} respectively) under similar conditions. The similarity in the single turnover DNA cleavage rate constants of E301W and wild type SgrAI at the different concentrations of activating PC DNA suggest that the NgoMIV-like interface is not formed in the activated form of SgrAI. This result also supports the cryo-EM structure for the oligomeric SgrAI which does not show an NgoMIV-like interface.

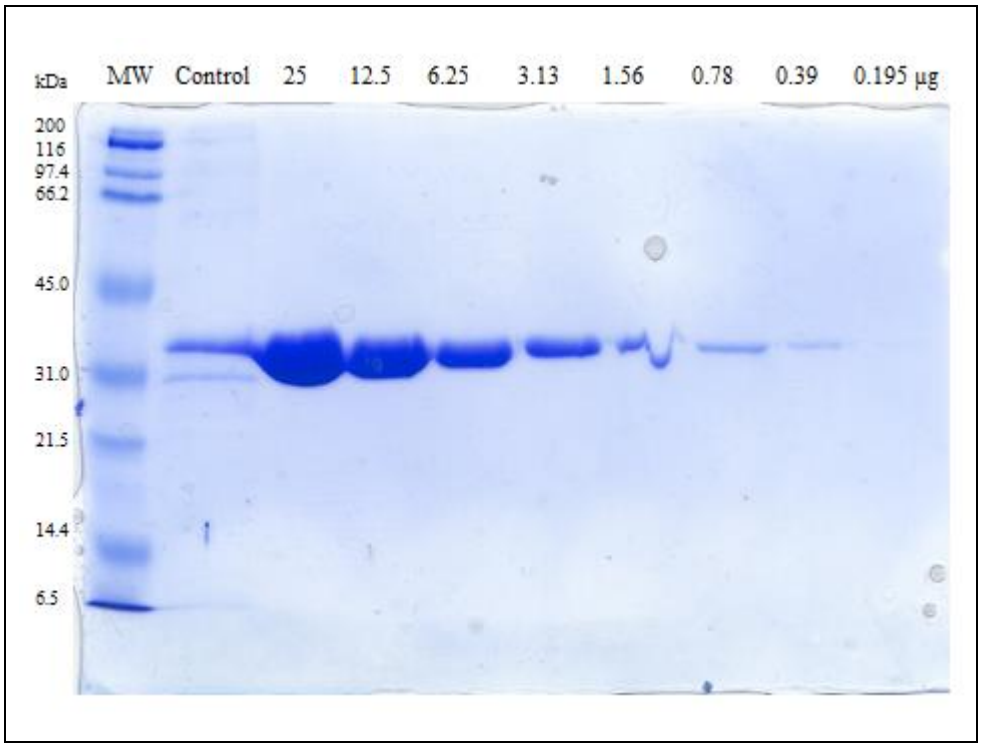


Figure 2.12: SDS-PAGE purity gel of SgrAI E301W mutant with decreasing amount of protein (25-0.195 µg). Previously purified WT SgrAI was used as the control.

Table 2.6: Single-Turnover DNA Cleavage Rate Constants using 1 µM E301W SgrAI and 1 nM ^{32}P-labeled 40 bp primary site (40-1AT) DNA at 37°C

Added unlabeled [PC]	Accelerated Rate constant (min^{-1})	% Cleaved, Accelerated	Nonaccelerated Rate constant (min^{-1})	% Cleaved, Nonaccelerated
0 nM	0.251±0.008	69.9±0.4		
10 nM	0.32±0.02	76±1		
100 nM	1.4±0.1	67.0±0.7	0.048±0.004	16.6±0.8
250 nM	4.7±1.3	66±1	0.039±0.007	18±2
500 nM	15.1±2.9	74±2	0.056±0.002	11±1
1 µM	17.2±2.4	76±1	0.055±0.002	10±1
2 µM	16.6±0.7	79.1±0.2	0.07±0.02	7.4±0.5

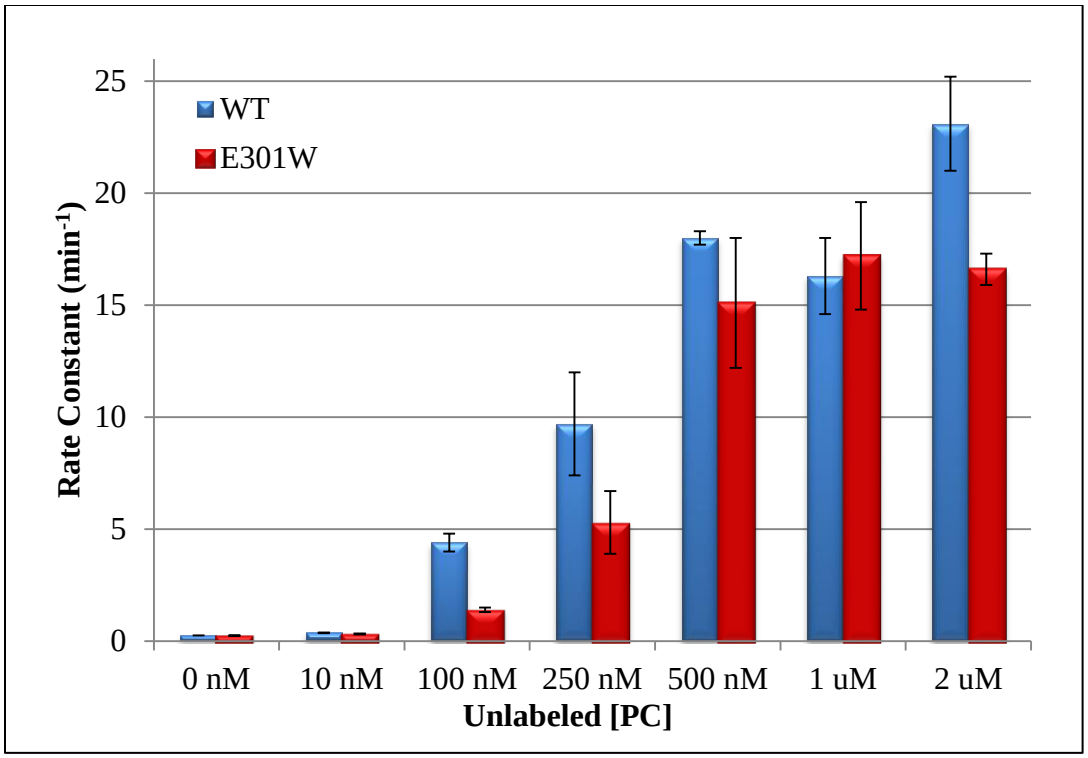


Figure 2.13: Comparison of WT and E301W SgrAI single turnover DNA cleavage rate constants.

SgrAI bound to 22-1HT DNA shows interactions with the neighboring DNA-bound-dimers similar to, but distinct from, those in the cryo-EM structure

The x-ray crystal structure of WT SgrAI bound to 22-1HT DNA (resolution = 3.5 Å, data statistics in Table 2.7) shows that the N-termini of neighboring SgrAI DBD come together to make many of the crystal lattice contacts, and that the flanking DNA of 22-1HT interact with the loops 56-60 and 122-140 from neighboring DBD in the crystal. Both the close approach of N termini of neighboring DBD (Figure 2.14-2.19), and contacts to the flanking DNA are reminiscent of those seen in the cryo-EM structure (Lyumkis *et al*. 2013), several important differences in these two distinct structures are seen. First, the neighboring SgrAI do not form the left handed run-on oligomer found in the cryo-EM structure (Figure 2.16). Second, the conformation of SgrAI in the crystal

structure is identical (within coordinate error) to the previously characterized low activity, dimeric, form, and does not exhibit the 10° rotation of one subunit relative to the other seen in the cryo-EM structure. In addition, the exact positions of the interacting N termini and the flanking DNA are different in the crystal and the cryo-EM structures (Figure 2.17-2.18). Unfortunately, due to the resolution limit (3.5 Å), it is difficult to conclude anything about the existence of domain swapping in the crystal structure, however, again, the N-termini are in sufficiently close proximity to make it a possibility (Figure 2.19). However, no indication of swapping is found in the electron density maps (Figure 2.14-2.15).

Table 2.7: Diffraction Data and Structure Refinement Statistics for WT SgrAI and 22-1HT	
Code	S634
Beamline	SSRL BL12_2
Processing Program	MOSFLM/SCALA
DNA	22-1HT
Space Group	C2
Cell	137.29Å, 50.68Å, 126.52Å; 100.78°
Resolution	3.5 Å
Total Observations	22256
% Complete	83.2%
I/Sigma	3.6
Multiplicity	2.6
R_{merge}[1]	30.1% (7.0% in top intensity bin)
R_{cryst}[2]	24.5 %
R_{free}[3]	37.6%
Overall B factor (Wilson Plot)	77.51 $Å^2$
RMSD-bonds	0.014 Å
RMSD-angles	1.42°
Asymmetric unit	2 SgrAI chains and 2 DNA strands
Number of waters	158
Number of divalent cations	4 Ca^{2+}

[1] $R_{merge}= \sum_{hkl}(|<I_{hkl}>-I_{hkl}|)/(\sum_{hkl} I_{hkl})$ where $<I_{hkl}>$ is the average intensity over symmetry related and equivalent reflections and I_{hkl} is the observed intensity for reflection hkl.

[2] $R_{cryst}=\sum_{hkl}(||F_{obs}|-|F_{calc}||)/(\sum_{hkl}|F_{obs}|)$ where $|F_{obs}|$ and $|F_{calc}|$ are the observed and calculated structure factor amplitude for reflection hkl. The sum is carried out over the 98% of the observed reflections which are used in refinement.

[3] R_{free} refers to the R factor for the test reflection set (2% of the total observed) which was excluded from refinement.

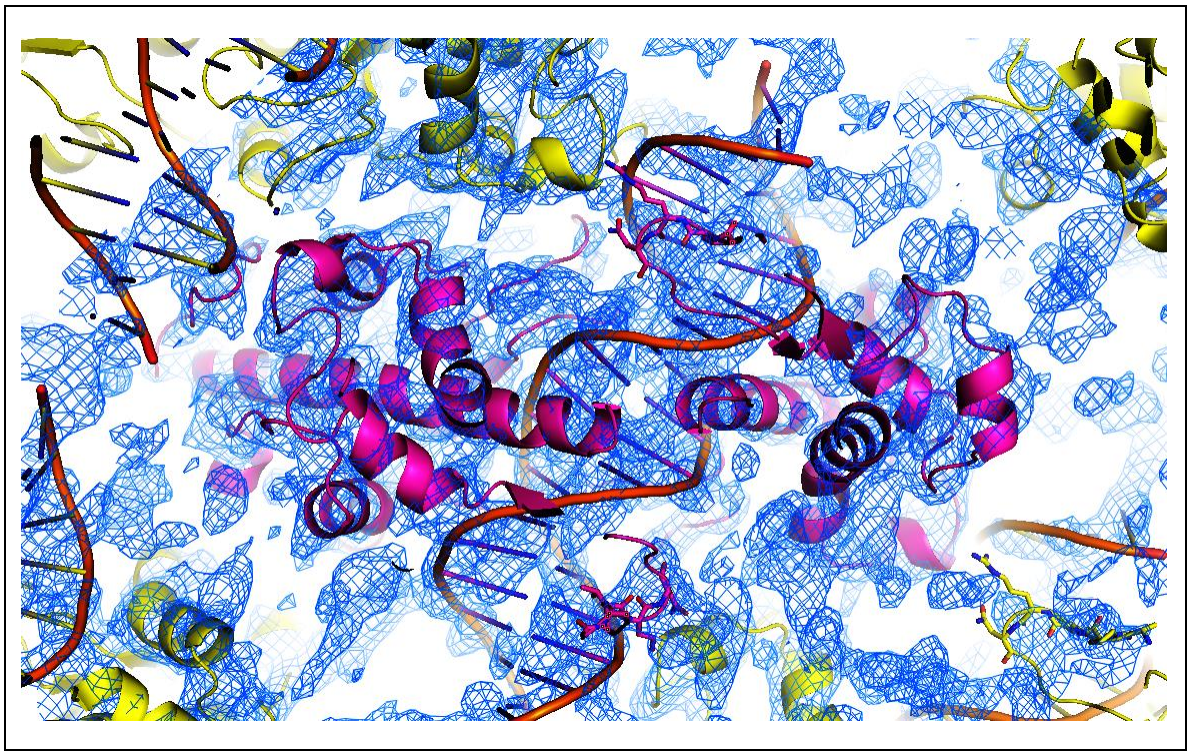


Figure 2.14: 2Fo-Fc electron density map of WT SgrAI bound to 22-1HT DNA. Figure courtesy of Prof. Nancy C. Horton.

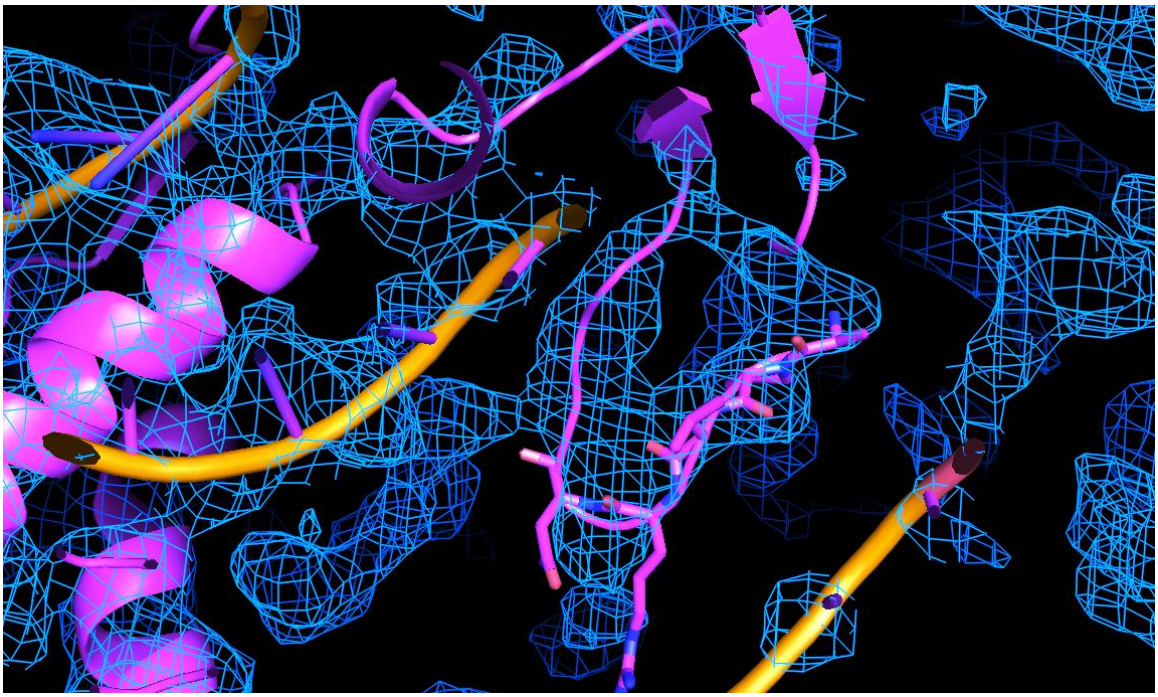


Figure 2.15: 2Fo-Fc electron density map of WT SgrAI bound to 22-1HT DNA showing the continuous electron density in the linker region shown as stick.

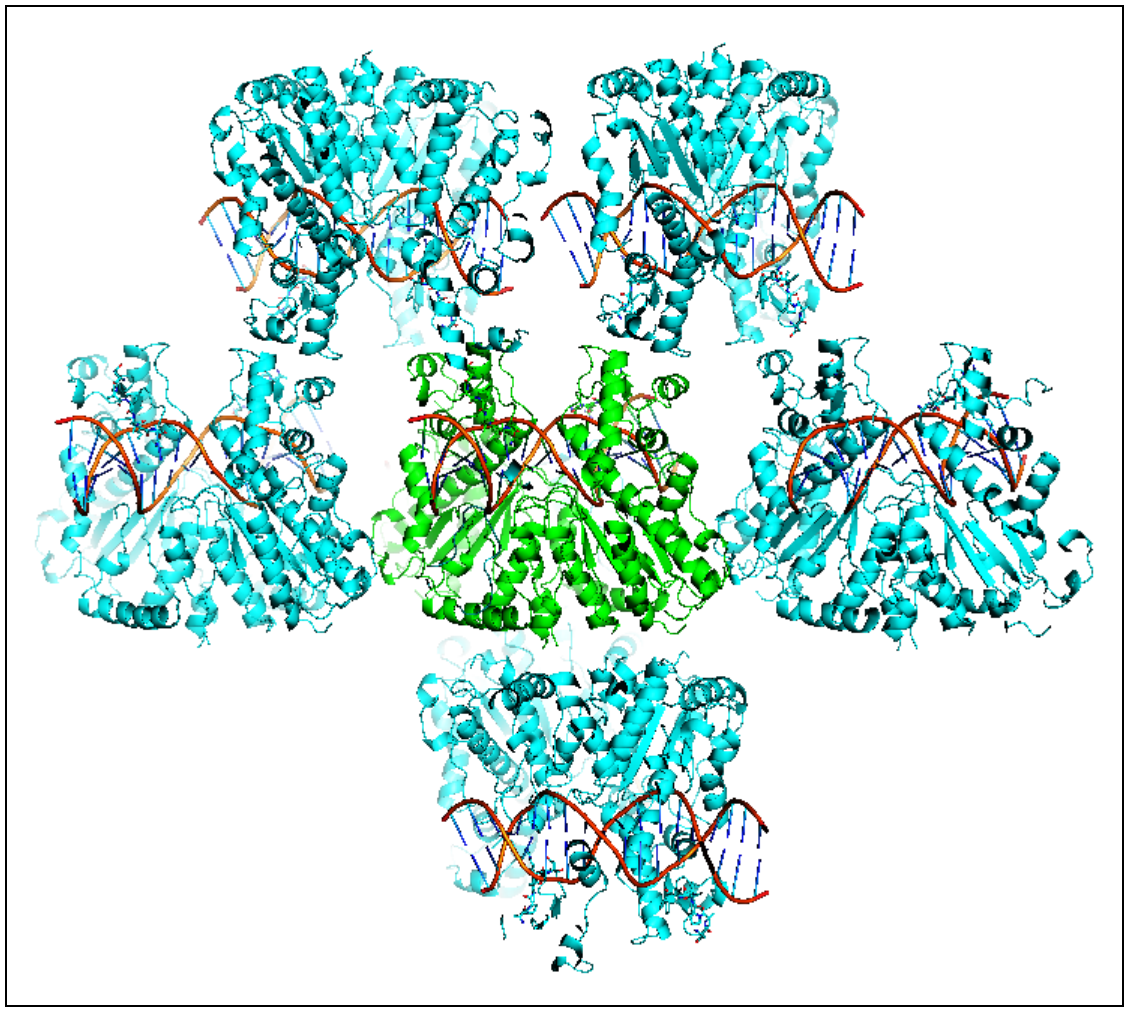


Figure 2.16: Symmetry related molecules (cyan) of WT SgrAI bound to 22-1HT (green) showing how the dimers are packed in the crystal. Figure courtesy of Prof. Nancy C. Horton.

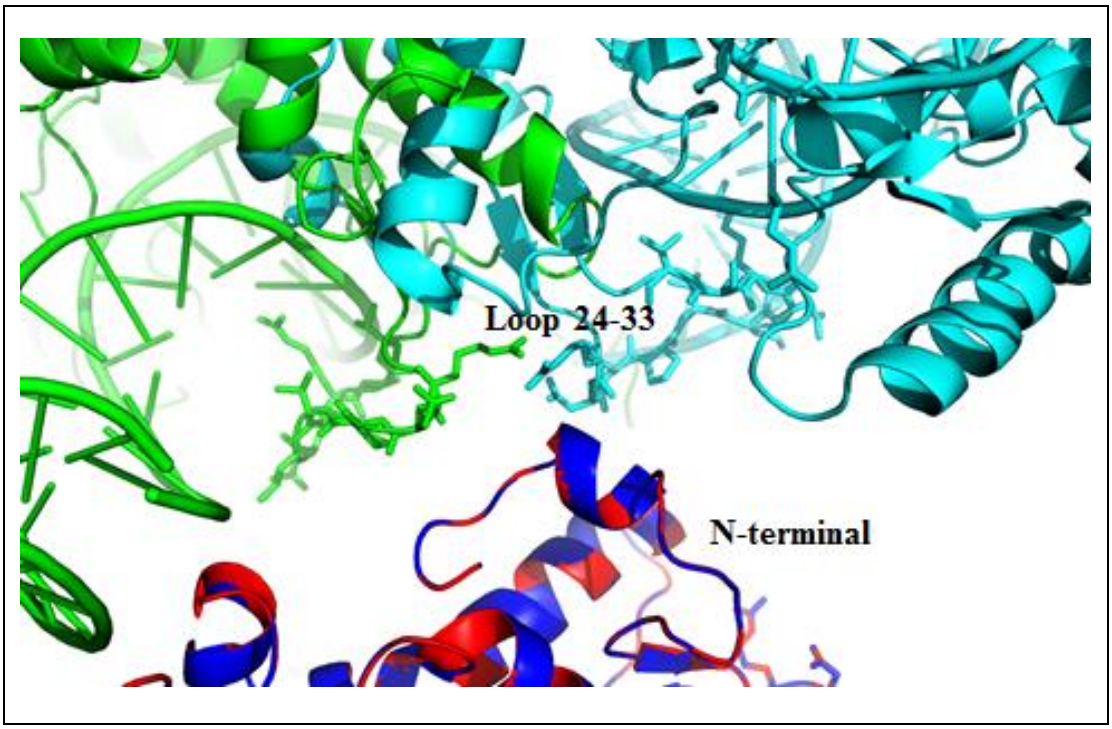


Figure 2.17: N-terminal contacts made in cryo-EM structure (blue and green) and crystal structure (red and cyan) show different orientation of loop 24-33 (linker). Loop 24-33 of symmetry related molecules are shown as stick.

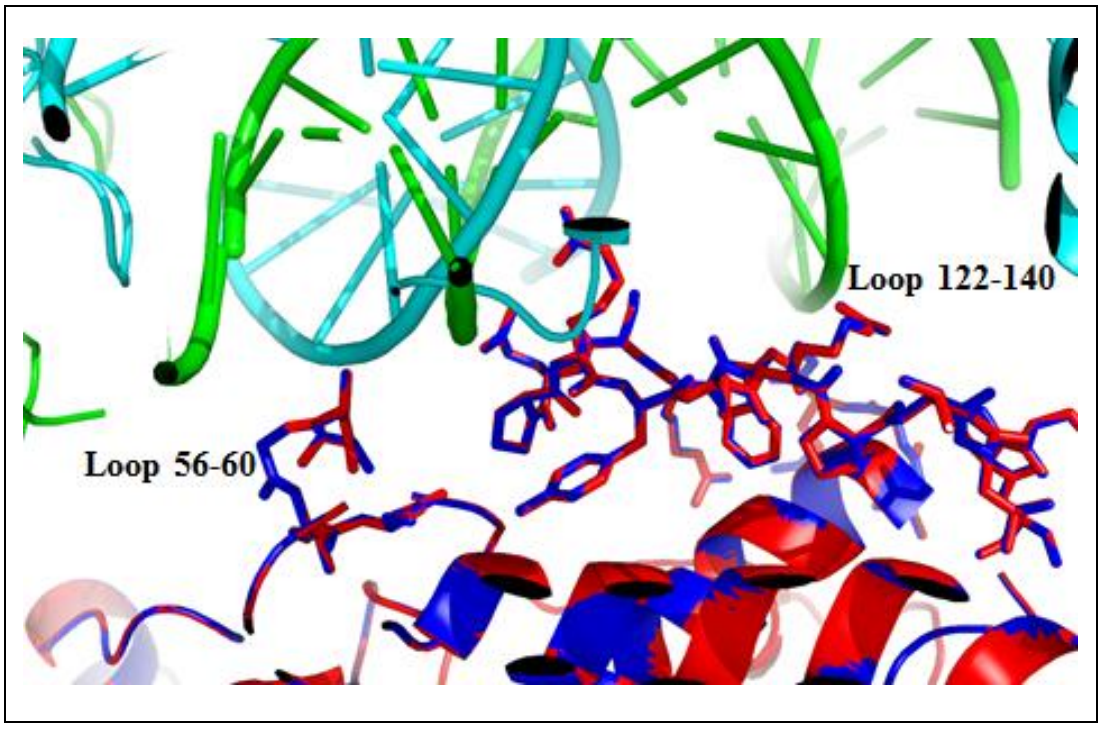


Figure 2.18: Loops 56-60 and 122-140 contacts made in cryo-EM structure (blue and green) and crystal structure (red and cyan) show similar contacts. Loops 56-60 and 122-140 are shown as stick.

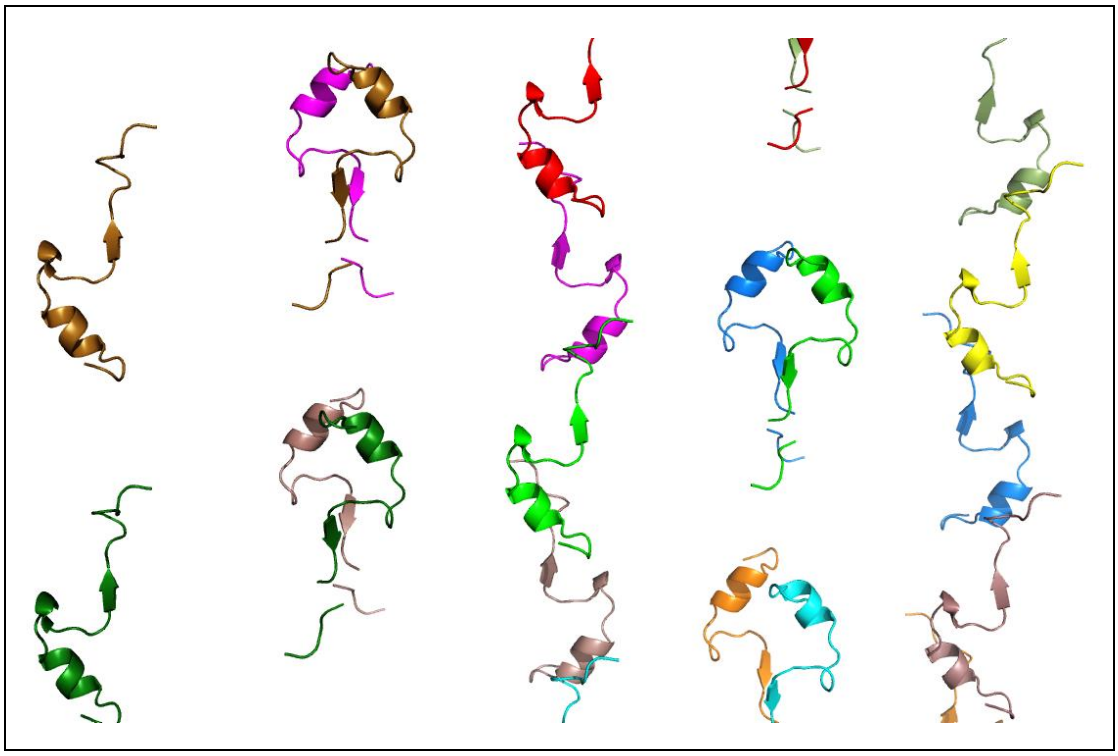


Figure 2.19: Possible domain swapping seen in crystal structure of WT SgrAI bound to 22-1HT DNA. Only the N-terminal domain is shown. Column 2 shows 2-way swap, whereas column 3 shows a 3-way swap. Figure courtesy of Prof. Nancy C. Horton.

Discussion

The effect of the R29A substitution of SgrAI activity

Based on the domain swapped structure of SgrAI (Figure 2.4B, Park *et al*. 2010a) arginine 29 lies in the linker region connecting the swapped domain (residues 1-22) to the rest of the protein, and makes a hydrogen bond to the swapped linker of a neighboring SgrAI subunit (Figure 2.5). The side chain guanidinium groups of arginine 29 make hydrogen bonds with the main chain oxygen of isoleucine 23, but only when the enzyme is in the swapped conformation. In order to test if domain swapping is involved in run-on oligomer formation, the R29A substitution was designed to disrupt this interaction thus destabilizing the domain swapped conformation. The binding assays, single turn-over assays, and crystal structure all indicate that the R29A substituted SgrAI behaves very similar to WT SgrAI, with the exception of its reduced solubility. The crystal structure shows a slight change in conformation of base adenine 2, but because no difference were seen in the activity of the enzyme as a result of the mutation, this slight change in conformation does not seem to affect the activity of the mutant. Thus, the R29 side chain does not appear to be important for the oligomeric SgrAI or activation of DNA cleavage by SgrAI, whether in domain swapping or other interactions in the run-on oligomer.

Involvement in NgoMIV-like tail-to-tail interface in activated SgrAI

Single turnover DNA cleavage rate constants of primary and secondary site DNA were measured with the E301W substituted SgrAI. They were found to be very similar to those of WT SgrAI, thus no destabilization of the activated form of SgrAI is found. The E301W results suggest that the activated form of SgrAI does not utilize the tail-to-tail

interface as seen in NgoMIV. The E301W SgrAI data also support the cryo-EM structure (Lyumkis *et al.* 2013) where the tail side does not make any contacts with either neighboring SgrAI or DNA.

The structure of WT SgrAI with activating DNA

The x-ray crystal structure of WT SgrAI bound to 22-1HT, a DNA found to result in some, albeit limited activation of SgrAI (Lyumkis *et al.* 2013) was solved at 3.5 Å. The crystal structure does not appear to have trapped the activated conformation of the enzyme, as the conformation of SgrAI is the same as in previously determined crystal structures (all assigned to the low activity conformation), and does not contain the 10º rotation between SgrAI subunits, as in the cryo-EM structure. Neither do the DBD in the crystal structure form the left handed helical run-on oligomer seen in the cryo-EM structure. However, interactions between DBD seen in the cryo-EM structure, namely those between N-termini and between flanking DNA and the loops 56-60 and 122-140, are found in this new crystal structure. The N-terminal domain appears to be making two-way and three-way contacts with its neighbors. Since longer flanking DNA stabilizes the oligomeric form, and provides more robust activation of DNA cleavage, attempts to crystallize the activated, oligomeric form of SgrAI should proceed with longer DNA.

Part C: Biological role of SgrAI run-on oligomer

Introduction

Restriction endonucleases (REs) are specific enzymes which cleave DNA at or near specific recognition sequences. They are part of restriction modification (R-M) systems that defend the host against foreign phage invasion (Arber *et al.* 1979, Pingoud *et al.* 2005). The host DNA is protected by methylation of N4 or C5 at cytosine or N6 at adenine. *Streptomyces griseus* has a complex multicellular development, and unlike other eubacterial chromosomes, its chromosome is linear consisting of 8,545,929 base pairs. It also has a high G+C content, i.e. 72.2% (Ohnishi *et al.* 2008). The large genome might hold clues to why SgrAI has evolved into such an unusual enzyme. The large genome results in more recognition sites than other bacterial species, which puts pressure on the methyltransferase activity to methylate so many sites. However, since the methyltransferase activity may be limited, possibly by the concentration of its cofactor *S*-adenosylmethionine (SAM) concentration, *S. griseus* may have evolved the SgrAI endonuclease to recognize a longer sequence (8 bp) and have a slow basal DNA cleavage rate. These effects reduce both the number of sites to be methylated and the likelihood of cleavage of any site by SgrAI prior to methylation by the cognate methyltransferase. However, the slow basal cleavage rate and rare recognition sequence may also limit effective phage protection. Activation of SgrAI (providing 200-1000 fold rate enhancement) and sequence-specificity expansion from 3 primary sites to total 17 different primary and secondary sites may have evolved to increase the anti-phage

activity. However, to prevent host DNA from being cleaved at the secondary sites presumably not methylated by the methyltransferase, the run-on oligomer could act to sequester activated SgrAI on the phage DNA and away from the host genome. This hypothesis was tested using a plasmid sequestration assay. Two types of plasmids were used: one with only secondary sites (pSK1, simulating the host DNA, since all the primary sites of host DNA will be protected by methylation), and the other with mixture of primary and secondary sites (pBR322, simulating the phage DNA since the phage DNA will not be methylated by host methyltransferase) (Figure 2.20). First, individual plasmids were cleaved by SgrAI and electrophoresed in 1% agarose gels in order to identify their cleavage patterns. The sequestration assay was conducted with an equimolar mixture of the two plasmids with sufficient SgrAI to saturate all primary and secondary sites. The electrophoresis pattern of the mixed plasmid conditions was then compared with the individual plasmid digests to determine if only one type or both types of plasmids were cleaved at the secondary sites. The hypothesis predicts that at physiological concentrations of SgrAI and DNA, SgrAI will cleave the secondary sites of the plasmid containing the primary site (i.e. pBR322, the phage-like plasmid) only, leaving the host-like plasmid pSK1 untouched.

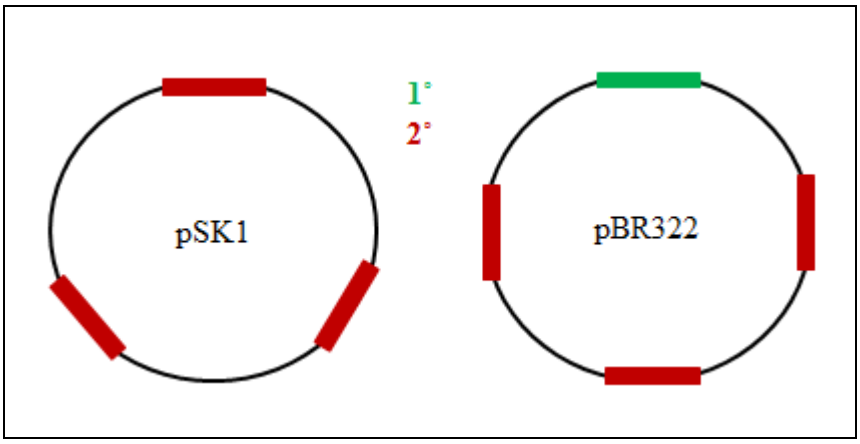


Figure 2.20: Plasmid map of pSK1 (2969 bp) and pBBR322 (4361 bp) showing the number of primary (1º, green) and secondary (2º, red) sites present in the plasmids. The relative positions of the sites are not drawn to scale as present in the actual plasmids.

Results and Discussion

Although the physiological concentration of SgrAI within *Streptomyces griseus* is not known, it is the concentration of unmethylated primary site DNA that is more relevant since activation is dependent on the concentration of primary site bound SgrAI (Park *et al*. 2010b). The only unmethylated primary site DNA in the cell should be that on the invading phage DNA, and with one phage genome per cell, this concentration should be in the low nanomolar range. The plasmid sequestration assay was performed with 50 nM plasmid DNA and 350 nM SgrAI. Complete cleavage of all sites, primary and secondary, on pBR322 occurs at these concentrations resulting in the expected cleavage products (red 1-3, Figure 2.22). However, mostly nicking and only cleavage at a single site per plasmid late in the timed assay (120 min) is found with pSK1 (Figure 2.21). When both plasmids are present (Figure 2.23), the band corresponding to supercoiled, intact pSK1 (green *, Figure 2.23) persists throughout most of the assay (becoming linear in the last two lanes), while cleavage of all four sites on pBR322 is

clear (red letters 1-3, Figure 2.23). None of the products from cleavage of secondary sites on pSK1 are present (these would occur between 900-1000 bp). This shows that activated cleavage by SgrAI, which results in cleavage at secondary sites as well as primary, is sequestered on the plasmid containing the primary site (pBR322). These results predict that secondary site cleavage should be sequestered to the invading phage DNA, and away from the host DNA, *in vivo*. It is expected that at higher concentrations of SgrAI, the cleaved primary sites would act *in trans* and activate SgrAI cleavage of secondary sites in pSK1, just as PC DNA activates the cleavage of secondary sites in 18-2A, 40-2A, and 40-2B (Park *et al*. 2010b and Table 2.1). Using Table 2.1 as a guide, this would be expected to occur at concentrations of SgrAI and primary site DNA of approximately 500 nM. Such studies have yet to be performed to confirm this prediction. This sequestration assay should also prove useful to further test the proposed biological role of run-on oligomer formation by SgrAI, by testing the single site substitutions which have been shown to reduce oligomer stability. Such mutated SgrAI should show sequestration at higher concentrations of SgrAI and primary site DNA than wild type SgrAI, since higher concentrations will be needed to favor oligomer formation.

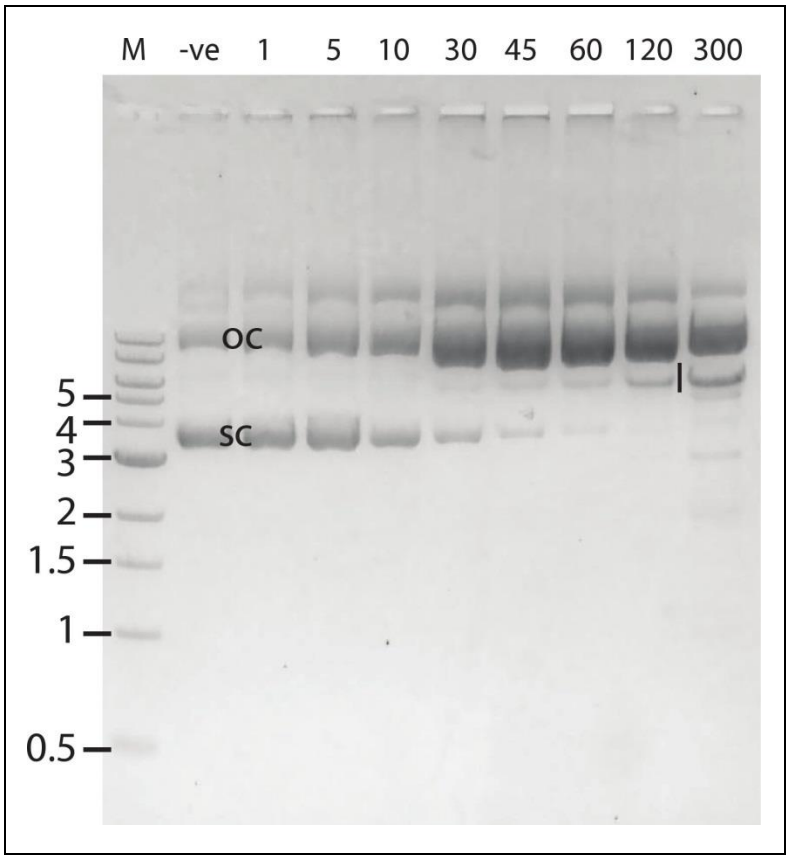


Figure 2.21: Digestion pattern with 50 nM pSK1 and 350 nM WT SgrAI. Size markers on the left are in kilobases. OC = nicked open circle, SC = supercoiled, l = linear DNA. Incubation times are in minutes. pSK1 is 2969 bp, and products of secondary site cleavage by SgrAI would give 1090, 964, and 915 bp. This preparation of pSK1 appears to have isolated a dimer of the plasmid.

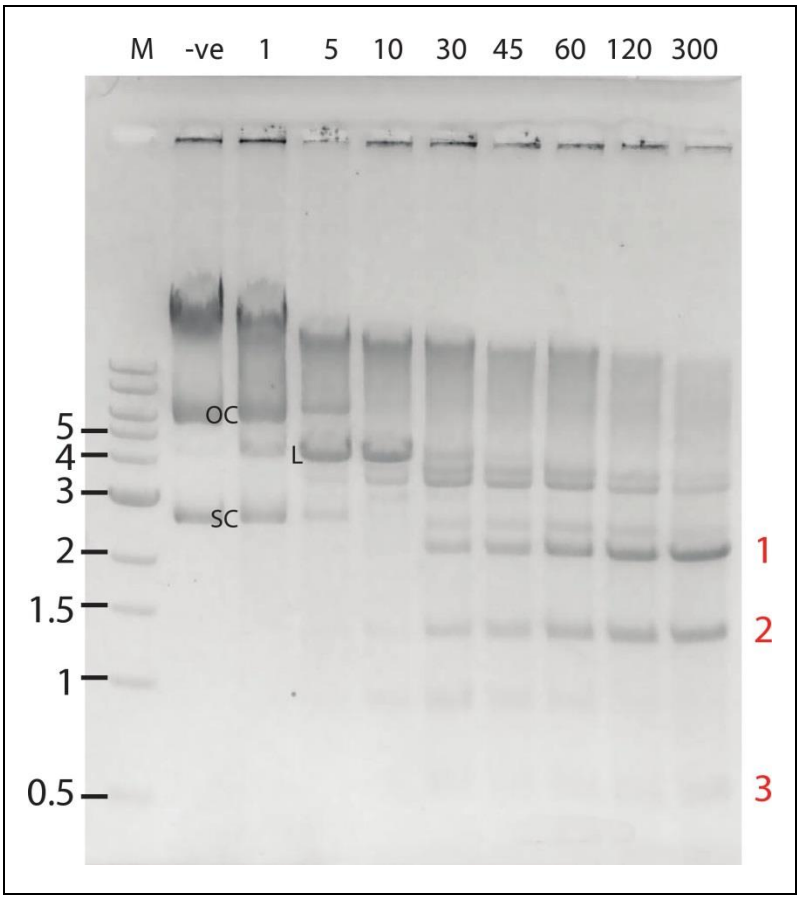


Figure 2.22: Digestion pattern with 50 nM pBR322 and 350 nM WT SgrAI. Size markers are in kilobases. OC = nicked open circle, SC = supercoiled, L = linear DNA. Incubation times are in minutes. pBR322 is 4361 bp, and products of primary and secondary site cleavage by SgrAI give 2157 (red number 1), 1325 (red number 2), 513 (red number 3), and 360 bp (not visible, run past edge of gel).

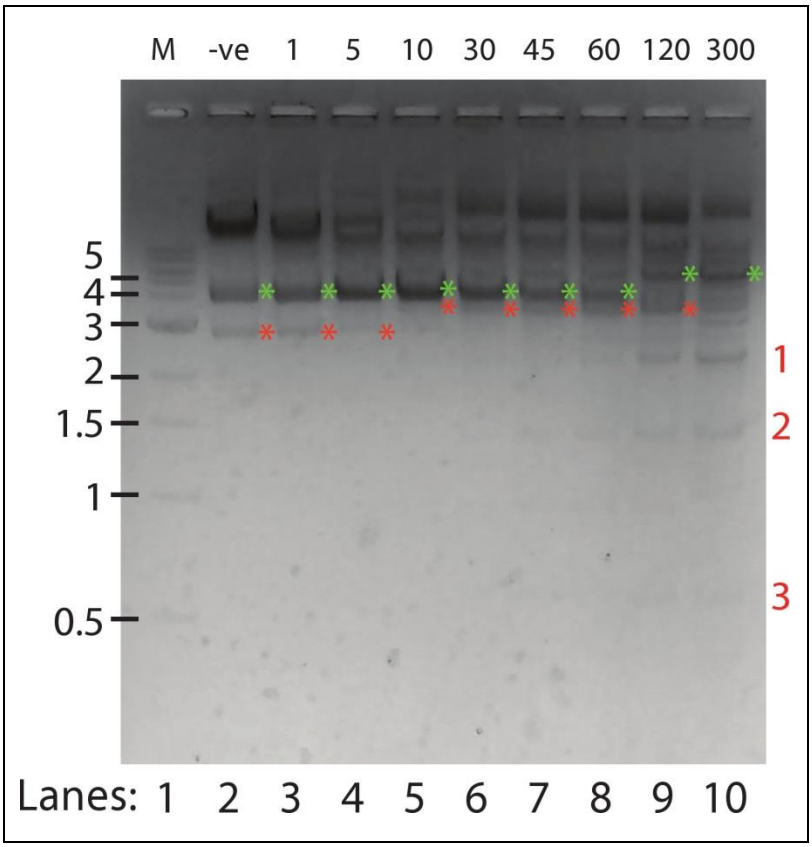


Figure 2.23: Sequestration assay using 50 nM pSK1 and 50 nM pBR322 and 350 nM WT SgrAI. Size markers are in kilobases. Incubation times are in minutes. Green * indicate uncleaved, supercoiled pSK1 (lanes 2-8) or linearized pSK1 (lanes 9-10), and red * indicate uncleaved, supercoiled pBR322 (lanes 2-4) or linearized pBR322 (lanes 5-9). Red numbers at right indicate cleavage products of pBR322, as in Figure 2.22.

Materials and Methods:

Protein Purification

All mutations in the SgrAI gene (R29A, E301W) were prepared at New England Biolabs by collaborators Dr. Alka Agrawal (E301W) or Dr. Jurate Bitinaite (R29A). Wild type, R29A, and E301W SgrAI were prepared as described (Dunten *et al.* 2008) with some modifications. Briefly, pET21a_SgrAIR (containing either WT or mutant sequences) plasmid were transformed into T7 Express Competent *E. coli* cells (High Efficiency, New England Biolabs) which constitutively express the MspIM methyltransferase to protect the host genome. The cells were grown in five batches of 18-L fermenter at 37°C until the OD_{600} reached 0.5 (Cary-50 Bio, Varian). The cells were then induced with 0.4 mM IPTG and was left to grow overnight at 17°C. The cells were then harvested using Avanti J-20 centrifuge (Beckman Coulter) at 4,800 x g for 30 minutes, frozen in liquid nitrogen and stored at -80°C. The frozen cell pellets were resuspended in breaking buffer (100 mM Tris (pH 8.0), 800 mM NaCl, 10 mM EDTA, 10 mM β-Me, and 100 mM PMSF) and sonicated using Branson Sonifier 450 (VWR Scientific). The lysed cells were centrifuged for 1 hour in Sorvall – Superspeed RC2-B at 16,300 x g followed by centrifugation at 190,000 x g for 1 hour in L8-70M Ultracentrifuge (Beckman Coulter). The enzyme was purified using FPLC (GE Healthcare Biosciences) chromatography and the following chromatographic resins: Heparin FF Sepharose (GE Healthcare Biosciences), SP FF Sepharose (GE Healthcare Biosciences), Q FF Sepharose (GE Healthcare Biosciences), followed by a second Heparin FF Sepharose (GE Healthcare Biosciences) chromatographic step. Finally, SgrAI

enzyme was dialyzed into storage buffer (20 mM Tris-OAc (pH 8.0), 50 mM KOAc, 0.1 mM EDTA, 1 mM DTT, 50% glycerol), aliquoted into single use aliquots, flash frozen in liquid nitrogen, and stored at -80°C.

His-tagged protein (WT) was purified using talon resin followed by heparin resin. The cell lysate was incubated with talon resin in lysis buffer (50 mM NaH_2PO_4 pH 8.0, 800 mM NaCl, 10 mM Imidazole, and 1 mM β-Me) overnight. The unbound cell lysate was washed off using wash buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 20 mM Imidazole, and 1 mM β-Me) followed by high salt wash buffer (50 mM NaH_2PO_4 pH 8.0, 2 M NaCl, 20 mM Imidazole, and 1 mM β-Me). Finally, the protein was eluted using elution buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 250 mM Imidazole, and 1 mM β-Me). The partially purified protein was then excessively dialyzed into Heparin A buffer (20 mM Tris-HCl pH 8.0, 50 mM NaCl, 0.1 mM EDTA, 10 mM β-Me, and 20% glycerol). The dialyzed protein was purified using a Heparin FF column connected to FPLC. The protein was eluted off using Heparin B buffer (20 mM Tris-HCl pH 8.0, 1 M NaCl, 0.1 mM EDTA, 10 mM β-Me, and 20% glycerol). Purity of the protein was confirmed using SDS-PAGE. The purified protein was then aliquoted into single use aliquots, flash frozen in liquid nitrogen, and stored at -80°C.

DNA Preparation:

The oligonucleotides were made synthetically and purified using C18 reverse phase HPLC or denaturing PAGE (Aggarwal *et al.* 1990). The concentration was measured spectrophotometrically, with an extinction coefficient calculated from standard values for the nucleotides (Fasman *et al.* 1975). The self-complimentary DNA strands, or

equimolar quantities of complementary DNA, were annealed by heating to 90°C for 10 minutes at a concentration of 1 mM, followed by slow-cooling to 4°C over 4-5 hours in a thermocycler. Sequences of the DNA used are shown below (cleavage sites are identified with arrows):

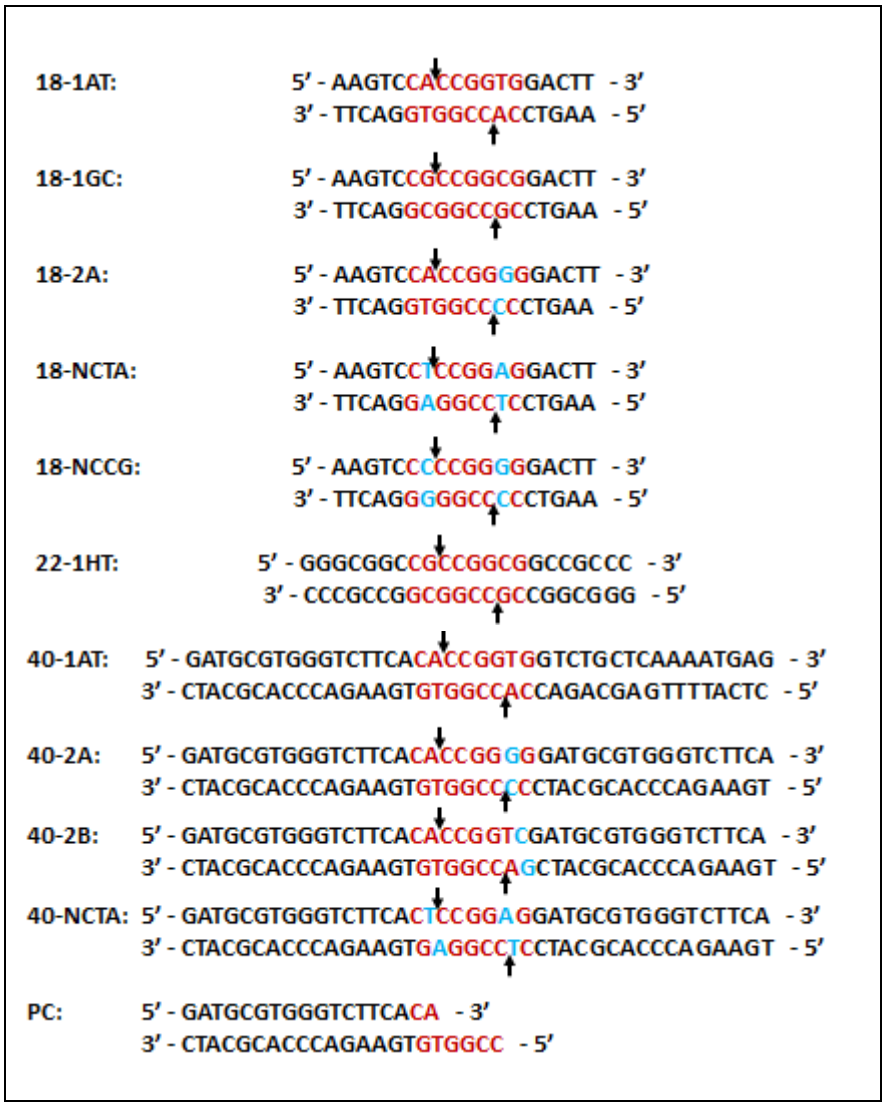


Table 2.8: DNA constructs used in determining dissociation constants, cleavage rate constants and crystal structure. Recognition sequences are shown in red, whereas the variations in secondary site sequences are shown in cyan. Arrows indicate the cleavage sites.

Because freeze-thawing altered the concentration of double stranded DNA used in the assays, DNA used for single turnover assays was treated very carefully to minimize this problem. Such DNA samples were aliquoted into small amounts, flash frozen in liquid nitrogen, stored at -20°C (in water), and used only once after removing from the freezer. DNA was 5´ end labeled with ^{32}P using T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P]-ATP (Perkin-Elmer, Inc.), and excess ATP removed using G-30 spin columns (Bio-Rad Laboratories, Inc.).

Single Turnover DNA Cleavage Assay

Single turnover kinetic measurements of DNA cleavage were performed using 5´ end ^{32}P-labeled oligonucleotides substrates (typically 1 nM), under conditions of excess enzyme (typically 1 μM), with and without the addition of unlabeled DNA. For reactions using 40-2A-Top and 40-2A-Bottom, either the top or bottom strands respectively were labeled before being annealed. For all the other reactions the complimentary strands were annealed before being labeled. All reactions were performed at 37°C in 20 mM Tris-OAc (pH 8.0), 50 mM KOAc, 10 mM $Mg(OAc)_2$, and 1 mM DTT. 5 μl aliquots were withdrawn at specific time intervals after mixing the enzyme and labeled DNA (100 μl total reaction volume), quenched by addition to 5 μl of quench (80% formamide, 50 mM EDTA, 1 mg/ml XCFF dye, and 1 mg/ml BPB dye), and electrophoresed on 20% denaturing polyacrylamide (19:1 acrylamide:bisacrylamide, 4 M urea, 89 mM Tris, 89 mM boric acid, 2 mM EDTA) gels. Autoradiography of gels was performed without drying using a phosphor image plate exposed at 4°C for 12-17 hours. Densitometry of phosphor image plates was performed with a Typhoon Scanner (GE Healthcare Life

Sciences), and integration using ImageQuant (GE Healthcare Life Sciences) or ImageJ (Abramoff 2009). The percent of product formed as a function of time was determined by integrating the density of both cleaved and uncleaved DNA bands, and normalizing to the total amount cleaved. The fraction of cleaved DNA was plotted against time. This plot obtained was then fit into either single or the sum of multiple single exponential functions to determine the single turnover rate constant of DNA cleavage:

$$\text{Percentage of product} = C_1 + C_2 * (1 - e^{-kt})$$

where C_1 is a constant fitting the baseline, C_2 is the total percent of DNA predicted to be cleaved by SgrAI, k is the cleavage rate constant, and t is the length of incubation in minutes. The data from some reactions fit poorly to a single exponential function, suggesting multiple phases. These were found to fit well to the sum of two exponential functions:

$$\text{Percentage of product} = C_1 + C_2 * (1 - e^{-k1t}) + C_3 * (1 - e^{-k2t})$$

where C_1 is a constant fitting the baseline, C_2 is the total percent of DNA predicted to be cleaved by SgrAI with rate constant k_1, C_3 is the total percent of DNA predicted to be cleaved by SgrAI with rate constant k_2, and t is the length of incubation in minutes. The rate constants are presented as k_1 and k_2, where k_1 is the greater (accelerated) of the two rate constants k_1 and k_2, and k_2 is the lower (non-accelerated). For reactions fit by a single rate constant, the rate constant is listed as accelerated if it was significantly greater than the unactivated rate constant of 0.094 ± 0.015 min^{-1}. Measurements were performed at least three independent times, and presented as the average±standard deviation.

Sequestration Assay

Two types of plasmids were used for the sequestration assay: one with only secondary sites (pSK1 (Bitinaite *et al.* 2001), simulating the host DNA, obtained from Dr. Jurate Bitinaite, New England Biolabs), and the other with mixture of primary and secondary sites (pBR322 simulating the phage DNA) (Figure 2.20). 50 nM of each pSK1 and pBR322 plasmids were incubated with 350 nM SgrAI, and 20 µl aliquots were stopped with 50 mM EDTA, 10% glycerol solution at various time points. The reaction was electrophoresed on a 1% agarose gel. The cleaved bands were compared with the reaction products that were obtained by incubating 50 nM of pSK1 or pBR322 individually with 350 nM SgrAI. Comparison of individually digested plasmid with the mixed plasmid digestion allowed assignment of cleaved bands to each plasmid.

Crystallization, Data Collection, Structure Solution, Refinement, and Analysis

The purified WT or R29A SgrAI protein was removed from -80°C storage, dialyzed overnight, and concentrated using Amicon Ultracel-10K filter centrifuged at 7,000xg in a Jouan-BR4 centrifuge. WT or R29A SgrAI and DNA duplex were mixed in a 1:2 molar ratio for the cocrystallization experiments. The ratio was chosen to ensure that there was an excess of DNA for DNA bound dimer formation. The initial screen was carried out using hanging drop method (McPherson 1976) with varied precipitating conditions (5 mM $CaCl_2$, 150 mM NaCl, 100 mM buffer (pH 4.5-8.5), and 5-20% PEG 4K) and 3.5-10 mg/ml protein at 17°C. Based on the initial screens follow up trials were set up which led to optimization of the crystallization conditions. For WT SgrAI bound to

22-1HT DNA, crystals grew best in about 4 to 5 months at 17°C using 10 mM $Ca(OAc)_2$, 150 mM NaCl, 100 mM imidazole buffer pH 6.5, and 15% PEG 4K as the precipitant and 3.6 mg/ml protein and protein:DNA of 1:1.5. The crystals were then exchanged into a cryoprotectant solution (10 mM $Ca(OAc)_2$, 300 mM NaCl, 100 mM imidazole buffer pH 6.5, and 25% PEG 4K and 30% glycerol) and flash-frozen in liquid nitrogen. X-ray diffraction was measured using synchrotron radiation at the Stanford Synchrotron Light Source (SSRL) BL12-2 (WT) or BL9-2 (R29A). Data collection was performed while maintaining the crystal at 100K. Image processing and data reduction were performed with MOSFLM (Leslie 1999) and SCALA (Evans 2006) (R29A) or XDS (WT) (Kabsch, 2010). The structure was solved by molecular replacement using 3DVO (Dunten *et al.* 2008). The structures were refined using REFMAC (Murshudov *et al.* 1999) in the CCP4 program suit (CCP4 1994) and the model building program Coot (Emsley *et al.* 2004) (R29A) or PHASER (McCoy, 2007), PHENIX (Adams *et al.* 2002), and PYMOL (DeLano 2002) (WT). All structure figures were prepared using PYMOL (DeLano 2002).

CHAPTER 3:

INDIRECT READOUT OF DNA SEQUENCE IN SGRAI K96A

ABSTRACT

Many DNA endonucleases that are involved in DNA repair, transposition, recombination, modification of DNA supercoiling, and bacterial immunity recognize specific sequences in DNA using direct and indirect readout. SgrAI, a Type IIF restriction endonuclease recognizes and cleaves an 8-bp sequence, $C_1R_2|C_3C_4G_5G_6Y_7G_8$. The crystal structure of SgrAI shows that the enzyme makes extensive direct contacts with the inner 4 bases ($C_3C_4G_5G_6$) but makes no direct readout contacts to the R_2Y_7 base pair. Also, there is only a single direct readout contact made per base pair to the outer two base pairs. A large DNA distortion is seen in the crystal structure at the Y_7G_8 base step which may be a distortion imposed or recognized by SgrAI to specific this base step. Recognition of DNA sequence using DNA distortion has been termed indirect readout. Lys 96 of SgrAI lies between the unstacked bases, in the minor groove, and may be important in stabilizing this distortion. To investigate the role of Lys 96 in indirect readout of the DNA sequence at the Y_7G_8 step, Lys 96 was mutated to alanine, and DNA binding specificity, cleavage preferences, and crystal structures with different DNA sequences were determined. The K96A SgrAI enzyme possesses stronger binding affinity for recognition sites modified such that Y_7 has been substituted with G, i.e. CRCCG<u>G</u>G, (also known as a secondary site) than for its recognition sequence CRCCGGYG (i.e. primary sites). However, K96A SgrAI shows no cleavage or very slow cleavage of the

secondary site and non-cognate sites, respectively, even after a 24-hour incubation at 37ºC. Intriguingly, the x-ray crystal structure of K96A SgrAI bound to primary site DNA shows no significant differences with that of the wild type SgrAI bound to the same DNA sequence (RMSD of 0.293 Å between the two structures).

INTRODUCTION

DNA binding proteins carry out many essential functions in living cells. Many of these proteins that cleave duplex DNA are divalent cation-dependent and are involved in DNA repair, transposition, recombination, modification of DNA supercoiling, and bacterial immunity (Cowan 1998, Horton *et al.* 2002, Etzkorn *et al.* 2004). Most of these enzymes recognize specific sequences in DNA using direct and indirect readout (Garvie *et al.* 2001). Direct readout involves base discriminating direct contacts between the protein side chain or main chain atoms and the DNA bases via hydrogen bonding and/or van der Waals interactions, typically in the minor groove of the DNA. Sequence specificity occurring in the absence of direct readout is considered to be evidence of indirect readout. Indirect readout is the ability of a protein to distinguish DNA sequences without direct readout contacts from the protein (Koudelka *et al.* 2006). Indirect readout involves contacts mediated by water (Otwinowski *et al.* 1998), other small molecules (Segal *et al.* 2006), sugar-phosphate backbone, and/or distortion of DNA that allows distinction of different DNA sequences (Rice *et al.* 1996, Koudelka *et al.* 2006, Blakaj *et al.* 2006, Little *et al.* 2008). The first proteins recognized to use an indirect readout mechanisms were *trp* repressor (Scheviz *et al.* 1985), TATA box binding protein (Kim *et al.* 1993), and 434 repressor (Rodgers *et al.* 1993).

Restriction endonucleases (REs) are bacterial enzymes that cleave DNA at specific target sequences, and are useful in studying direct and indirect readout mechanism. They are thought to protect bacteria from phage infection. Unlike the phage

DNA, the bacterial genomes are methylated by a cognate DNA methyltransferase and are thus, protected from endonuclease activity (Roberts *et al.* 1993, Pingoud *et al.* 2001). Since only specific target sequence should be cleaved to prevent damage to the host DNA, restriction endonucleases have evolved very efficient DNA recognition mechanisms. Thus, REs are a great model system for studying sequence-specific DNA recognition (Kovall *et al.* 1999). In addition, sequence-specific REs with a longer base-pair recognition sequence are of particular importance in genomic studies, as longer sequences occur less frequently in the genome they allow manipulation of larger DNA fragments (Tautz. *et al.* 1990b, Dunten *et al.* 2008). Type II REs recognize specific DNA sequences and cleave DNA within the binding site typically using Mg^{2+} as a cofactor (Sapranausks *et al.* 2000). They do not require ATP or *S*-adenosyl methionine (SAM) for their activity (Roberts *et al.* 2003, Mucke *et al.* 2003). The recognition sites are often 4-8 base-pair palindromic sequences, and the enzymes that recognize these sites are usually homodimers. SgrAI, a Type IIF restriction endonuclease native to *Streptomyces griseus,* recognizes and cleaves an 8-bp sequence, CR|CCGGYG (R = A or G, Y = C or T, | denotes cut site, numbering is $C_1R_2C_3C_4G_5G_6Y_7G_8$). This type of rare cutting endonuclease can be a useful tool for restriction mapping of genomic DNA (Tautz *et al.* 1990a, Tautz *et al.* 1990b). In addition to this rare cutting property, SgrAI also shows unusual biochemical phenomena: it cleaves plasmids with two sites much faster than ones having only a single cleavage site, and cleaves DNA at secondary sites as well (CR|CCGGY(A,C,T) and CR|CCGGGG) but only in the presence of at least one primary site (Bitinaite *et al.* 2002). SgrAI is appealing to scientists because REs such as Cfr10I

(R|CCGGY) (Siksnys *et al.* 1999) and NgoMIV (G|CCGGC) (Deibert *et al.* 2000) have similar target sites as SgrAI, however, they do not show any modulation of specificity as seen in cleavage activity of SgrAI. Moreover, Cfr10I and NgoMIV, like most other type IIF REs, form tetramers while SgrAI appears to form a dimer in the absence of DNA, and a species much larger than tetramer upon binding to DNA (Siksnys *et al.* 1999, Deibert *et al.* 2000). Preliminary studies suggest that this unique restriction endonuclease recognizes and cleaves at the target site using direct as well as indirect readout.

SgrAI binds to a specific DNA sequence $C_1R_2|C_3C_4G_5G_6Y_7G_8$, but makes no direct readout contacts to the Y_7R_2 base pair (Figure 3.1A). Further, it is unusual that only a single direct readout contact is made with the outer two base pairs of the 8 bp recognition sequence, while the inner base pairs have direct readout interactions at both the bases of each base pair, which is more typical in type II REs. The side chain of Lys 96 interacts with the DNA very near to an unusual distortion in the DNA. Since there are no direct contacts at the Y_7R_2 bp, yet specificity (although degenerate) occurs, indirect readout is assumed to provide the sequence specificity. The Lys 96 side chain contacts the outer G_8 of the SgrAI recognition sequence, $C_1R_2C_3C_4G_5G_6Y_7G_8$, from the minor groove. Atoms N_{zeta} and $C_{epsilon}$ of Lys96 are each 3.6 Å away from atoms N_3 and C_2 of base G_8, respectively (Figure 3.1B). The contacts appear to stabilize the unstacking of the T_7-G_8 bases, a distortion in the DNA. This DNA distortion has been proposed to provide the specificity for the pyrimidine base at position 7 in the recognition sequence via indirect readout (Dunten *et al.* 2008), as YR base steps have been found to be the least costly energetically to unstack (Friedman *et al.* 1995, Horton *et al*. 2002). Hence

non-YR sequences may be too costly to distort in this manner, thus greatly diminishing the binding affinity to SgrAI. To test this proposal, Lys 96 was mutated to Ala to remove the direct contact made between Lys 96 of SgrAI and the DNA base G_8. The proposal predicts that such a mutation should relax specificity at the Y_2R_7 base pairs. Sequence specificity in terms of sequence preferences for binding and cleavage were measured using a primary site DNA sequence ($C_1R_2C_3C_4G_5G_6Y_7G_8$) and several non-cognate sequences ($C_1N_2C_3C_4G_5G_6N_7G_8$), with both WT and K96A SgrAI enzymes. In addition, the X-ray crystal structures of WT and K96A SgrAI bound to primary site DNA were also determined.

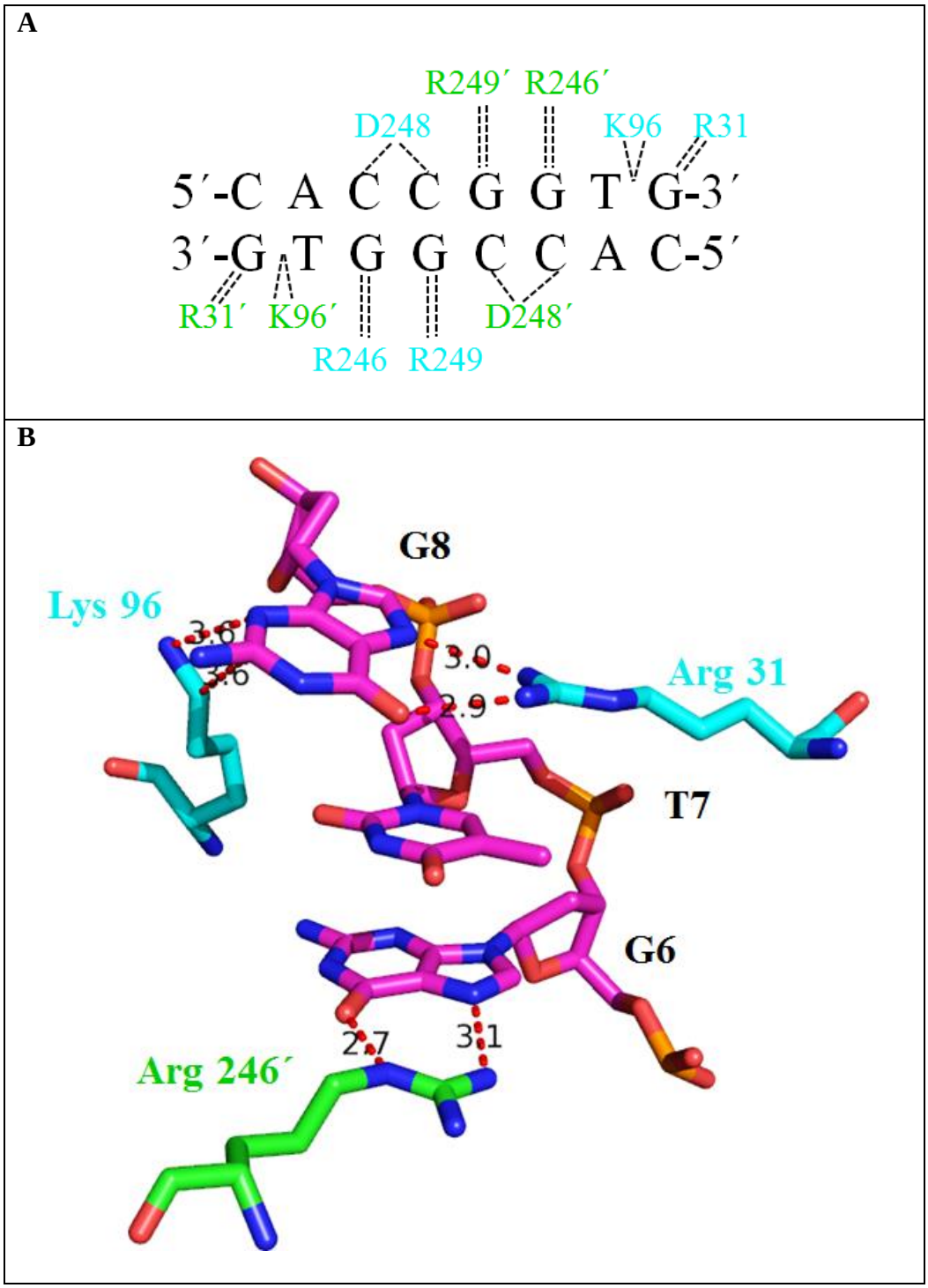


Figure 3.1: (A) Schematic diagram showing interactions of DNA recognition sequence and the neighboring residues. (B) Stick diagram showing interactions between the Lys 96 in wild type SgrAI and oligonucleotide G6, T7 and G8. Adapted from Dunten *et al.* 2008.

MATERIALS AND METHODS

Construction of SgrAI K96A Mutant

A modified USER™ friendly DNA engineering method (Bitinaite *et al.* 2007) was used to introduce the K96A codon change into the SgrAIR gene. The mutation was constructed by our collaborator Dr. Alka Agrawal at New England Biolabs.

The overlapping primers 5´*AGGCG*<u>*GCA*</u>*GX*TGCCGGCGATATTTTC and 5´*AC*<u>*TGC*</u>*CGCCX*TTGCATTTGAATCTACG were used to introduce the K96A (AAG → GCA) change into the sgrAIR gene. The overlapping sequence (shown in italic) is flanked on the 3´ end by a single 5-hydroxymethyluracil residue (indicated by "X") and contained the desired codon change (underlined). The entire pET21a_SgrAIR plasmid was amplified as a 7548 bp linear fragment using Phusion DNA polymerase and the overlapping primers. The 50 µl PCR reaction contained 10 ng of pET21a_SgrAIR template DNA, 0.4 mM dNTPs, 0.2 µM each primer, 3% DMSO, and 0.5 µl of Phusion Hotstart High-fidelity DNA polymerase (New England Biolabs). pET21a_SgrAI was amplified as follows: initial denaturation at 98°C for 30 sec; followed by 35 cycles of denaturation at 98°C for 10 sec, annealing at 65°C for 20 sec, and polymerization for 4 min at 72°C; with a final polymerization for 5 min at 72°C. Next, 10 µl of the reaction was digested with 20 units of DpnI restriction endonuclease (New England Biolabs) for 1 hour at 37°C in a volume of 20 µl. After incubation at 70°C for 20 min to inactivate DpnI, EndoVIII DNA glycosylase (10 units) and human SMUG1 DNA glycosylase (10 units, New England Biolabs) were added, and the reaction was incubated at 37°C for 15

min to excise 5-hydroxymethyluracil from the PCR product, followed by an additional 15 minute incubation at room temperature to allow annealing of complementary extensions. *Escherichia coli* T7 Express I^q competent cells (New England Biolabs) were transformed with 2.5 µl of the annealing reaction. Recombinants were selected by plating 50 µl of transformation reaction on LB plates containing 0.2 mg/ml ampicillin. To confirm nucleotide sequence, plasmid DNA was purified from 4 individual recombinant colonies and sequenced across the sgrAIR gene. No sequence changes, except for the anticipated K96A codon change, were observed.

Protein Expression and Purification

SgrAI mutant K96A protein was purified as described (Dunten *et al.* 2008) with some modifications. Briefly, pET21a_SgrAIR_K96A plasmid was transformed into T7 Express Competent *E. coli* cells (High Efficiency, New England Biolabs) which constitutively express the MspIM methyltransferase from the plasmid pBAKMspIM to protect the host genome. The cells were grown in five batches of 18-L fermenters at 37°C, the cells were induced at OD_{600} of 0.5 (Cary-50 Bio, Varian) with 0.4 mM IPTG, followed by incubation overnight at 17°C. Cells were harvested by centrifugation at 4,800 x g for 30 minutes using an Avanti J-20 centrifuge (Beckman Coulter), frozen in liquid nitrogen, and stored at -80°C. The frozen cell pellets were resuspended in breaking buffer (100 mM Tris (pH 8.0), 800 mM NaCl, 10 mM EDTA, 10 mM β-Me, and 100 mM PMSF) and sonicated using a Branson Sonifier 450 (VWR Scientific). The lysed cells were centrifuged for 1 hour in a Sorvall Superspeed RC2-B at 16,300 x g, followed

by centrifugation at 190,000 x g for 1 hour in a L8-70M Ultracentrifuge (Beckman Coulter). The enzyme was purified using FPLC (GE Healthcare Biosciences) chromatography and the following chromatographic resins: Heparin FF Sepharose (GE Healthcare Biosciences), SP FF Sepharose (GE Healthcare Biosciences), Q FF Sepharose (GE Healthcare Biosciences), followed by a second Heparin FF Sepharose (GE Healthcare Biosciences) chromatographic step. Finally, SgrAI enzyme was dialyzed into storage buffer (20 mM Tris-OAc, pH 8.0, 50 mM KOAc, 0.1 mM EDTA, 1 mM DTT, 50% glycerol), aliquoted into single use aliquots, flash frozen in liquid nitrogen, and stored at -80°C.

DNA Preparation

The oligonucleotides were made synthetically (Sigma Aldrich) and purified using C18 reverse phase HPLC or denaturing PAGE (Aggarwal *et al.* 1990). The concentration was measured spectrophotometrically, with an extinction coefficient calculated from standard values for the nucleotides (Fasman *et al.* 1975). The self- complementary DNA strands, or equimolar quantities of complementary DNA, were annealed by heating to 90°C for 10 minutes at a concentration of 1 mM, followed by slow-cooling to 4°Cover 4-5 hours in a thermal cycler. Because freeze-thawing altered the concentration of double stranded DNA used in the assays, DNA used for single turnover assays was treated very carefully to minimize this problem. Such DNA samples were aliquoted into small amounts, flash frozen in liquid nitrogen, stored at -20°C (in water), and used only once after removing from the freezer. DNA was 5´ end labeled with ^{32}P using T4

polynucleotide kinase (New England Biolabs) and [γ-^{32}P]-ATP (Perkin-Elmer, Inc.), and excess ATP removed using G-30 spin columns (Bio-Rad Laboratories, Inc.).

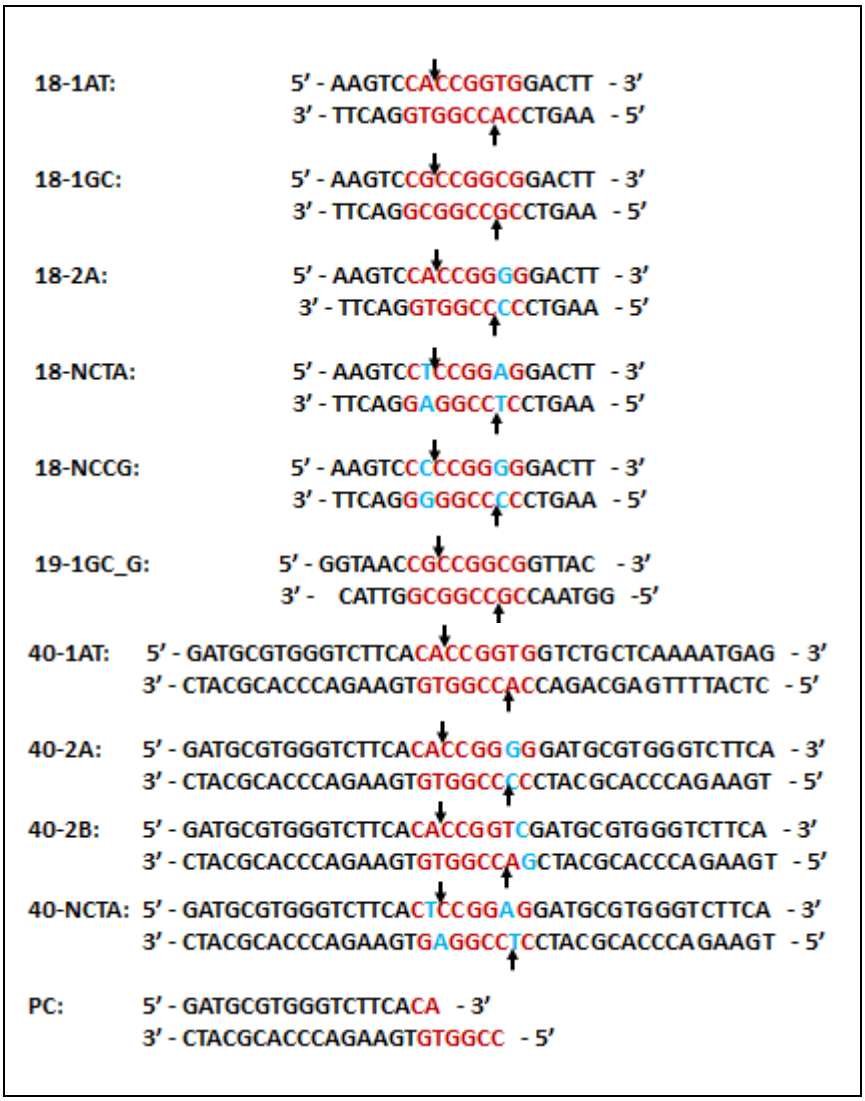


Table 3.1: DNA constructs used in determining dissociation constants, cleavage rate constants and crystal structures. Recognition sequences are shown in red, whereas the variations in secondary site sequences are shown in cyan. Arrows indicate the cleavage sites.

Binding affinity of K96A SgrAI using Fluorescence Polarization Anisotropy

The binding affinity of K96A SgrAI to various cognate and non-cognate DNA sequences was determined as described (Reid *et al.* 2001, Park *et al.* 2010b). The 5′end of both the strands of DNA duplexes were labeled with the fluorophore HEX (6-(4,7,2′,4′,5′,7′-hexachloro-(3′,6′-dipivaloylfluoresceinyl)-6-carboxamido)), obtained from Sigma Genosys. Each DNA has a 6-carbon spacer between the fluorophore and the 5′ end of the DNA. 1 nM fluorophore labeled DNA was titrated with increasing amounts of K96A SgrAI enzyme (1 nM to 1 μM) in 2 ml binding buffer (20 mM Tris-OAc pH 8.0, 50 mM KOAc, 10 mM $Ca(OAc)_2$, 1 mM DTT, and 10% glycerol). The excitation was carried out at 537 nm and the emission was measured using a 50.8 mm diameter 570 nm cut-off filter with a 580-2750 nm transmittance range and 1 mm slit widths. The experiment was carried out in a PCI (ISS) fluorometer with T format, automatic polarizers and temperature controlled at 4°C. The polarization of the emitted light as a function of added enzyme was fit to KaleidaGraph software. The data was fit to the following equation (Reid *et al.* 2001):

$$A = A_{min} + (A_{max} - A_{min}) [(P_T + O_T + K_d) - \{(P_T + O_T + K_d)^2 - (4P_TO_T)\}^{1/2}]/ (2O_T)$$

where A is the anisotropy at a given protein concentration, A_{max} is the predicted anisotropy of fully bound DNA, A_{min} is the anisotropy with no protein binding, O_T is the total DNA concentration, P_T is the total protein concentration, and K_d is the dissociation constant to be determined.

Single turn-over rate constants

The single turnover DNA cleavage rate constant was determined as described (Park *et al.* 2010b). In order to carry out the single turnover assay, the protein concentration was required to be much higher than the DNA concentration and the Kd, such that more than 99% of the DNA was bound by the protein. Normally, the protein concentration is chosen about ten-fold above the binding affinity (Kd).

Reactions were initiated by mixing 50 μl of 2 nM of 5′-end ^{32}P-labeled DNA with 50 μl containing K96A SgrAI (2 μM) to give a total of 100 μl reaction volume. The reaction was carried out at 37°C with or without the addition of PC DNA (Table 3.1) in kinetic buffer (20 mM Tris-OAc pH 8.0, 50 mM KOAc, 10 mM $Mg(OAc)_2$, and 1 mM DTT). At various time points, from 0 minutes to 24 hours, 5 μl of the mixture was withdrawn and quenched with 5 μl of quench solution (80% formamide, 50 mM EDTA, 1 mg/ml XCFF dye, and 1 mg/ml BPB dye). These quenched aliquots were then electrophoresed on a 20% denaturing polyacrylamide gel (20% 19:1 acrylamide:bisacrylamide, 4 M urea, 89 mM Tris, 89 mM boric acid, 2 mM EDTA) for 3 hours at 300 V. The gels were left to expose overnight on phosphor image plate at 4°C. The next day the gels were scanned on a Typhoon Scanner (GE Healthcare Life Sciences), and the density of each band was calculated using ImageQuant software (GE Healthcare Life Sciences). The single turnover rate constant of DNA cleavage was then determined by plotting the percentage of DNA that was cleaved vs. time using KaleidaGraph software and fitting into the following equation:

$$\text{Percentage of Product} = C_1 + C_2 * (1 - e^{-kt})$$

where C_1 is a constant fitting the baseline, C_2 is the total percent of DNA predicted to be cleaved by SgrAI, k is the cleavage rate constant, and t is the length of incubation in minutes. The data from some reactions fit poorly to a single exponential function, suggesting multiple phases. These were found to fit well to the sum of two exponential functions:

$$\text{Percentage of product} = C_1 + C_2 * (1 - e^{-k1t}) + C_3 * (1 - e^{-k2t})$$

where C_1 is a constant fitting the baseline, C_2 is the total percent of DNA predicted to be cleaved by SgrAI with rate constant k_1, C_3 is the total percent of DNA predicted to be cleaved by SgrAI with rate constant k_2, and t is the length of incubation in minutes. The rate constants are presented as k_1 and k_2, where k_1 is the greater (accelerated) of the two rate constants k_1 and k_2, and k_2 is the lower (non-accelerated). For reactions fit by a single rate constant, the rate constant is listed as accelerated if it was significantly greater than the unactivated rate constant of 0.094 ± 0.015 min^{-1}. Measurements were performed at least three independent times, and presented as the average±standard deviation.

Crystallization, Data Collection, Structure Solution, Refinement, and Analysis

The purified WT or K96A SgrAI protein was removed from -80°C storage, dialyzed overnight, and concentrated using Amicon Ultracel-10K filter centrifuged at 7,000 x g in a Jouan-BR4 centrifuge. WT or K96A SgrAI and DNA duplex were mixed in a 1:2 molar ratio for the cocrystallization experiments. The ratio was chosen to ensure that there was an excess of DNA for DNA bound dimer formation. The initial screen was

carried out using hanging drop method (McPherson 1976) with varied precipitating conditions (5 mM $CaCl_2$, 150 mM NaCl, 100 mM buffer (pH 4.5-8.5), and 5-20% PEG 4K) and 3.5-10 mg/ml protein at 17°C. Based on the initial screen, follow up trials led to optimization of the crystallization conditions. Crystals grew best in 1 to 15 weeks at 17°C using 5 mM $CaCl_2$, 150 mM NaCl, 100 mM imidazole buffer pH 6.0, and 20% PEG 4K as the precipitant and 7.5 mg/ml protein bound to 192 µM DNA. The crystals were then exchanged into a cryoprotectant solution (5 mM $CaCl_2$, 300 mM NaCl, 100 mM imidazole buffer pH 6.0, and 25% PEG 4K and 30% glycerol) and flash-frozen in liquid nitrogen. X-ray diffraction was measured using synchrotron radiation at the Stanford Synchrotron Light Source (SSRL) BL12-2. Data collection was performed while maintaining the crystal at 100K. Image processing and data reduction were performed with MOSFLM (Leslie 1999) and SCALA (Evans 2006). The structure was solved by molecular replacement using 3DVO (Dunten *et al.* 2008) for data set S506B2_K96A_GC (K96A SgrAI with the 18-1GC DNA, Tables 3.1 and 3.4) and 3N7B (Little *et al.* 2011) for data sets S501B7_WT_GC and S489C6_K96A_GC (wild type and K96A SgrAI with 19-1GC_G and 18-1GC DNA, respectively, Tables 3.1 and 3.4). The structure was refined using REFMAC (Murshudov *et al.* 1999) in the CCP4 program suit (CCP4 1994), and the model building program Coot (Emsley *et al.* 2004). All structure figures were prepared using PYMOL (DeLano 2002).

RESULTS

K96A shows stronger binding affinity for secondary site compared to primary sites

The binding affinities of K96A SgrAI to various cognate and non-cognate DNA sequences were determined using fluorescence polarization anisotropy (Table 3.2). 1 nM fluorophore labeled DNA was titrated with increasing amounts of K96A enzyme (1 nM to 1 μM) in 2 ml binding buffer at 4°C. The dissociation constant of K96A bound to 18-1(AT) was determined to be 53.4±1.4 nM, which was very similar to that with 18-1(GC), (43.1±1.9 nM). However, it was surprising to see that 18-2A had a tighter binding (22.7±3.1 nM) than either of the primary sequences bound to K96A. The dissociation constant of K96A bound to 40-1AT was determined to be 12.3±3.6 nM, indicating that the longer flanks enabled formation of tighter enzyme-substrate complex. The binding affinities of WT bound to various DNAs show similar trend with approximately 10 fold greater affinity (Table 3.2).

Table 3.2: Binding affinities of K96A and WT SgrAI with various DNA sequences at 4°C

	K96A K_D (nM)	**WT K_D (nM)***
18-1AT	53.4±1.4	2.5±0.9
18-1GC	43.1±1.9	-
40-1AT	12.3±3.6	0.9±0.2
18-2A	22.7±3.1	1.5±0.2

***** Data from Park *et al.* 2010b.

K96A SgrAI has reduced cleavage activity with secondary and non-cognate sites

The single turnover DNA cleavage rate constants of K96A SgrAI for primary and secondary site DNA were measured (Table 3.1 and Table 3.3). The DNA cleavage studies make use of ^{32}P labeled DNA of varied lengths containing either primary or secondary site sequence (Table 3.1). Cleavage studies with wild type SgrAI show that the rate of cleavage of both primary and secondary site sequences is increased when added precleaved primary site DNA (PC DNA) is added (Park *et al.* 2010). The assays contain excess SgrAI (1 μM), and only 1 nM of radiolabeled DNA, therefore the additional unlabeled PC DNA is expected to bind SgrAI and raise the concentration of SgrAI bound to PC DNA, which in turn results in accelerating the DNA cleavage of other SgrAI bound to primary or secondary site DNA (Park *et al.* 2010).

In contrast, K96A SgrAI was found to exhibit very little cleavage activity of primary (40-1AT, Table 3.1) and secondary (40-2A, 40-2B, Table 3.1) site DNA in the absence of added PC DNA, even after 24 hours at 37ºC (Table 3.3). However, in the presence of 10 nM unlabeled PC DNA, K96A SgrAI cleaves the 40 bp primary site DNA (40-1AT, Table 3.1) with a rate constant of 0.205±0.005 min^{-1} (18% of the DNA cleaved), which is increased to 6.2±0.1 min^{-1} (61% of the DNA cleaved) with 1 μM PC DNA (Table 3.3), showing a similar, although more modest, capacity for stimulation as the wild type enzyme (Park *et al.* 2010). Interestingly, K96A SgrAI shows only very limited cleavage with secondary site DNA (40-2A, CACCGGGG, Table 3.1); even in the presence of 1-2 μM added PC DNA (Table 3.3). Only 2-6% of this DNA is cleaved, and

the rate constants are limited to 0.21±0.07 and 2.2±0.9 min^{-1}). Similarly, with the other secondary site DNA (40-2B, CACCGGTC, Table 3.1) and 1-2 μM added PC DNA, K96A SgrAI exhibits DNA cleavage rate constants comparable to that of the 40-2A DNA (Table 3.1) (0.09±0.06 and 0.11±0.03 min^{-1} respectively, Table 3.3), and only about 4% of DNA is cleaved in this case (Table 3.3).

Similarly, K96A SgrAI shows very little to no cleavage of non-cognate sites either 18mer or 40mer (Table 3.3). The percentage of cleaved product is also very low along with the low cleavage rate constants. Due to low cleavage product the errors are also significantly high. Low cleavage rate constants with non-cognate DNA indicates that K96A mutant gains specificity rather than becoming more relaxed due to the loss of the lysine side chain.

Table 3.3: Single-turnover DNA cleavage rate constants using 1 μM K96A SgrAI at 37°C					
^{32}P-labeled DNA (1 nM)	**Added unlabeled [DNA]**	**Accelerated Rate Constant (min^{-1})**	**% Cleaved, Accelerated**	**Non Accelerated Rate Constant (min^{-1})**	**% Cleaved, Non Accelerated**
40-1AT	10 nM PC			0.205±0.005	18±3
	1 μM PC	6.2±0.1	61.1±0.3		
40-2A	0 nM PC			NC*****	
	1 μM PC			0.21±0.07	6±2
	2 μM PC	2.2±0.9	2.0±0.5	0.016±0.007	1.0±0.6
40-2B	0 nM PC			NC*****	
	1 μM PC			0.09±0.06	4.1±0.5
	2 μM PC			0.11±0.03	3.7±0.6
18-NCTA[§]	0 nM PC			0.0002±0.0016	2.2±13.5
18-NCCG[§]	0 nM PC			0.000075±0.005	10±657
18-NCCG[§]	1 μM PC			0.11±0.6	1.6±0.2
40-NCTA[§]	0 nM PC			0.0018±0.0044	0.5±0.3
40-NCTA[§]	1 μM PC			0.0011±0.0006	6.2±1.0

*****No cleavage of the DNA was observed after 24 hours at 37º.

[§] Based on single trial.

Similarity in crystal structure of WT and K96A SgrAI with primary site DNA

Crystallization trials were performed using the hanging drop method (McPherson 1976) to obtain crystals with suitable diffraction in order to determine the crystal structure of the WT and K96A SgrAI bound to canonical DNA. A single, large crystal was obtained with K96A SgrAI bound to 18-GC (Figure 3.2). X-ray diffraction data was collected at Stanford Synchrotron Light Source (SSRL) BL12-2. Image processing and data reduction were performed with MOSFLM (Leslie 1999) and SCALA (Evans 2006). Because the previous crystal structure of wild type SgrAI was determined with a primary site containing AT at the R_2Y_7 positions, and the crystallographic data obtained for K96A SgrAI containing primary site DNA with GC at these positions, crystals of wild type SgrAI with 18-1GC were also prepared and analyzed (S501B7_WT_GC, Table 3.4). The structures were solved by molecular replacement using a single monomer chain of SgrAI from 3DVO (Dunten *et al.* 2008) as the search model in the case of the S506B2_K96A_GC data set (Table 3.4), and a single monomer chain from 3N7B (Little *et al.* 2011) as the search model in the case of the data sets S501B7_WT_GC and S489C6_K96A_GC (Table 3.4). The structures were refined using REFMAC (Murshudov *et al.* 1999) in the CCP4 program suit (CCP4 1994), and the model building program Coot (Emsley *et al.* 2004). All structure figures were prepared using PYMOL (DeLano 2002). Diffraction data and structure refinement statistics of the wild-type (S501B7_WT_GC with 19-1GC_G DNA) and two K96A mutant SgrAI (S506B2_K96A_GC and S489C6_K96A_GC containing the 18-1GC DNA) (Table 3.1), are shown in Table 3.4.

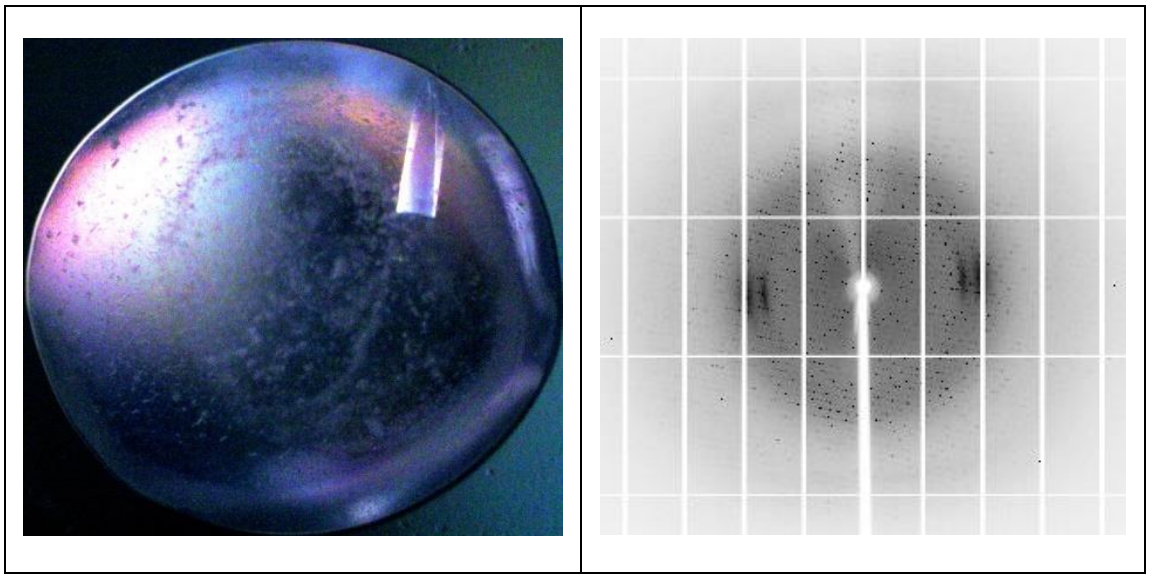


Figure 3.2: Crystal and diffraction pattern of K96A bound to 18-1GC.

Table 3.4: Diffraction Data and Structure Refinement Statistics of K96A and WT SgrAI bound to 18-1GC and 19-1GC_G, respectively.

Code	S506B2_K96A_GC	S501B7_WT_GC	S489C6_K96A_GC
Beamline	SSRL BL12-2	SSRL BL12-2	SSRL BL12-2
Processing program	MOSFLM/SCALA	MOSFLM/SCALA	MOSFLM/SCALA
DNA	18-1GC	19-1GC_G	18-1GC
Space Group	$P2_1$	$P2_12_12_1$	$P2_12_12$
Cell	58.91 Å, 117.52 Å, 62.43 Å, β=112.076°	62.13 Å, 102.7 Å, 129.98 Å	61.75 Å, 102.77 Å, 117.28 Å
Resolution	2.187 Å	1.802 Å	1.894 Å
Total Observations	119,952	266,447	248,710
Unique Observations	39,253	75,261	57,806
% Complete	96.2 % (86.1 %)	97.9 % (96.6 %)	96.6 % (91.7 %)
I/sigma	12.3 (2.1)	8.4 (1.6)	5.6 (1.8)
Multiplicity	3.1 (2.9)	3.5 (3.5)	4.3 (3.9)
Rmerge[1]	5.1 % (35.3 %)	6.6 % (46.4 %)	9.2 % (42.2 %)
Rcryst[2]	18.9 %	18.2 %	24.6 %
Rfree[3]	26.7 %	22.5 %	30.5 %
Overall B factor (Wilson plot)	37.15 $Å^2$	23.38 $Å^2$	21.70 $Å^2$
RMSD-bonds	0.018 Å	0.024 Å	0.022 Å
RMSD-angles	1.89°	2.09°	2.10°
Asymmetric unit	2 SgrAI chains & 2 DNA strands	2 SgrAI chains & 2 DNA strands	2 SgrAI chains & 2 DNA strands
Numbers of waters	345	667	501
Number of divalent cations	4 Ca^{2+}	4 Ca^{2+}	4 Ca^{2+}

[1] $R_{merge}=\sum_{hkl}(|<I_{hkl}>-I_{hkl}|)/(\sum_{hkl} I_{hkl})$ where $<I_{hkl}>$ is the average intensity over symmetry related and equivalent reflections and I_{hkl} is the observed intensity for reflection hkl.

[2] $R_{cryst}=\sum_{hkl}(||F_{obs}|-|F_{calc}||)/(\sum_{hkl}|F_{obs}|)$ where $|F_{obs}|$ and $|F_{calc}|$ are the observed and calculated structure factor amplitude for reflection hkl. The sum is carried out over the 98% of the observed reflections which are used in refinement.

[3] R_{free} refers to the R factor for the test reflection set (2% of the total observed) which was excluded from refinement.

No significant differences between the structure of SgrAI bound to CACCGGTG primary site (Dunten *et al.* 2008) and SgrAI bound to CGCCGGCG (this work, S501B7_WT_GC, Table 3.4) were found (*data not shown*). Similarly, no large differences were observed between the two K96A SgrAI structures containing CGCCGGCG primary site DNA. The electron density around the K96A mutation clearly shows the absence of lysine side chain and confirms the mutation of Lys 96 to Ala (Figure 3.3). The superimposition of structure S501B7_WT_GC and S506B2_K96A_GC (Figure 3.4-3.5) shows that the WT and K96A SgrAI also have very similar structures (RMSD = 0.293). Surprisingly, the unstacking of bases C_7-G_8 present in the WT structure is also found in the K96A SgrAI structure, despite the absence of the Lys 96 side chain (beyond the beta carbon). The distance between C_{beta} of Ala96 and C2 of base G8 is approximately 6.6 Å, seemingly too far to influence the unstacking of the bases C_7-G_8.

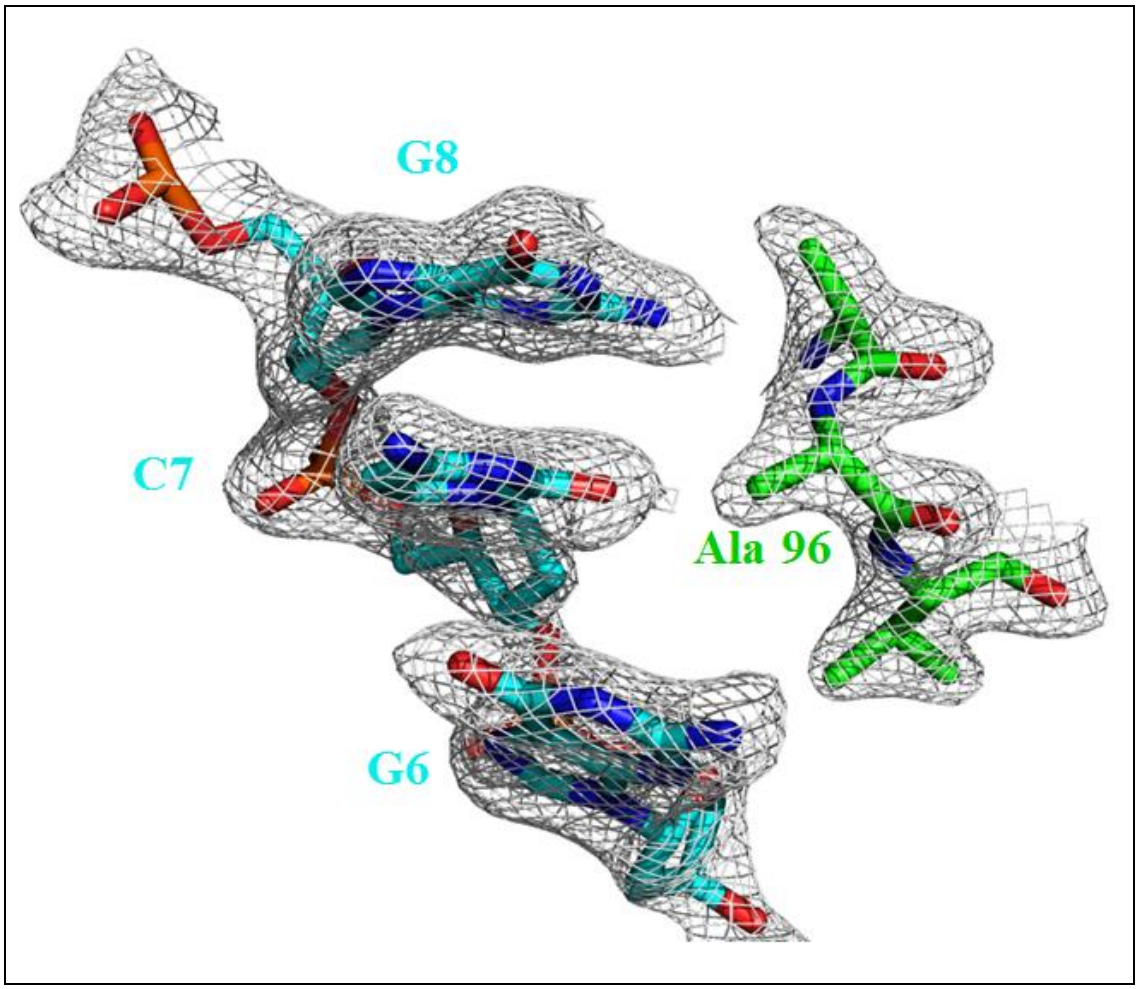


Figure 3.3: K96A SgrAI (green) electron density map showing the absence of electron density around the mutated lysine side chain. DNA is shown in cyan. Electron density contoured at 1.0 sigma.

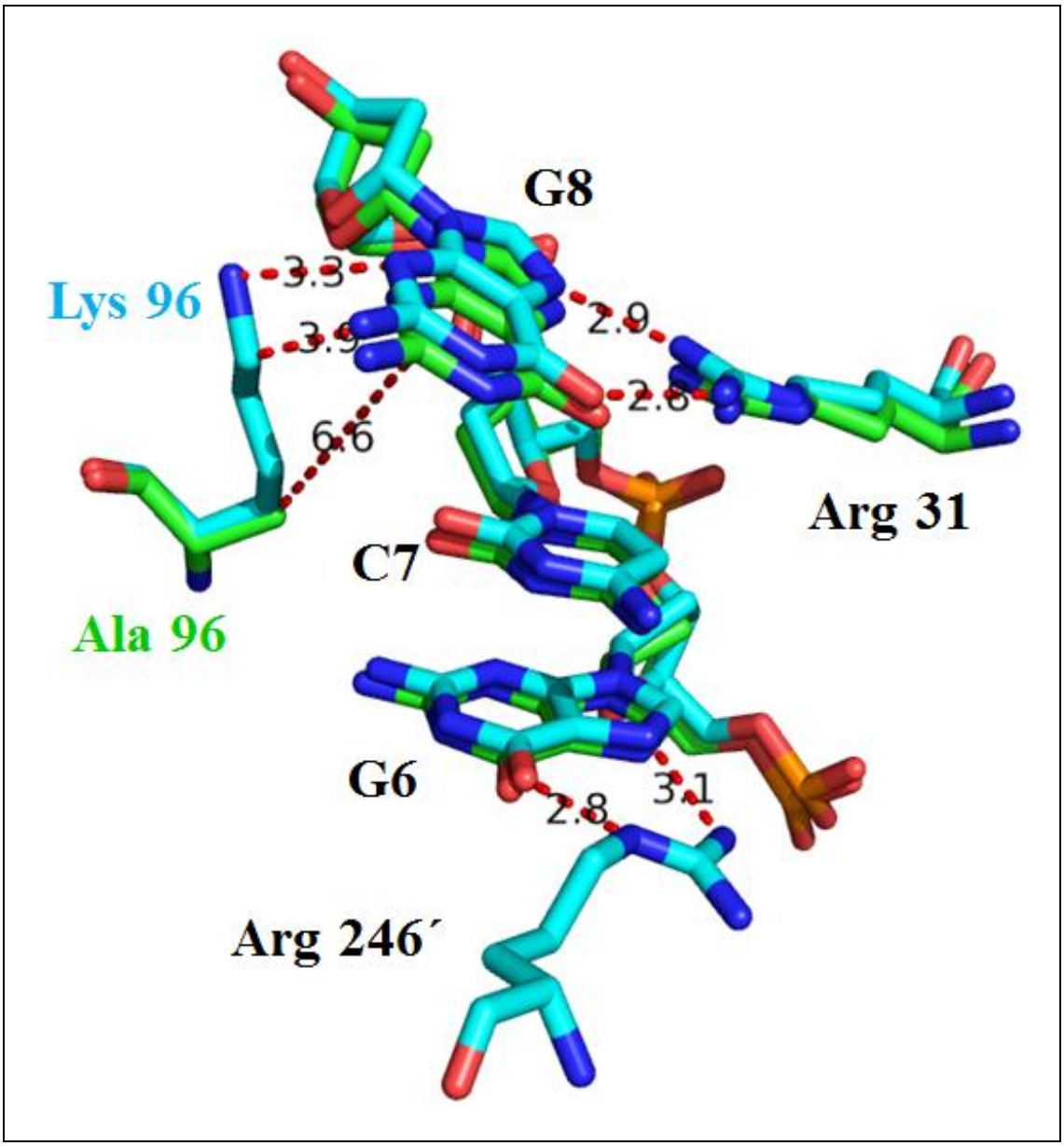


Figure 3.4: Superimposed diagram showing interactions between residue 96 (K96 in wild type SgrAI and A96 in the mutant) and oligonucleotide G6, C7 and G8. RMSD = 0.293.

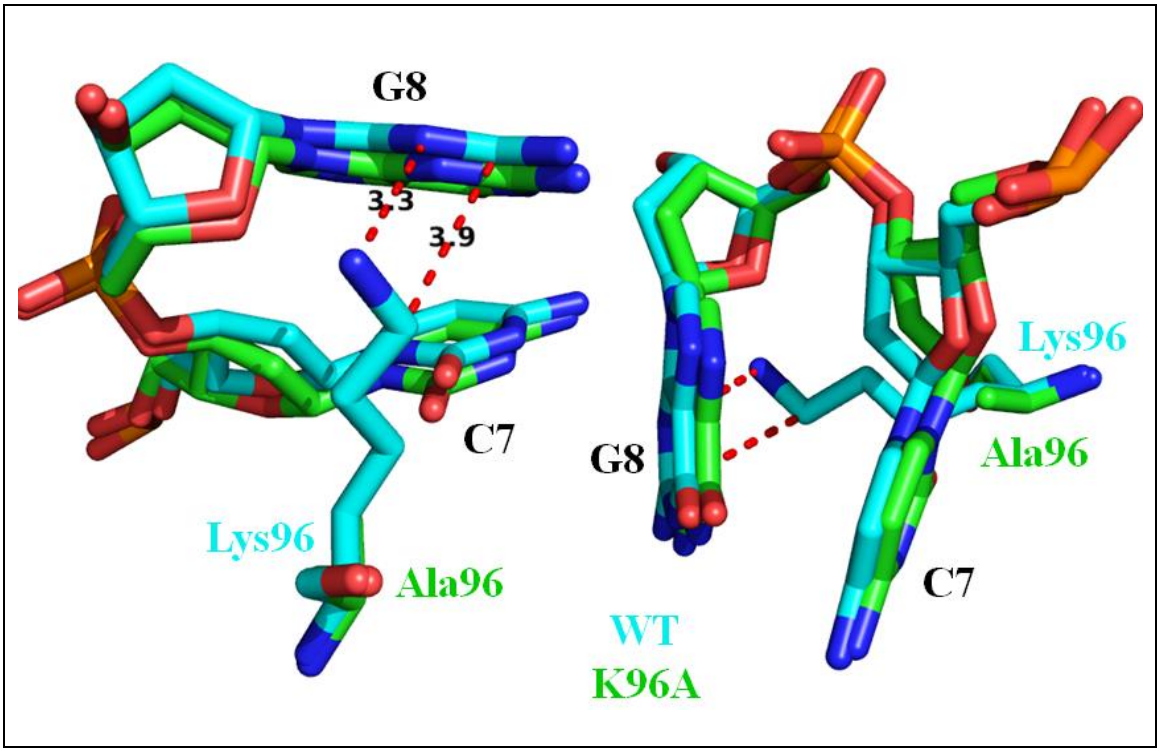


Figure 3.5: Superimposed diagram showing interactions between the residue 96 (K96 in wild type SgrAI and A96 in the mutant) and oligonucleotide C7 and G8.

DISCUSSION

Lys 96 was mutated to Ala to understand how the specificity was maintained by the enzyme for the recognition sequence at the degenerate (YR) site of the primary recognition sequence CRCCGGYG in the absence of any direct readout contacts made to these bases. It has been proposed that the specificity was provided by the distortion of DNA which was recognized by the intercalation of Lys side chain between Y_7 and G_8 (Dunten *et al.* 2008). Thus, removal of the side chain is predicted to relax the specificity at this base pair, and the enzyme is predicted to recognize any of the four bases at the degenerate sites (i.e. $C_1\underline{N_2}C_3C_4G_5G_6\underline{N_7}G_8$) equally. Binding assays and activity assays were carried out with various canonical and non-canonical DNA sequences to determine

the effects of the mutation, in order to better understand the indirect readout mechanism in SgrAI. Crystal structures of WT and K96A SgrAI bound to primary site DNA were also obtained to aid the functional studies. The crystal structures clearly show that WT and K96A SgrAI assume a similar conformation when bound to primary site DNA, and the distortion of the DNA bases at C_7G_8 persists even in the absence of the Lys side chain beyond the beta carbon. However, the mutant enzyme does exhibit altered binding and cleavage activity with primary and secondary site DNA sequences. First, the binding affinity to both types of sites is greatly reduced relative to that of wild type (20-100 fold in the case of 18-1AT, 200 fold in the case of 40-1AT, and 10-20 fold in the case of 18-2A, Table 3.2), which is likely due to the removal of two positive charges in the SgrAI dimer. Second, the DNA binding sequence preference appears to be reversed by the mutation, for example, K96A binds two fold more tightly to the secondary site 18-2A compared to the primary site of the same length, 18-1AT (Table 3.2). This is in contrast to the WT enzyme; where there appears to be very little difference in the binding affinity to these two sequences (Park *et al.* 2010). Finally, the DNA cleavage rate constants measured for K96A SgrAI also show differential effects of the mutation with the two types of DNA sequences. With primary site DNA, K96A SgrAI shows activated DNA cleavage (i.e. in the presence of PC DNA) that is only 2-3 fold reduced relative to WT (Table 3.3). However, almost no cleavage of secondary site or non-cognate DNA is observed with K96A SgrAI (Table 3.3). The wild-type enzyme cleaves secondary site DNA nearly as rapidly as it cleaves primary site DNA under fully activated conditions (Chapter 2). Therefore, K96A SgrAI shows DNA cleavage behavior similar to the wild

type enzyme on primary site DNA, but very different behavior with secondary site and non-cognate DNA, making the enzyme more specific rather than less. A low cleavage rate constant with secondary site and non-cognate DNA indicates that K96A mutant gains specificity rather than promiscuity due to the loss of the lysine side chain. The similar structure of K96A and wild type SgrAI bound to primary site is consistent with the observed activity on primary site DNA. It appears that the K96 side chain is not required by SgrAI to recognize its primary site sequence CRCCGGYG, however, this side chain may be critical for specificity expansion. In the absence of the K96 side chain, SgrAI appears incapable of cleaving secondary site sequences CACCGGGG and CACCGGTC or non-cognate sites CTCCGGAG and CGCCGGCG. Perhaps the YG of the primary site unstacks readily, such that the lysine side chain is not needed, while the more difficult GG and TC base steps require the side chain for proper DNA distortion and optimal alignments of the protein DNA interface. The structure of K96A SgrAI bound to secondary site DNA would be useful to assess this possible mechanism. A structure of SgrAI in the activated form bound to primary and secondary site DNA is especially needed, with wild type and K96A SgrAI, to understand how the K96A mutation prevents SgrAI bound to secondary site from achieving the activated conformation productively. Efforts to obtain the structure of wild type SgrAI with primary site DNA in the activated form are currently in progress, and are left to future studies.

CHAPTER 4

ROLE OF MK0566 IN NUCLEOTIDE EXCISION REPAIR

ABSTRACT

Nucleotide excision repair (NER) is one of the major DNA repair pathways carried out by all living organisms. Defects in NER in humans may lead to diseases such as xeroderma pigmentosum (XP), Cockayne's syndrome (CS), and trichothiodystrophy (TTD). The archaeal DNA repair pathways show similarities to both bacterial and eukaryotic processes. Some archaea have both UvrABC (bacterial repair system) and homologues of the eukaryotic NER pathway. However, most of the hyperthermophilic archaea, including *Methanopyrus kandleri,* do not have UvrABC and are also missing homologs of XPA and XPC. Since these proteins are required for initial damage recognition, it is unclear how DNA damage is recognized in any NER type of pathway in archaea. The NER pathway has not been well studied in archaea, and in fact it is not known if an NER pathway is even operative in these organisms. Mk0566, from *M. kandleri,* is a homolog of human FEN-1. While humans have several FEN-1 homologues, including the NER protein XPG, *M. kandleri* has only Mk0566. The different human FEN-1 enzymes have specialized roles in replication, recombination, and repair, which the singular Mk0566 may be required to perform. Mk0566 protein was prepared recombinantly in order to address its role in M. kandleri, by determining its structure and substrate specificity. DNA cleavage experiments with potential DNA substrates, representing substrates that would be found in the different potential biological roles were

performed by Amanda Stiteler for her senior thesis. Analytical ultracentrifugation and x-ray crystallographic analysis was also performed and described herein. These studies showed that Mk0566 is a 42 kDa monomeric protein with specificity towards 5´ flap and blocked flap endonuclease activity. These substrates are more typical of those of FEN-1, occurring in replication and recombination pathways. However, unlike the NER pathway protein hXPG, it does not cleave a bubble substrate. The three dimensional crystal structure of Mk0566 showed a structure very similar to FEN-1 and previous archaeal FEN-1 like proteins.

INTRODUCTION

Nucleotide Excision Repair

DNA nucleases perform various biological functions such as replication, recombination, and repair. DNA nucleases have high specificity and efficiency in performing activities such as endonucleolytic cleavage of double and single stranded DNA, and 5´ or 3´ exonucleolytic cleavage. Failure to perform these critical functions properly leads to diseases such as premature aging and premature cancer development (Thoms *et al.* 2007). It could also lead to autoimmune diseases (Nagata 2007) and repeat expansion diseases, a group of human genetic disorders caused by long and highly polymorphic tandem repeats (Fry *et al.* 1999, Parrish *et al.* 2006). One of the major DNA repair pathways carried out by DNA nucleases is called nucleotide excision repair (NER).

In general, NER takes place when bulky DNA damage leads to a distorted DNA helix. The NER mechanism can remove a wide range of DNA damage by dual incisions on both sides of the lesion (Petit *et al.* 1999). It can also be triggered by UV-induced DNA damage, such as cyclobutane-pyrimidine dimers or pyrimidine-(6-4)-pyrimidone photoproducts, as well as polycyclic aromatic carbohydrates found in tobacco smoke and crosslinking agents. Cisplatin, used to treat prostate cancer, is one such crosslinking agent (Mitchell *et al.* 1989, Buschta-Hedayat *et al.* 1999, Wood 1999). NER is a versatile and universal mechanism found in all living organisms tested so far, from the smallest free-living life form *Mycoplasma genitalium* to humans (Sancar 1996, Petit *et al.* 1999). The NER mechanism was recognized in the 1960s (Setlow *et al.* 1964, Boyce *et al.* 1964,

Pettijohn *et al.* 1964) when the UV-induced damage of thymine dimers were found to be repaired in *E. coli*.

In *E. coli*, only three proteins, UvrA, UvrB and UvrC, are needed to perform NER (Petit *et al.* 1999). The $UvrA_2B$ complex locates the lesion by tracking along the DNA. Formation of transient $UvrA_2B$-DNA complexes kinks and unwinds the DNA in an ATP-dependent reaction. UvrA, the molecular matchmaker, dissociates allowing formation of a stable UvrB-DNA complex. Once UvrC binds to the UvrB-DNA complex, UvrB makes the 3′ incision first, followed by the 5′ incision made by UvrC. UvrC makes the 5′ incision shortly (Sancar 1996, Petit *et al.* 1999, de Laat *et al.* 1999). In contrast, human NER requires 18 different proteins, including 7 xeroderma pigmentosum (XPA to XPG) proteins, the trimeric replication protein A (RPA), and the multisubunit (7-10) general transcription factor (TFIIH), to repair such damage (Sancar 1996).

Defects in NER in humans may lead to diseases such as xeroderma pigmentosum (XP), Cockayne's syndrome (CS), and trichothiodystrophy (TTD). Symptoms of XP are photodermatoses and neurological abnormalities. Photodermatoses include extreme photosensitivity to sunlight with manifestations ranging from erythema to xerosis and skin atrophy with a high probability of acquiring skin cancer. Neurological symptoms include mental retardation, progressive ataxia, and deafness (Cleaver *et al.* 1989). CS patients can suffer from cachectic dwarfism, photosensitivity, mental retardation, and progressive neurological symptoms caused by demyelination (Nance *et al.* 1992). TTD patients have sulfur-deficient brittle hair and suffer from dental caries, ichthyosis, skeletal

abnormalities, and progressive mental retardation (Stefanini *et al.* 1993, de Laat *et al.* 1999).

The NER pathway has not been studied well in archaea. The only published NER-like patch repair pathway that has been studied is that from *M. thermautotrophicus*, which has a complete set of UvrABC genes (Ogrunc *et al.* 1998). In contrast, many other archaea including *M. kandleri* do not possess the prokaryotic NER, and have only limited eukaryotic NER homologues. Therefore, the study of NER-like repair pathway in *M. kandleri* is very important to shed light on the functions of the homologous proteins, to understand the repair pathway and how it relates to bacterial and eukaryotic repair pathways evolutionarily. Studying NER in archaea may provide an opportunity to better understand the more complex version of the eukaryotic NER system. This might also aid in learning the evolutionary aspects of the DNA repair. Moreover, for practical purposes, archaeon enzymes are easier to express, purify, and crystallize than the eukaryotic equivalent.

Relationship between FEN-1, XPG and Mk0566

FEN-1 and XPG are structure specific nucleases and they belong to the FEN-1 family (Lieber 1997). FEN-1 plays a vital role in DNA replication in the processing of Okazaki fragments, and in base excision repair (Liu *et al.* 2004, Henneke *et al.* 2003). FEN-1 is highly expressed in all proliferating tissues, and is even expressed at higher levels in many cancers directly related to tumor aggressiveness (Finger *et al.* 2010). Thus, FEN-1 specific inhibitors have potential in chemotherapy (Tumey *et al.* 2005). FEN-1 is

also implicated in mitochondrial genome maintenance (Kalifa *et al.* 2009). XPG plays an essential role in nucleotide excision repair by cleaving on the 3´ end of the DNA damage (Chapter 1- Figure 1.9, Gillet *et al.* 2006, Friedberg *et al.* 2005). It also plays structural role in NER complex formation since it is required by ERCC1-XPF to make the 5´ end incision (Wakasugi *et al.* 1997, Mu *et al.* 1997). It cleaves DNA structures that contain ss/ds DNA junctions, including bubbles and splayed arm substrates (O'Donovan *et al.* 1994, Cloud *et al.* 1995). FEN-1 cleaves 5´ flap in sequence independent manner using its divalent metal ion-dependent phopshodiesterase activity (Tomlinson *et al.* 2010). However, FEN-1 only cleaves substrates that have free 5´ single strand DNA ends (Robins *et al.* 1994, Murante *et al.* 1995). FEN-1 nuclease family members have conserved N-terminal domain (N) and internal domain (I) (Lieber 1997). The N and I domains are separated by <70 amino acids in most of the FEN-1 family members, however, N and I domains in XPG are separated by more than 600 amino acids known as "spacer region". The domain regions form a helical arch which lies above the active site and contribute to DNA binding and catalysis (Storici *et al.* 2002, Scherly *et al.* 1993).

Sequence alignment of Mk0566 and human XPG shows that there are two conserved N domains in human XPG (Figure 4.1). The sequence 1-150 of human XPG has sequence identity of 29% and an E value of $1x10^{-4}$ and the sequence 711 to 996 has sequence identity of 26% and an E value of $2x10^{-15}$ compared to Mk0566. Similarly, the human FEN-1 sequence has 42% sequence identity and E value of $3x10^{-69}$ with Mk0566. This suggests that Mk0566 is more closely related to human FEN-1 than to human XPG.

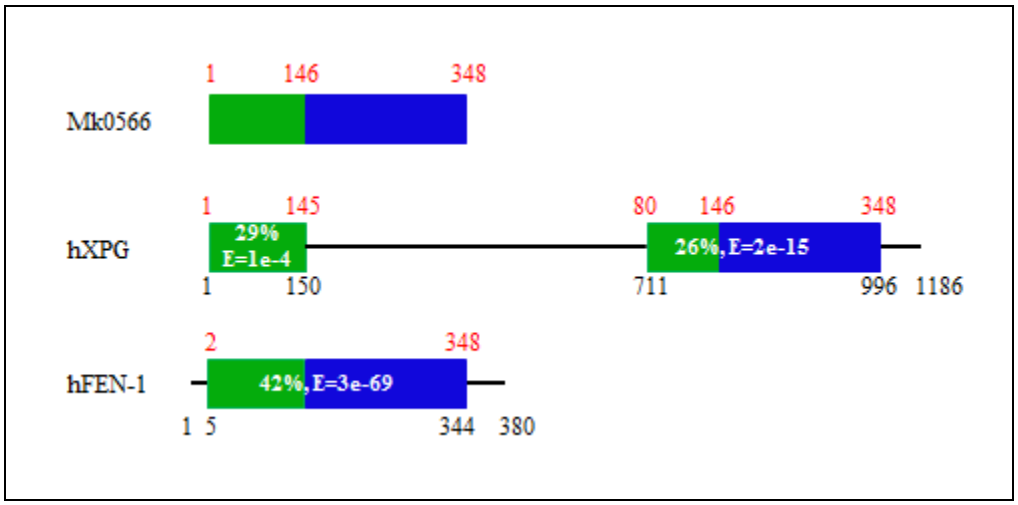


Figure 4.1: Sequence alignment of Mk0566 from *Methanopyrus kandleri* with hXPG and hFEN-1 from *Homo sapiens*. The green rectangles are the conserved N-terminal domain and the blue rectangles are the conserved internal domain. Percent refer to percent identity between the aligned regions, and E refers to the expect value. Numbers in red are sequence number corresponding to Mk0566. Sequence alignment by Prof. Nancy C. Horton.

At the time of the work described in this chapter, the only crystal structures of FEN-1 enzymes were those from *Archaeoglobus fulgidis* (Chapados *et al.* 2004), *Methanococcus jannaschii* (Hwang *et al.* 1998), *Pyrococcus furiosis* (Hosfield *et al.* 1998), *Pyrococcus horikoshii* (Matsui *et al.* 2002), and human (Sakurai *et al.* 2005). However, none of the structures contain bound DNA, with the exception of that from *A. fulgidis*. However, the DNA in that structure was not the complete substrate, nor did it occupy the full DNA binding site of the enzyme. Therefore the question remained as to how the flap portion of the DNA substrate bound the enzyme, including whether or not it threaded through an opening in the protein termed the "arch". Superimposition of the structure of T4 RNase H with a DNA substrate onto AfFEN-1 suggested that the 5´ flap threaded through a structure formed by the "thumb" in Figure 4.2 termed the "arch", although the arch is disordered in the actual crystal structure. Therefore, we sought to

determine the crystal structure of Mk0566 with substrate DNA fully occupying its active site and DNA binding sites, to better understand the origin of the enzyme's specificity for particular DNA substrates.

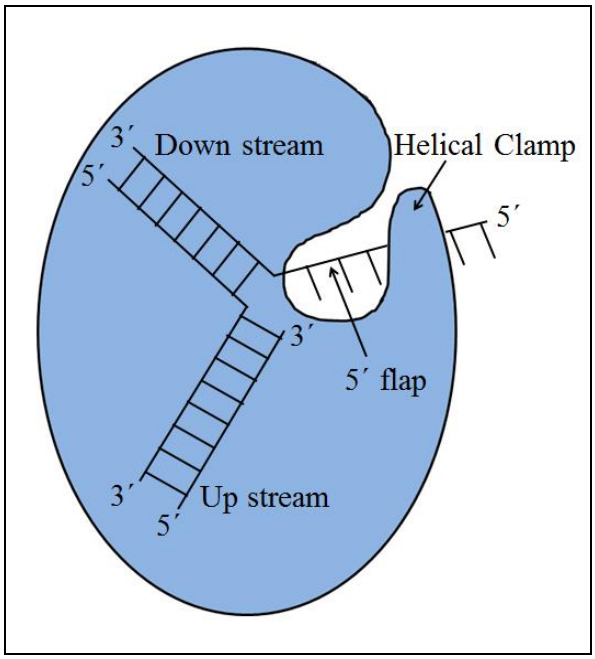


Figure 4.2: Cartoon showing interaction of *A. fulgidus* FEN-1 and DNA substrate Adapted from Chapados *et al.* 2004.

More recently, x-ray crystal structures of human FEN-1 bound to DNA show full binding of the single-stranded 5′ flap DNA, and that 5′ flap threads through the helical arch in the active site (Tsutakawa *et al.* 2011). The DNA recognition is carried out in a structure-specific, sequence independent manner by recognizing 5′ flap DNA by its ability to form a sharp 100° bend with dsDNA, thus, restricting only dsDNA with a break or a flap to be recognized. These are the only type of dsDNA capable of bending to that extent at a single phosphodiester link. The conserved motifs bind and open dsDNA, direct the target region into the helical gateway, allowing only the damaged ssDNA into

the active site which then gets cleaved out. The binding of the DNA allows the secondary structures of FEN-1 to be more ordered indicating a substrate-induced, disorder-to-order transition (Tsutakawa *et al.* 2011).

Prior results (from the senior thesis of Amanda Stiteler):

DNA cleavage assays were performed on ^{32}P 5´-labeled substrates (Figure 4.3 left and Figure 4.4). The assays contained 10 nM DNA with varied concentrations of Mk0566 and carried out at 55°C for 2 hours. Figure 4.3 right shows an example of cleavage by Mk0566, where only the top strand containing the 5´ flap is radiolabeled (* indicate radiolabeled strand in Figure 4.3, left). From Figure 4.3, right it is evident that the quantity of 20 nt cleaved product increased, and the 50 nt reactant decreased as the concentration of Mk0566 was increased. The dsDNA substrate is also visible despite the denaturing urea in the polyacrylamide gel (this was later eliminated by increasing the formamide concentration in the loading buffer and the gel), however it is also seen to decrease as the DNA is cleaved (Figure 4.3, right). Figure 4.4 summarizes the observed specificities of Mk0566, and compares them to those of human XPG and human FEN-1. XPG and FEN-1 cleave all four types of substrates whereas Mk0566 cleaves only 5´ flap and blocked flap DNA, but not pseudo-Y or bubble DNA.

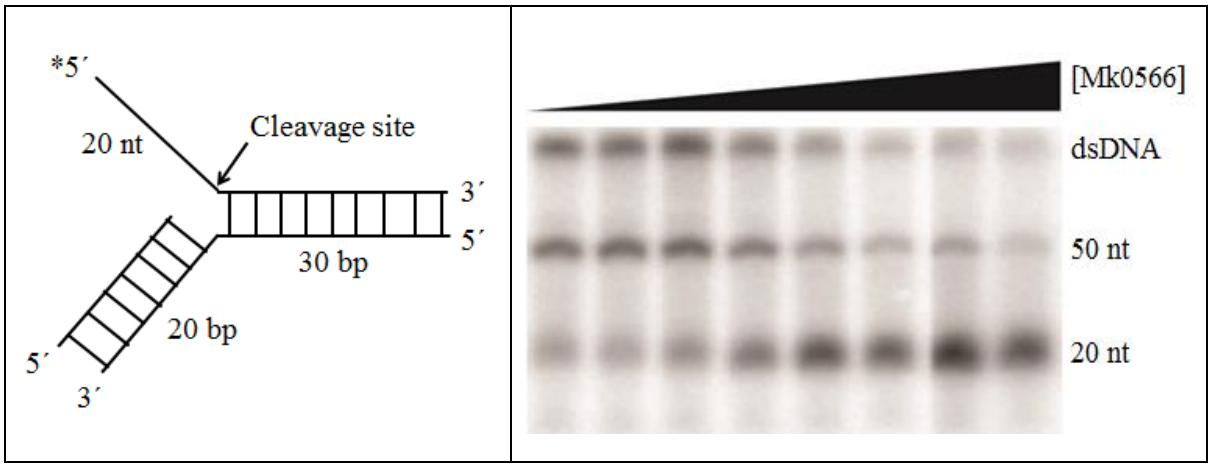


Figure 4.3: Mk0566 shows cleavage activity on 5´ flap DNA. (Left) 5´ flap DNA construct, the * indicates the strand with ^{32}P label. (Right) Denaturing gel showing cleavage of 5´ flap DNA with increasing concentration of Mk0566.

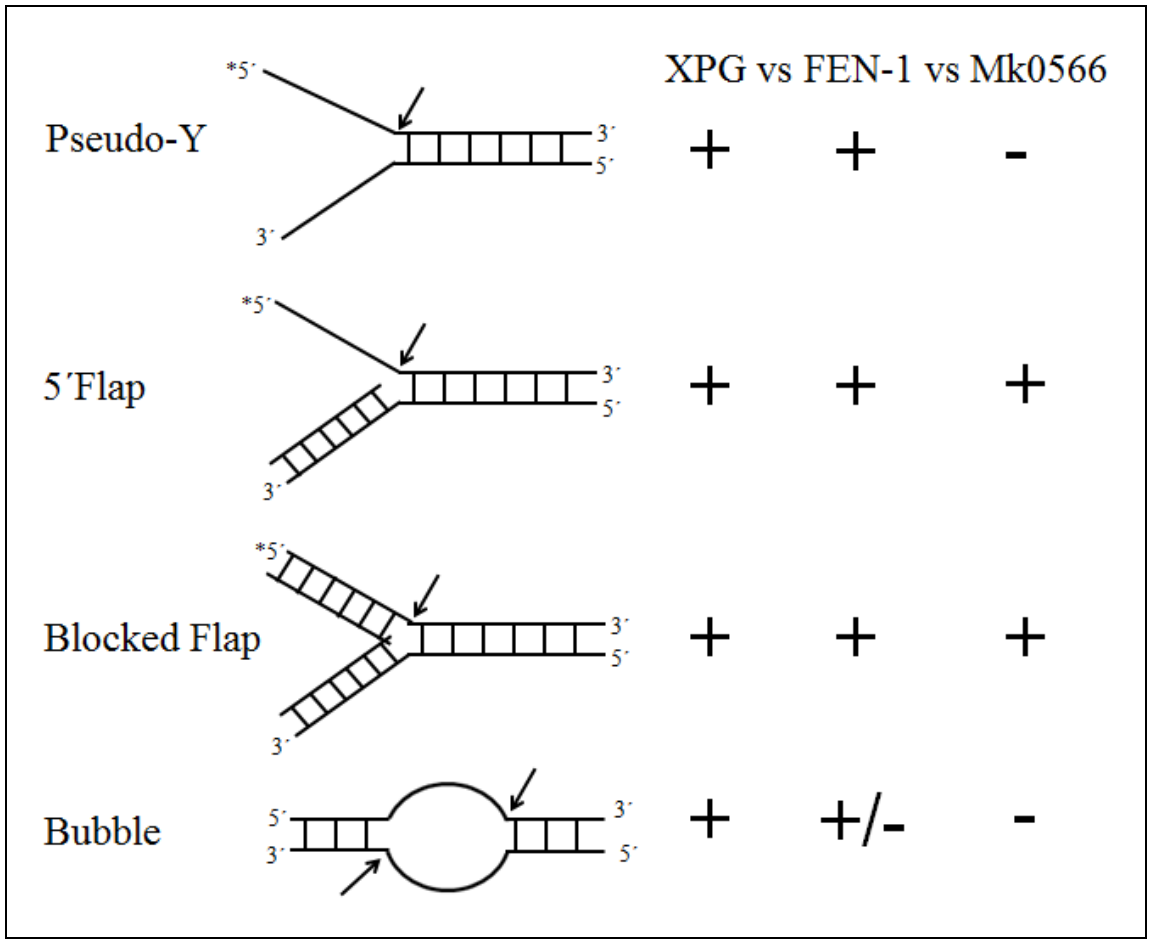


Figure 4.4: Cleavage specificity comparison of XPG, FEN-1 and Mk0566. Mk0566 cleaves 5´ flap and blocked flap DNA but not pseudo-Y or bubble DNA.

Materials and Methods:

Protein Purification

A synthetic codon optimized gene for Mk0566 (coding sequence based on *Methanopyrus kandleri* genome sequence, Genbank number AE009439.1) was purchased (Biobasic, Inc.) already placed in the pMAL-c4X expression vector (New England Biolabs, Inc.). This vector is controlled by the lac operator, and therefore can be induced by the addition of the gratuitous inducer IPTG. The protein is expressed as a fusion with

an N-terminal *E. coli* maltose binding protein (MBP) separated from the Mk0566 protein amino acids by a linker containing the cleavage site for the protease Factor Xa. The expression plasmid was transformed into Top 10 *E. coli* cells (F^- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80l*acZ*ΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*) 7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G λ-) grown overnight in a 3 ml culture, and then mixed with glycerol, flash frozen, and stored at -80°C. Cells from the glycerol stock (Horton lab glycerol stock #217) were later streaked on an LB agar plate containing 50 µg/ml ampicillin and grown overnight at 37ºC. A 3-ml culture was grown from a single colony and used to grow an 18-L culture at 37°C. Expression of Mk0566 was induced with 0.4 mM IPTG when the OD_{600} of the culture reached 0.5 (Cary-50 Bio, Varian), followed by overnight incubation at 37ºC. The cells were harvested by centrifugation using an Avanti J-20 centrifuge (Beckman Coulter) at 4,800 x g for 30 mins. The cells were flash frozen using liquid nitrogen and then stored at -80°C until needed. Mk0566 was purified by first resuspending the frozen cell pellets in amylose buffer A (50 mM Tris-HCl pH 7.5, 2 M NaCl, 2 mM 2-mercaptoethanol and 100 µM PMSF) and sonicated using Branson Sonifier 450 (VWR Scientific). The lysate was centrifuged using a Sorvall Superspeed RC2-B centrifuge, and then batch purified using amylose resin (New England Biolabs, Inc.). The protein was eluted from the resin using amylose buffer B (50 mM Tris-HCl pH 7.5, 50 mM NaCl, 30 mM maltose and 2 mM 2-mercaptoethanol. The maltose binding protein (MBP) tag was removed by first incubating with Factor Xa protease (New England Biolabs), followed by a second round of amylose resin chromatography. FPLC with Q-Sepharose fast flow column, followed by a Butyl fast flow was used to further

purify the protein (Figure 4.6). The protein was assessed to be >99% pure via coomassie stained SDS-PAGE (Figure 4.7). Western blot using anti-MPB antibody (New England Biolabs Inc.) was used to monitor the purification as well as pre and post cleavage steps. The purified protein was dialyzed into storage buffer (20 mM Tris-OAc pH 8.0, 50 mM KOAc, 1 mM EDTA, 1 mM DTT, 50% glycerol), aliquoted, and stored at -80°C.

DNA preparation

The oligonucleotides were made synthetically and purified using C18 reverse phase HPLC or denaturing PAGE (Aggarwal *et al.* 1990). The concentration was measured spectrophotometrically, with an extinction coefficient calculated from standard values for the nucleotides (Fasman *et al.* 1975). The self-complementary DNA strands, or equimolar quantities of complementary DNA, were annealed by heating to 90°C for 10 minutes at a concentration of 1 mM, followed by slow-cooling to 4°C over 4-5 hours in a thermocycler.

For crystallization trials, Mk DNAs (Table 4.1, Figure 4.5) were prepared. Variations of Mk DNAs were made by truncating one or two nucleotides off of the 3´ or 5´ ends of either Mk1, Mk2 or Mk3 to obtain optimum length of DNA for crystallization. The DNA sequences used are shown below:

Mk1: 5′ -GCTTGCAGGTGCTCGTCC- 3′ Mk2: 5′ -GTACCTGCAAGC- 3′ Mk3: 5′ -GGACGAGCC- 3′

Table 4.1: DNA sequences used for setting up crystallization trials.

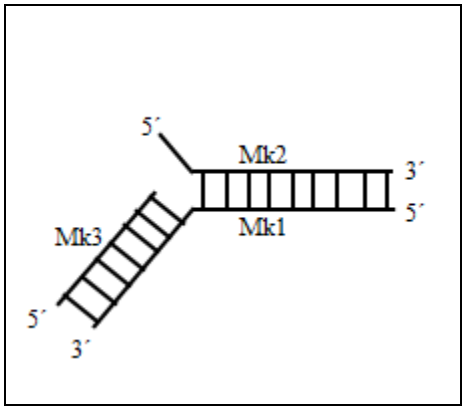


Figure 4.5: DNA construct used for setting up crystallization trials.

Sedimentation Velocity (performed by Dr. Chad K. Park)

Sedimentation velocity experiments were performed to determine the oligomeric form of Mk0566 in the presence and absence of DNA using a Beckman Coulter XL-I instrument with monochromator and interference scanning optics (632 nm), automated scanning capability, and a Ti-50 rotor. Approximately 400 µl of sample containing 6 µM Mk0566 was loaded in one of the sectors of the two-sector sedimentation velocity cells. The other sector was loaded with 425 µl of buffer (20 mM Tris-OAc (pH 8.0), 150 mM NaCl, 10 mM $Ca(OAc)_2$ and 1 mM DTT). The sample was allowed to equilibrate at 4°C for at least 1 hour in the mounted rotor. The sample was then spun at 40,000 rpm

(115,000 x g) and absorbance scans were taken continuously at 280 nm for 10 hours. Data obtained from the scans were fit to a sedimentation coefficient distribution, c(s), using SEDFIT (Schuck *et al.* 2000). The viscosity of sample buffers were measured by Cannon-Fenske viscometer and density was measured using the Mettler-Toledo densitometer.

Sedimentation Equilibrium (performed by Dr. Chad K. Park)

Sedimentation equilibrium experiments were done as in sedimentation velocity experiment with some modifications: 110 µl of 6 µM Mk0566 were loaded into three chambers of a six-chamber cell, and buffer was loaded into the other three chambers. The samples were allowed to equilibrate at 4°C for at least 1 hour in the mounted rotor. The samples were spun at three different speeds: 12,600, 18,100 and 23,700 rpm for 72-96 hours. The absorbance scans were taken at 280 nm every 4 hours until there was no observable trend in the difference between subsequent runs. Data obtained from the scans were then fit using SEDPHAT (Vistica *et al.* 2004). The density was measured using the Mettler-Toledo densitometer.

Crystallization, Data Collection, Structure Solution, Refinement, and Analysis

The hanging drop vapor diffusion method (McPherson 1976) was used to screen crystallization conditions for Mk0566. Drops were composed of 1.5 µl Mk0566 protein at varied concentrations mixed with a 1.5 molar excess of DNA and 1.5 µl of the precipitant solution. The drop was mixed on a cover glass, and then placed over 1 ml of precipitant solution in the well. Crystallization trials were set up with varying PEG 4K (Hampton

Research, Inc.) concentration from 5–25% and the pH from 4.5–8.5. After several subsequent optimizations, crystals were obtained using 15% PEG 4K, 100 mM Tris buffer pH 8.5, 150 mM NaCl, 10 mM $CaCl_2$ as the precipitating solution, and with MkL8R9 DNA (a variant of Mk DNA with one less nucleotide on the 3´-end of Mk2 and 5´-end of Mk1). Crystals reached full size in about 2 weeks at 17°C. The crystals were then exchanged into a cryoprotectant solution (25% PEG 4K, 100 mM Tris buffer pH 8.5, 300 mM NaCl, 10 mM $CaCl_2$ and 30% glycerol) and flash-frozen in liquid nitrogen. X-ray diffraction was measured using synchrotron radiation at the Stanford Synchrotron Radiation Lightsource (SSRL) BL9-2. Data collection was performed while maintaining the crystal at 100K. Image processing and data reduction were performed with MOSFLM (Leslie 1999) and SCALA (Evans 2006) respectively. The structure was solved by molecular replacement using PHASER (Storoni *et al.* 2001) by Dr. Pete Dunten at SSRL. 1RXW (FEN-1 from *Archeoglobus fulgidis*, contains 52% sequence identity with thumb domain deleted (Chapados *et al.* 2004)) was used as search model in molecular replacement. Refinement was carried out using PHENIX (Adams *et al.* 2002) and the model was built using Coot (Emsley *et al.* 2004) starting with the 1RXW model. All structure figures were prepared using PyMOL (DeLano 2002). Electrostatic surface was drawn using PyMOL using “Show” > “Surface” and coloring by element.

Results:

Purification of Mk0566

Mk0566 was purified using amylose resin (New England Biolabs, Inc.), and Q-Sepharose (GE, Inc.) and Butyl fast flow columns (GE, Inc.). The MBP tagged Mk0566 was first bound to the amylose resin, and then eluted using a buffer containing 30 mM maltose. The MBP tag was separated from Mk0566 using Factor Xa protease, followed by a second round of amylose resin batch purification which allowed the cleaved Mk0566 protein to flow through while the retaining MBP bound to the resin. The pre- and post- cleavage protein is shown in Figure 4.6. The Mk0566 protein was further purified using FPLC and Q-Sepharose and Butyl fast flow columns in sequence to obtain >99% pure Mk0566 as shown in Figure 4.7.

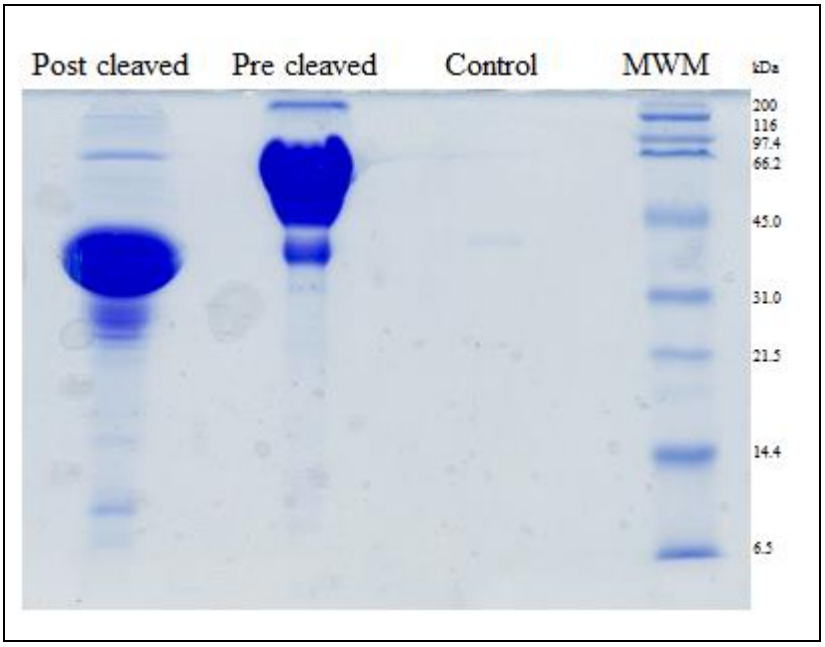


Figure 4.6: SDS-PAGE gel showing purified Mk0566 pre and post cleaved by Factor Xa. Previously purified Mk0566 (Lot 2, purified by Amanda Stiteler and Dr. Andrea Babic) was used as positive control. MWM refers to molecular weight marker.

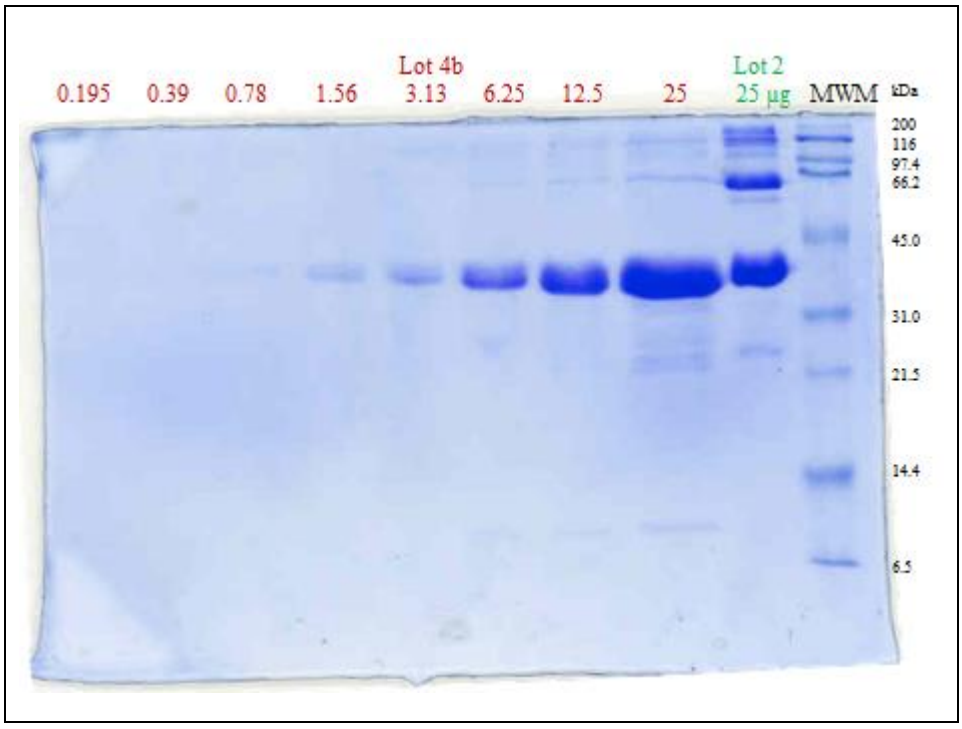


Figure 4.7: SDS-PAGE purity gel of Mk0566 (Lot 4b) showing increasing amount of protein 0.195-25 µg. Previously purified Mk0566 (Lot 2, purified by Amanda Stiteler and Dr. Andrea Babic) was used as positive control. MWM refers to molecular weight marker.

Sedimentation Velocity & Equilibrium

In order to characterize Mk0566, sedimentation velocity and equilibrium experiments were performed using an analytical ultracentrifuge. Sedimentation velocity experiments were run at 40,000 rpm, whereas the sedimentation equilibrium experiments were run at three different speeds: 12,600, 18,100 and 23,700 rpm. Data from the sedimentation velocity run was fit to sedimentation coefficient, c(s), and boundary spreading allowed determination of coefficient of diffusion, D, using SEDFIT (Schuck *et al.* 2000). Taking measured density as well as viscosity into consideration, the enzyme

was determined to be 42.0 kDa which is the predicted monomeric weight of Mk0566 (Figure 4.8). Similarly, global fitting of data obtained from sedimentation equilibrium determined the molecular weight of the enzyme to be 39,471 kDa (Figure 4.9) which is very close to the one predicted from the amino acid sequence of the enzyme (40,251 kDa).

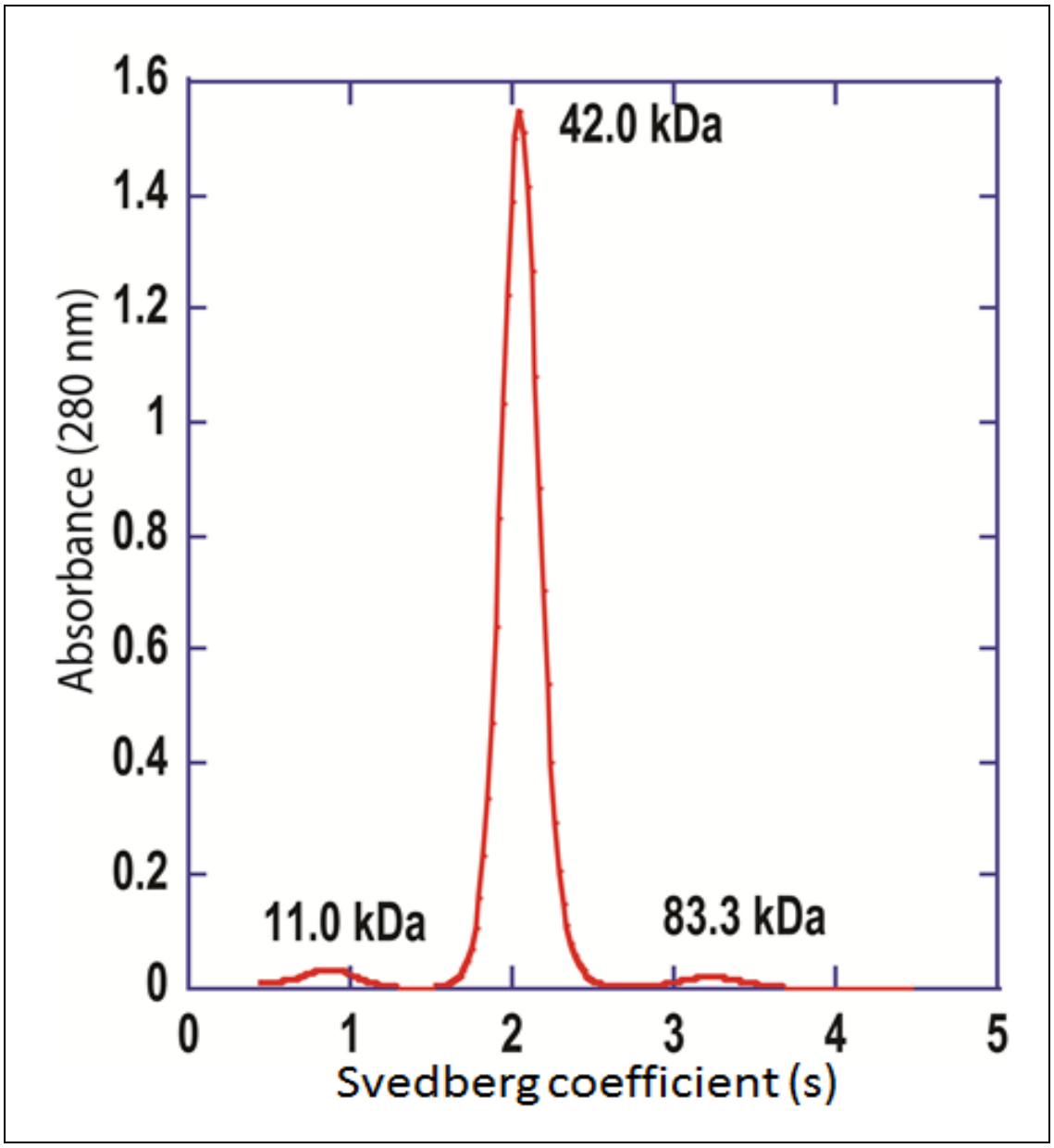


Figure 4.8: Sedimentation velocity of Mk0566 using analytical ultracentrifugation. Molecular weight was calculated by fitting sedimentation diffusion and the shape parameter by measuring density and viscosity. Figure courtesy of Dr. Chad K. Park, Analytical Biophysics Core Facility.

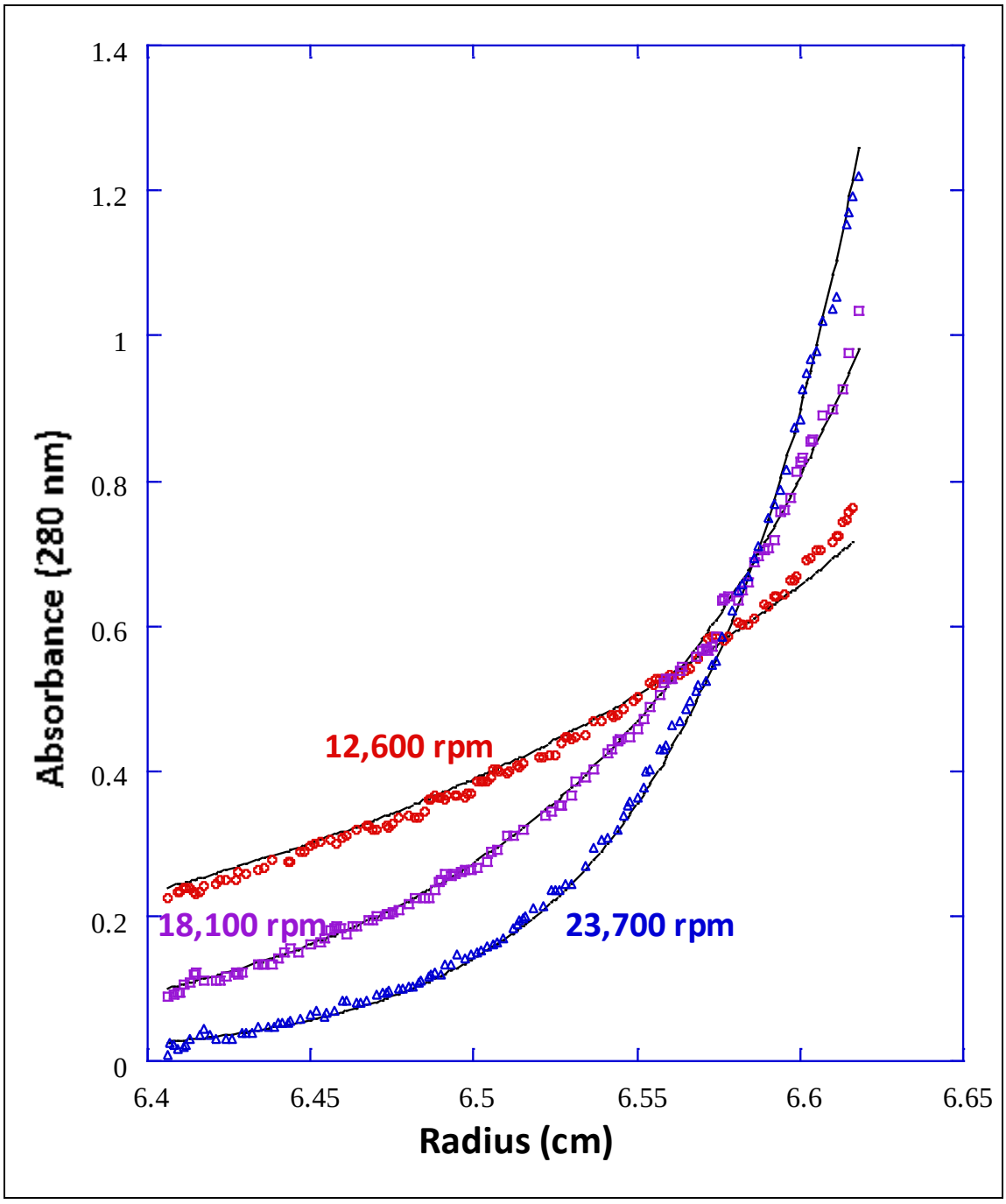


Figure 4.9: Sedimentation equilibrium of Mk0566 using analytical ultracentrifugation. Molecular weight obtained by global fitting was 39,471 g/mol and the enzyme was in monomeric form in absence of DNA. Figure courtesy of Dr. Chad K. Park, Analytical Biophysics Core Facility.

Crystal Structure of Mk0566

Purified Mk0566 was used in an attempt to obtain the crystal structure DNA containing variations of 5′ flap DNA (Figure 4.5). The crystals obtained diffracted to 3Å resolution with in-house x-ray generating machine (Figure 4.10 right). The space group was determined to be primitive orthorhombic ($P2_12_12_1$). Later, a 2.48 Å diffraction data set was collected at SSRL beamline 9-2 (Table 4.2). Crystals were obtained using 15% PEG 4K, 100 mM Tris buffer, pH 8.5, 150 mM NaCl, 10 mM $CaCl_2$ as the precipitant solution and the hanging drop vapor diffusion method. The hanging drops were composed of 8 mg/ml Mk0566 mixed with a 1.5 fold molar excess of DNA. Processing of the diffraction images and subsequent data reduction were performed with MOSFLM (Leslie 1999) and SCALA (Evans 2006), respectively.

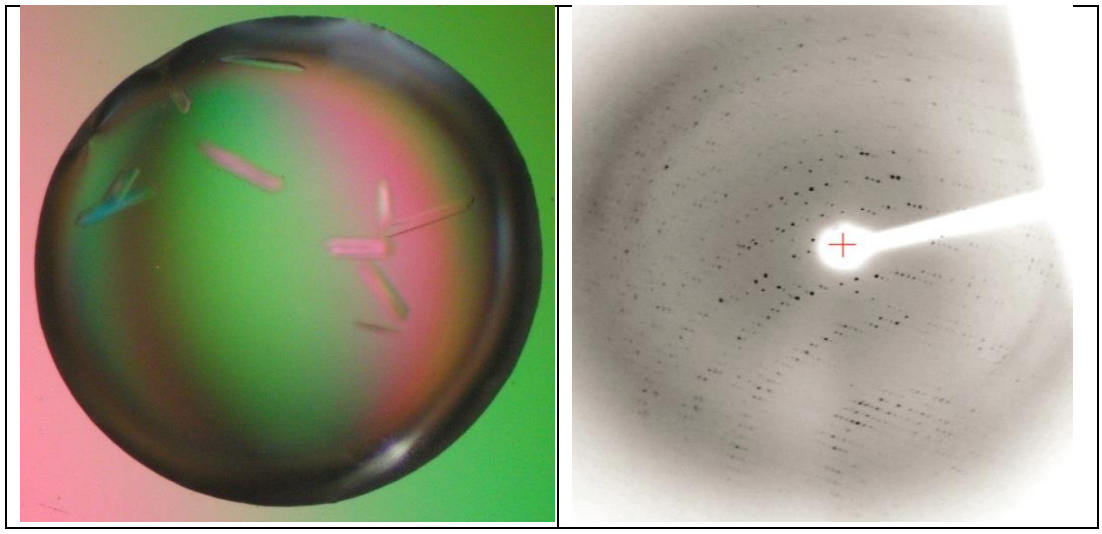


Figure 4.10: (left) Crystals of Mk0566 and (right) the diffraction pattern obtained from the crystals after being shot by the x-ray beam.

The space group of the diffraction data was determined to be $P2_12_12_1$ with cell dimensions of 55.33 Å, 86.19 Å, and 147.62 Å. The final merged, scaled, and reduced data show an I/sigma of 7.7 and R_{merge} of 7.4% (Table 4.2). The structure was solved by molecular replacement using PHASER (Storoni *et al.* 2001) with 1RXW (FEN-1, 52% sequence identity with thumb domain deleted (Chapados *et al.* 2004)) as the search model. Refinement was carried out using PHENIX (Adams *et al.* 2002) and the model was built using Coot (Emsley *et al.* 2004). All structure figures were prepared using PYMOL (DeLano 2002).

The asymmetric unit contains 2 monomers, and iterative refinement, density modification and model building was able to produce a model with all residues with the exception of residues 38-56, 81-128, 186-207 and 255-271. The truncated regions 38-56 and 81-128 contain the gateway and hydrophobic wedge region of FEN-1.

The sequence of the search model, *A. fulgidis* FEN-1, is 52% identical and 69% similar over 354 residues to that of Mk0566, with an E value of $3x10^{-91}$. As expected the structure of Mk0566 is very similar to that of *A. fulgidis* FEN-1. The structure alignment between FEN-1 (3Q8K, Tsutakawa *et al.* 2011) and Mk0566, with distorted arch region truncated from Mk0566, gives an RMSD of 1.090.

The Mk0566 monomer is a mixed α/β mixed structure composed of fifteen alpha helices and eight twisted parallel beta sheets sandwiched in the core (Figure 4.12). An electrostatic surface of Mk0566 (Figure 4.13) shows the likely DNA binding site. In addition, additional positive electrostatic regions may occur in the truncated portion of the structure, also indicative of the DNA binding site.

Table 4.2: Diffraction Data and Structure Refinement Statistics for Mk0566	
Code	Mk32
Beamline	SSRL BL 9-2
Processing program	MOSFLM/SCALA
Space Group	$P2_12_12_1$
Cell	55.33 Å, 86.19 Å, 147.62 Å,
Resolution	2.48 Å
Total Observations	150,824
Unique Observations	25,380
Completeness	98.4 %
I/sigma	7.7
Multiplicity	5.9
Rmerge[1]	7.4 %
Rcryst[2]	21.1 %
Rfree[3]	28.0 %
Overall B factor (Wilson plot)	42.53 $Å^2$
RMSD-bonds	0.02 Å
RMSD-angles	1.78°
Asymmetric unit	2 monomers
Numbers of waters	111

[1] $R_{merge}=\sum_{hkl}(|<I_{hkl}>-I_{hkl}|)/(\sum_{hkl} I_{hkl})$ where $<I_{hkl}>$ is the average intensity over symmetry related and equivalent reflections and I_{hkl} is the observed intensity for reflection hkl.

[2] $R_{cryst}=\sum_{hkl}(||F_{obs}|-|F_{calc}||)/(\sum_{hkl}|F_{obs}|)$ where $|F_{obs}|$ and $|F_{calc}|$ are the observed and calculated structure factor amplitude for reflection hkl. The sum is carried out over the 98% of the observed reflections which are used in refinement.

[3] R_{free} refers to the R factor for the test reflection set (5% of the total observed) which was excluded from refinement.

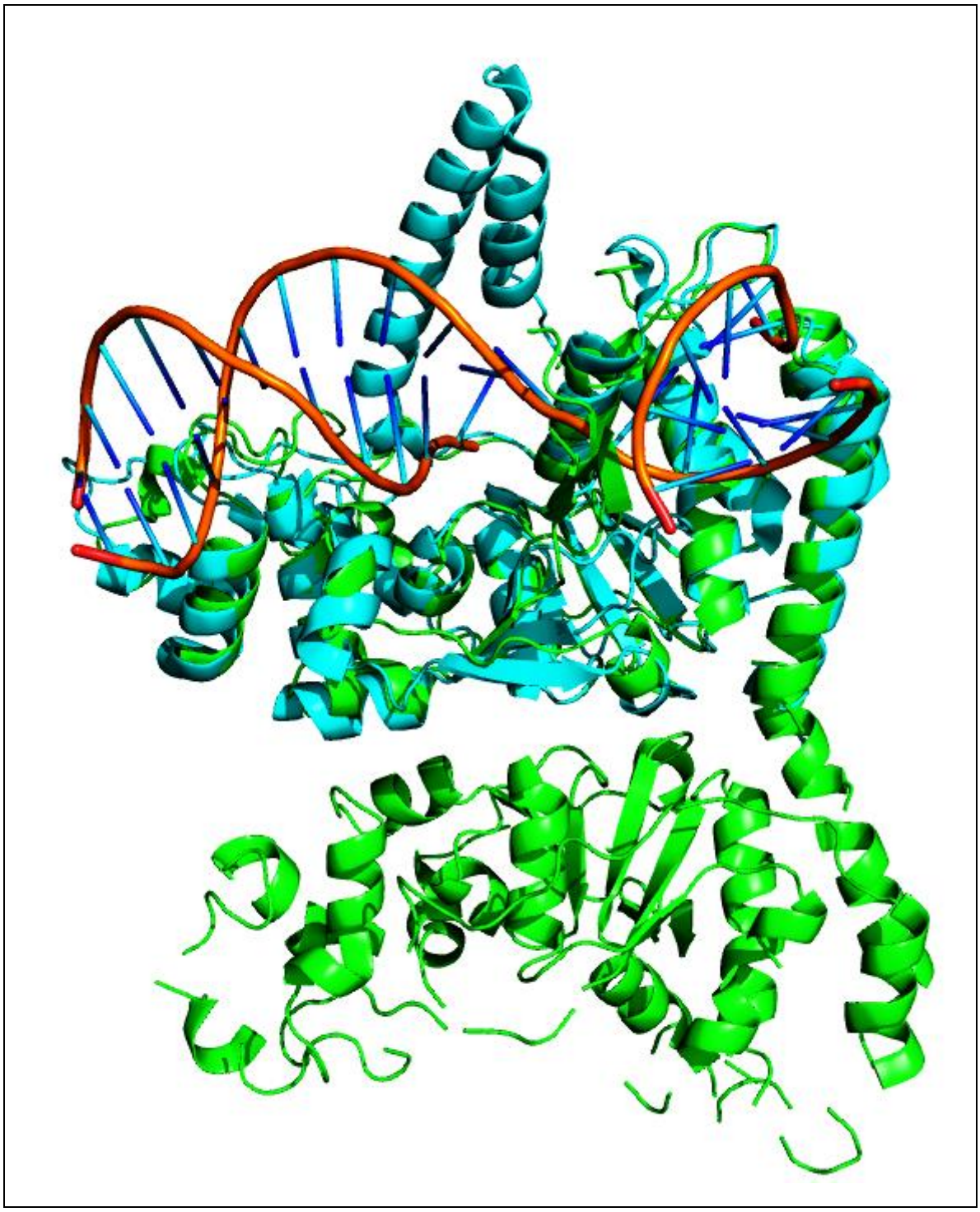


Figure 4.11: Ribbon structure of FEN-1 (3Q8K, cyan) superimposed on Mk0566 (green). RMSD = 1.028. FEN-1 structure shows the location of the arch.

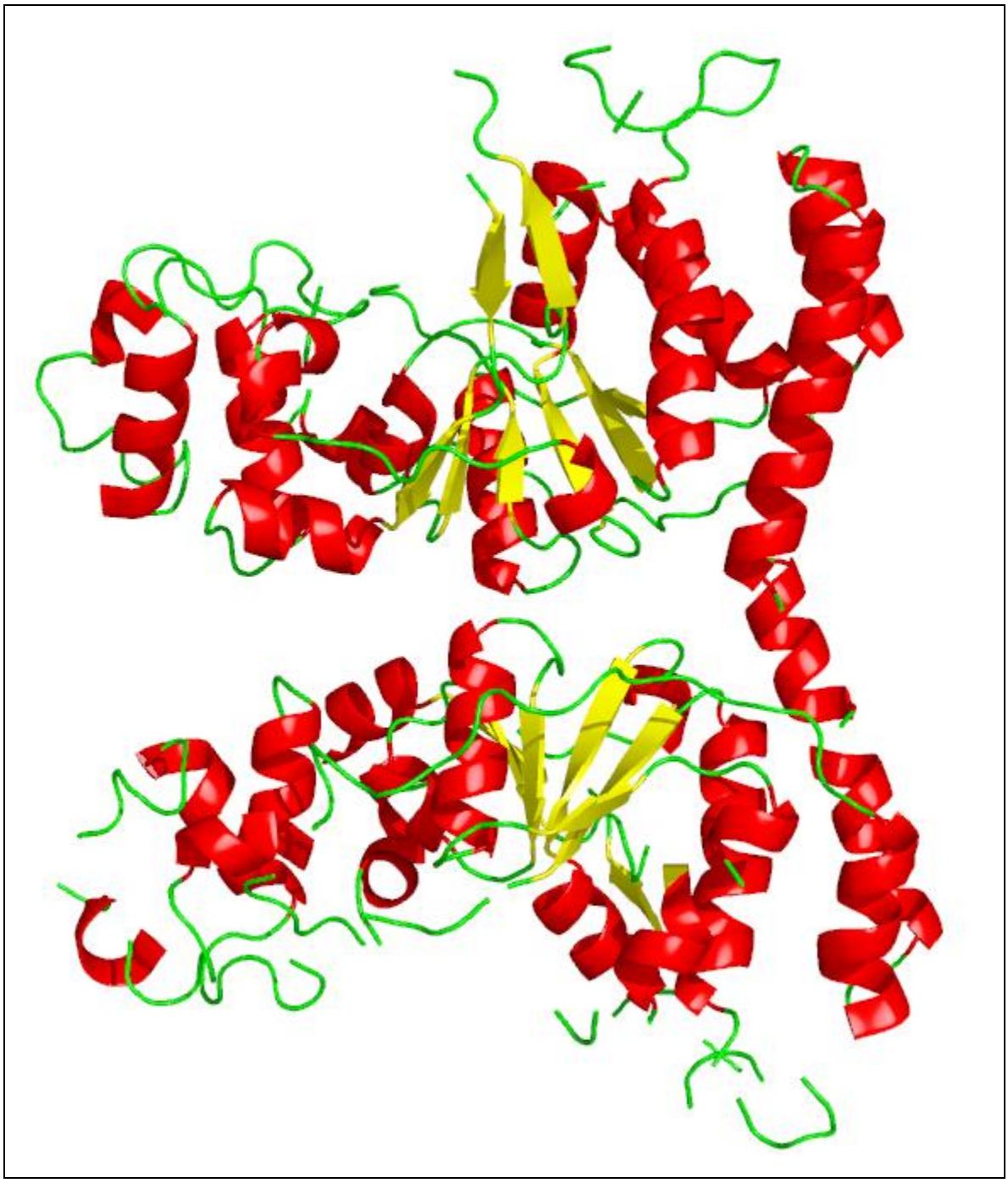


Figure 4.12: Ribbon diagram of Mk0566 dimer showing alpha helices (red), beta sheets (gold) and loops (green) of Mk0566.

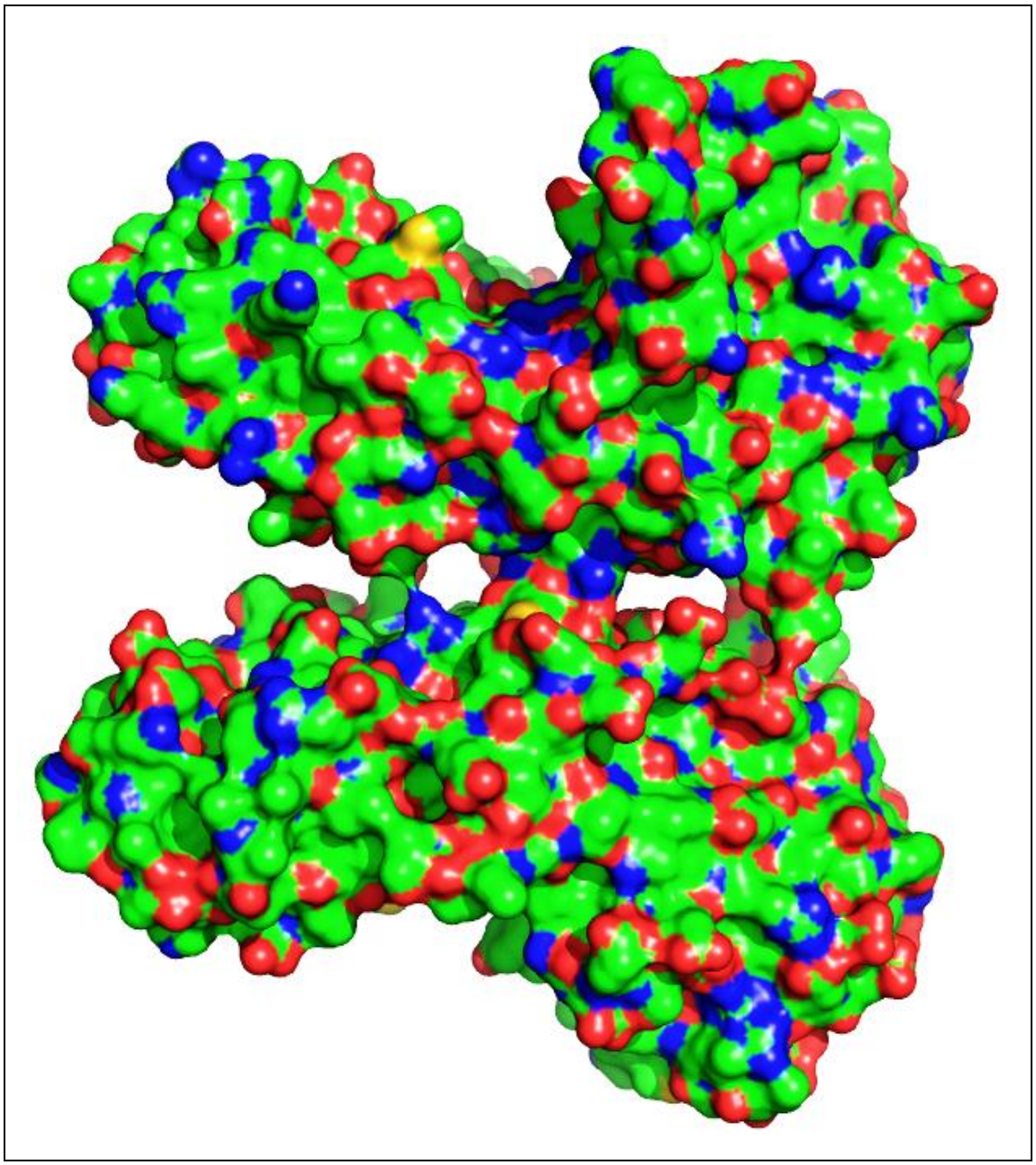


Figure 4.13: Electrostatic surface of Mk0566 and cations showing possible DNA binding site. Coloring done by element with carbon in green, nitrogen in blue, oxygen in red and sulfur in gold.

Discussion:

Initially, Mk0566 was purified using amylose resin in an Econo column (BioRad, Inc.). The MBP tagged Mk0566 bound, and eluted from the amylose resin in this column without problems. However, after cleavage of the MBP tag and attempted repurification using the same column (prewashed and re-equilibration), very little Mk0566 was recovered. Therefore Q-Sepharose and Butyl fast flow columns were used with FPLC (GE, Inc.) to further purify the protein. Later it was deduced that the bottom frit of the Econo column was binding most of the Mk0566, and therefore subsequent amylose purifications utlized batch methods in 50 ml tubes. Combined with the FPLC chromatography, this switch to batch purification allowed a large amount of Mk0566 to be purified.

Multiple sedimentation velocity and equilibrium experiments were required before meaningful data was acquired. Moreover, all the attempts to obtain sedimentation velocity and equilibrium data with enzyme bound to DNA was unsuccessful, showing only the free protein and free DNA. Although extensive DNA cleavage studies had previously been carried out by prior students (see Results), the newly purified protein showed no DNA cleavage activity (data not shown). The origin of these differing results was never determined.

Mk0566 is the only homolog of human FEN-1 in *M. kandleri*, while humans contain four different homologues with different biological functions reflected by their different substrate specificities. Mk0566 was purified and studied in order to address the

possibility that this one enzyme could be performing the many functions performed by the multiple FEN-1 homologues in humans. The substrate specificity was first studied to determine which specificity it possessed, which could then shed light on its possible biological role. DNA cleavage assays with the different flap and overhang substrates (Figure 4.4) showed that Mk0566 only cleaves 5´ flap and blocked flap. This shows a preference for substrates containing duplex DNA both upstream and downstream from 5´ single stranded flap, closer to, but more limited than FEN-1. However, unlike hFEN-1 or hXPG it does not cleave the pseudo-Y or bubble substrates. Its inability to cleave the bubble substrate could be due to the absence of the double stranded region upstream of the 5´ flap, or the lack of the free 5´ end. Although Mk0566 cannot cleave the bubble substrate, it could still function in NER pathway by being aided by other enzymes. Mk1678, a homolog of XPF nuclease, could make the incision which would create a flap structure for Mk0566. Similarly, it is also possible that Mk0566 can cleave the 5´ flap DNA only after new DNA has been polymerized once the 3´ end is created by XPF cleavage. It is known that replication protein A (RPA) aids hFEN-1 in the cleavage of bubble substrates (Zheng *et al.* 2005). Similarly, Mk1441 contains a homolog to one of the three RPA subunits, p70, which might aid Mk0566 in the cleavage process. (Chapados *et al.* 2004, Friedrich-Heineken *et al.* 2004).

If Mk0566 is involved in any repair mechanisms, it must ensure cleavage takes place effectively and specifically at the correct location to prevent damaging DNA cleavage to the genome. It must therefore exhibit some sort of specificity to recognize a particular type of structure specific DNA. The current crystal structure of Mk0566 is

missing any bound DNA molecule and also much of the arch region. Given the strong sequence similarity, and similar cleavage specificity, it is very likely that Mk0566 will have similar DNA contacts and helical arch conformation as found in the FEN-1/DNA complex structure (Tsutakawa *et al.* 2011). Pertaining to the fact that both FEN-1 and Mk0566 belong to the same superfamily, and neither cleaves the bubble substrate, we can expect the bound DNA to be severely bent, and the 5´ flap would pass through the arch in Mk0566 as well. Also the arch region would be more ordered upon binding of the DNA. It could also achieve specificity by a unique double-base flipping of the unpaired bases on the 5´ flap DNA (Tsutakawa *et al.* 2011). Just like FEN-1, EXO-1, XPG and GEN-1, Mk0566 could also be a junction specific enzyme that would require three separate binding sites for DNA to carry out the nucleotide excision. Similarly, the presence of conserved α2 helix might confer Mk0566 the flap incision activity, whereas the absence of spacer region (~ 650 residues) similar to XPG might explain why Mk0566 cannot cleave the bubble substrate. Thus, comparing structure-function similarities and differences among the FEN-1 superfamily provides insight into the potential role of Mk0566 in DNA metabolism.

CHAPTER 5

SIGNIFICANCE AND FUTURE DIRECTIONS

Streptomycetes are one of the most abundant and ubiquitous soil-dwelling, filamentous Gram-positive bacteria that have a variety of metabolic processes and biotransformations. They supply over two-thirds of all the naturally derived antibiotics used today e.g. streptomycin, tetracycline and chloramphenicol (Hodgson 2000, Kong *et al.* 2000, Bentley *et al.* 2002, Ohnishi *et al.* 2008). Unlike most bacteria, streptomycetes have complex multicellular development (Bentley *et al.* 2002). *Streptomyces griseus* has a linear chromosome consisting of 8,545,929 base pairs; one of the largest bacterial genome discovered so far and almost twice the size of most bacterial genomes. It also has a high G+C content of about 72.2%. The genome contains a large proportion of regulatory genes that are mainly involved in responding to external stimuli and stresses. It has about 7,138 genes that could, in theory, be translated into proteins (Ohnishi *et al.* 2008, Bentley *et al.* 2002). The large genome might be essential for survival in the highly competitive soil environment allowing for a more complex life cycle, to exploit a greater variety of nutrient sources, to adapt to multitude of stress conditions and to transport variety of substrates such as sugars, amino acids, peptides, metals and ions in the cellular level (Bentley *et al.* 2002). Moreover, streptomycetes are under constant evolutionary pressure in order to combat phage invasion. SgrAI, a restriction endonuclease from *Streptomyces griseus*, is thought to be activated upon phage invasion, undergoing DNA recognition specificity expansion and forming large polymers to sequester its own DNA

from the activated SgrAI. Enzyme regulation involving polymer formation is very rare and is a new paradigm in enzyme and restriction endonuclease regulation. Although there are few enzymes that form run-on oligomers that alter the enzyme activity, they do not show alteration of substrate specificity in the run-on oligomer or reversible domain swapping as shown by SgrAI (Korennykh *et al.* 2009, Kim *et al.* 2010, Ingerson-Mahar *et al.* 2010, Li *et al.* 2012).

SgrAI has been shown to be an unusual and interesting enzyme. It cleaves plasmids having two recognition sites faster than the ones with only one recognition sites (Bilcock *et al.* 1999, Daniels *et al.* 2002). In addition to its primary cleavage activity it also cleaves secondary sequences CR|CCGGGG or CR|CCGGY(A/C/T) in the presence of the primary sequence (Bitinaite *et al.* 2006). The current working model of SgrAI DNA cleavage reaction mechanism is shown in Figure 5.1. Primary site DNA (1°**)** binds to inactive or low activity conformation of SgrAI dimer (grey boxes, Figure 5.1) and forms a primary site DNA bound SgrAI dimer (DBD). The DBD oligomerizes quickly and forms a run-on oligomer. The run-on oligomer stabilizes the activated conformation of SgrAI and facilitates attachment of additional DBDs. The activated form of run-on oligomer (red boxes, Figure 5.1) then cleaves DNA at an accelerated rate. Although, secondary site DNA by itself does not allow run-on oligomer formation, the secondary site DNA cleavage activation occurs in a similar way in the presence of primary site DNA. The secondary site DBD oligomerize with the primary site DBD to form a run-on oligomer and get activated resulting in an accelerated cleavage of DNA. The activated

conformation favors oligomerization and oligomerization in turn favors the activated conformation.

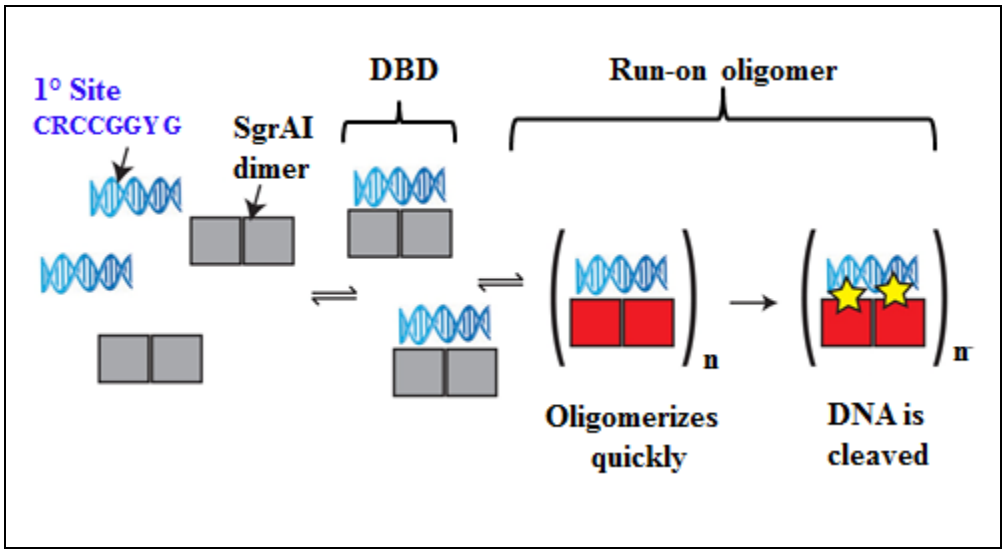


Figure 5.1: Model of SgrAI DNA cleavage reaction mechanism. Grey and red boxes indicate low and high activity conformations of SgrAI, respectively. Stars indicate DNA cleavage. DBD = DNA bound SgrAI dimers. Figure courtesy of Prof. Nancy C. Horton.

The run-on oligomer activates cleavage as well as alters the sequence specificity of SgrAI (Park *et al.* 2010a, Park *et al.* 2010b, and Lyumkis *et al.* 2013). The run-on oligomers are different than the single stranded DNA binding proteins (Kozlov *et al.* 2012, Liu *et al.* 2006) since they bind to specific recognition sequences and the bound DNA loops in and out of the oligomer rather than binding along the oligomer (Lyumkis *et al.* 2013). SgrAI is also unusual in a sense that the uncleaved reactant as well as the cleaved product can stimulate the formation of run-on oligomer (*personal communication with Michael Piperakis and Richard Cosstick*). The crystal structure of SgrAI also shows involvement of reversible domain swapping in formation of oligomer (Park *et al.* 2010a). All these unusual phenomena: activation and specificity modulation by run-on

oligomerization, stimulation by reaction products and reversible domain swapping has not been seen in any other enzyme, and is an unusual mechanism of enzyme regulation. Other enzymes show some of these phenomena but not all of them. Nucleotide triphosphatases (NTPases) such as actin (Wegner 1976, Bindschadler *et al.* 2004), tubulin (Kueh *et al.* 2009), DnaseA (Duderstadt *et al.* 2011), and RecA (Chen *et.al*. 2008) form oligomers but they do not show domain swapping or product based stimulation. They do not change specificity for their NTP either. Similarly, some amyloid proteins form domain swapped run-on oligomeric structures but are often nonfunctional (Staniforth *et al.* 2001, Bennett *et al.* 2006, Stroud *et al.* 2012). Acetyl-CoA carboxylase also forms a run-on oligomer but it does not possess other unusual phenomena such as product stabilization of the activated form, activation, specificity expansion and domain swapping as SgrAI (Brownsey *et al.* 2006). Thus, activation of SgrAI through run-on oligomerization is a new paradigm in enzyme activity. Study of this enzyme holds great potential in understanding run-on oligomerization and its significance in understanding other systems, such as deposition diseases. This type of mechanism may be more common in metabolic enzymes that form filaments or large self-assemblies *in vivo* (Narayanswamy *et al.* 2009, Noree *et al.* 2010). Lessons learned from the SgrAI system could be used in biomedical applications where DNA cleavage could be targeted to multiple recognition sites, or DNA can be silenced by sequestering the run-on oligomer. Similarly, the reversible domain swapping by SgrAI can be used in the biotechnology field in nanofabrication to design reversible bridge, gate or translocation device (Nagarka *et al.* 2010).

SgrAI has shown several unusual mechanistic characteristics and holds great promise for learning more from its unusual properties. The flanking DNA dependent activation, asymmetric cleavage activity of SgrAI on different types of secondary sites, the domain swapped structure, and the cryo-EM structure showing the run-on oligomerization have been novel findings which have kept us captivated and craving for more. Moving forward, the information obtained from 8.6 Å cryo-EM structure will be tested using site directed mutagenesis. The cryo-EM structure showed that the flanking DNA made seemingly important contacts to neighboring SgrAI, especially with loops 56-60 and 122-140 (Lyumkis *et al.* 2013). Various mutations in the loops: S56E, S56Q, A57E, or A57Q in Loop 56-60 and R127A, R131A, or R134A in Loop 122-140 have been designed. Single turn-over rate constants and ability of run-on oligomer formation by the mutants S56E, S56Q, A57E, A57Q, R127A, R131A, and R134A will be tested and compared with the wild type SgrAI in the future. The results obtained from these experiments will allow testing of the information obtained from the cryo-EM structure. There are also other residues (e.g. F3, T4, I7, E8, R11, R24, N27, I50, I59, M62, and R84) at the interface of the amino termini that can be mutated to either alanine or arginine or glutamate (large charged groups) and tested for its possible role in oligomer stabilization. Since these mutations are made in the loops, there is a possibility that the loops might become more ordered and form different secondary structures. Thus, circular dichroism experiments can be done at far-UV range to see if any secondary structures have been altered compared to WT SgrAI.

Similarly, the single-turnover cleavage assays carried out using K96A SgrAI mutants have shown some interesting results, especially with secondary site DNA. Removal of lysine side chain seems to make the mutant enzyme more specific. K96A SgrAI does not show cleavage of secondary site DNA. Thus, obtaining a crystal structure of K96A SgrAI bound to secondary site DNA and other biochemical assays using secondary and non-cognate DNA will help understand the gain in specificity by the mutant.

The crystal structure of SgrAI WT bound to 22-1HT DNA (included in this thesis) partially supports the cryo-EM structure. It shows contacts made by the flanking DNA to the SgrAI and interactions between N-terminal domains. However, it is still in the low activity conformation as seen in other SgrAI structures obtained so far (Dunten *et al.* 2008). Crystallization trials with activating, longer flanking PC DNA variants have been initiated to obtain a high resolution activated, oligomeric form of SgrAI. Similarly, in order to confirm the existence of domain swapped structure (Park *et al.* 2010a) crosslinking experiments can be performed. The domain swapped structure could be captured using cysteine crosslinking to map the domain swapped regions in the run-on activated oligomer. Either naturally occurring cysteine (Cys 19) or strategically positioned cysteine point mutations will be made for the experiment. Two sets of residues could be mutated: the first set will consist of residues that are in the N-terminal domain swapped region (residue 1-26) and the second set will consist of residues within the proximity of residues in first set of another SgrAI chain based on the domain swapped

structure (Park *et al.* 2010a). The evidence obtained from the experiment will help confirm the unusual enzyme mechanism implemented by SgrAI.

Other biochemical experiments can also be conducted to figure out the mechanism adapted by SgrAI. Forster resonance energy transfer (FRET) experiments using doubly labeled primary site DNA with fluorescein on the 5´ end of one strand and rhodamine-X on the 5´ end of the other strand can be used in DNA cleavage experiments. Changes in FRET signal due to emission from the donor and the acceptor fluorophore will be used to measure the DNA binding, bending and product release rates. Similarly, FRET experiments can be designed using rhodamine labeled 18-1 and fluorescein labeled PC DNA to measure the rates of run-on oligomer formation and dissociation and product release. These experiments will allow answering of several fundamental questions; such as, does the association of the run-on oligomer limit the DNA cleavage rate? Does the run-on oligomer slow product release during enzyme turnover? And can the cleaved DNA be released from the run-on oligomer? Or is dissociation of the run-on oligomer to DBD required before product can be released? Answering these questions will allow better understanding of enzyme mechanism adapted by this unusual and fascinating enzyme, SgrAI.

APPENDIX

APPENDIX A – LIST OF PRIMERS USED FOR CLONING

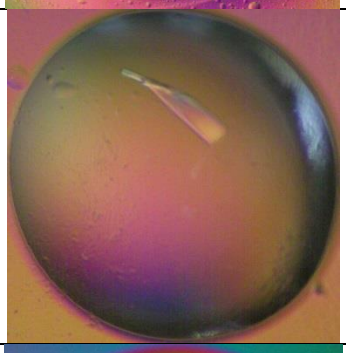
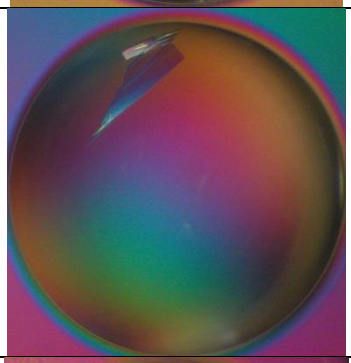
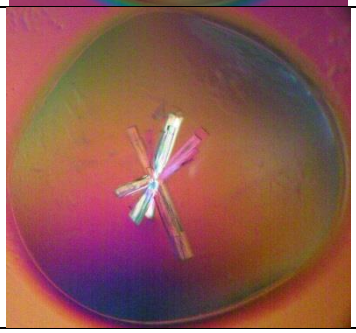
Construct	Primer 1 (Forward)	Primer 2 (Reverse)	Restriction Sites
K96A [x]	AGGCGGCAGXTGCCGGCGATATTTTC	ACTGCCGCCXTTGCATTTGAATCTACG	
SgrAI-His [Δ]	TAGCATGACTGGTGGACA	TGCATCTCGAGTGCTCATGACTCATCAGTGGTGGTGGTGGTGGTGGCCGTTCACCAGCTCC	BamHI, XhoI
R127A [Δ]	GTGGGTGAAGGGTGGgccTCTCAGCCGCGGTAC	GTACCGCGGCTGAGAggcCCACCCTTCACCCAC	BamHI, XhoI
R131A [Δ]	TGGCGCTCTCAGCCGgcgTACAGTAGGCCGACG	CGTCGGCCTACTGTAcgcCGGCTGAGAGCGCCA	BamHI, XhoI
R134A [Δ]	CAGCCGCGGTACAGTgcgCCGACGCTGAGTCCG	CGGACTCAGCGTCGGcgcACTGTACCGCGGCTG	BamHI, XhoI

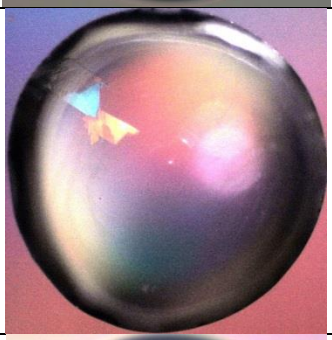
[x] Cloned by Alka Agrawal

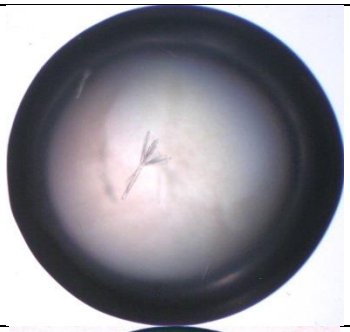
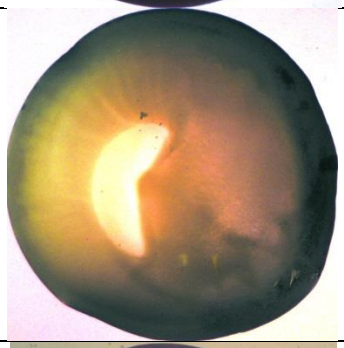
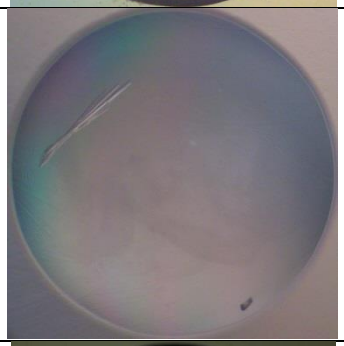
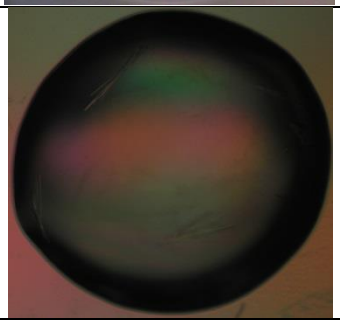
[*] Cloned by Jurate Bitinaite

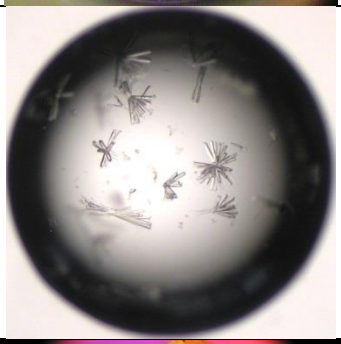
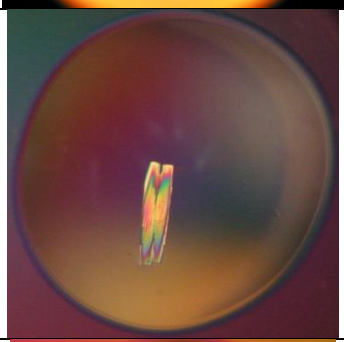
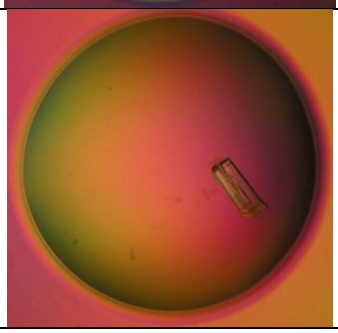
Δ Cloned by Jonathan Sanchez

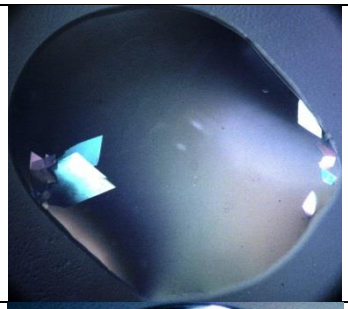
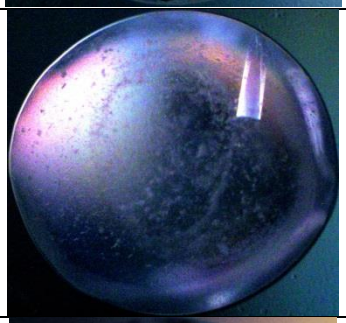
APPENDIX B – CRYSTALLIZATION CONDITIONS

Tray ID	Enzyme	DNA	Crystals	Precipitant
S483_B4	WT 3.0 mg/ml	18-1AT		150 mM NaCl, 100 mM Imidazole pH 6.5, 20% PEG4K
S483_D4	WT 3.0 mg/ml	18-1AT		150 mM NaCl, 100 mM Imidazole pH 6.5, 10% PEG4K
S440_B4	WT 3.5 mg/ml	18-1GC		5 mM $CaCl_2$, 150 mM NaCl, 100 mM Imidazole pH 6.5, 15% PEG4K
S476_D4	WT 12 mg/ml	18-1GC		10 mM $CaCl_2$, 150 mM NaCl, 100 mM Imidazole pH 6.5, 10% PEG4K

S550_C6	WT 5 mg/ml	18-1 Mixed		10 mM $CaCl_2$, 150 mM NaCl, 100 mM Tris pH 8.5, 10% PEG4K
S551_B4	WT 4 mg/ml	18-1 Mixed		10 mM $CaCl_2$, 150 mM NaCl, 100 mM Imidazole pH 6.5, 8% PEG4K
S501_B4	WT 7 mg/ml	19-1 GC_G		10 mM $CaCl_2$, 150 mM NaCl, 100 mM Imidazole pH 6.5, 20% PEG4K
S584_D6	WT 2.4 mg/ml	PC		10 mM $Mg(OAc)_2$, 150 mM NaCl, 100 mM Tris pH 8.5, 5% PEG4K

S601_D4	WT 5 mg/ml	PC-3		100 mM Imidazole pH 6.5, 40% 4M $(NH_4)_2SO_4$
S630_D2	WT 7 mg/ml	PC-4		10 mM $Ca(OAc)_2$, 10 mM $NaNO_3$, 150 mM NaCl, 100 mM NaOAc pH 5.5, 1% PEG6K
S627_D2	WT 6 mg/ml	PC-5		10 mM $Ca(OAc)_2$, 10 mM $NaNO_3$, 150 mM NaCl, 100 mM NaOAc pH 5.5, 5% PEG6K
S556_D5	WT 5 mg/ml	PC-6 HT		10 mM $Ca(OAc)_2$, 150 mM NaCl, 100 mM HEPES pH 7.5, 5% PEG4K
S564_C4	WT 3 mg/ml	PC-6 HT		10 mM $Ca(OAc)_2$, 10 mM $NaNO_3$, 150 mM NaCl, 100 mM Imidazole pH 6.5, 10% PEG4K

S574_B6	WT 5.5 mg/ml	16-2B		10 mM $Ca(OAc)_2$, 150 mM NaCl, 100 mM Tris pH 8.5, 15% PEG4K
S599_D5	WT 4.5 mg/ml	18-2B		100 mM HEPES pH 7.5, 40% 4M $(NH_4)_2SO_4$
S604_C2	WT 5.2 mg/ml	18-2B		100 mM Tris pH 9.5, 40% 4M $(NH_4)_2SO_4$
S534_A4	E301W 5 mg/ml	18-1GC		10 mM $CaCl_2$, 150 mM NaCl, 100 mM Imidazole pH 6.5, 8% PEG4K
S534_A5	E301W 5 mg/ml	18-1GC		10 mM $CaCl_2$, 150 mM NaCl, 100 mM Imidazole pH 6.5, 6% PEG4K

S489_A3	K96A 5 mg/ml	18-1GC		10 mM $CaCl_2$, 150 mM NaCl, 100 mM Imidazole pH 6.0, 15% PEG4K
S489_A5	K96A 5 mg/ml	18-1GC		10 mM $CaCl_2$, 150 mM NaCl, 100 mM buffer pH 7.0, 15% PEG4K
S506_B4	K96A 7 mg/ml	18-1GC		10 mM $CaCl_2$, 150 mM NaCl, 100 mM Imidazole pH 6.5, 15% PEG4K
S506_C4	K96A 7 mg/ml	18-1GC		10 mM $CaCl_2$, 150 mM NaCl, 100 mM Imidazole pH 6.5, 10% PEG4K

APPENDIX C – SOLVED X-RAY CRYSTAL STRUCTURES

Crystal ID	Enzyme	DNA	Resolution (Å)	Comments
Mk32	Mk0566	L8R9	2.48	No DNA
S451	SgrAI R29A	18-2, PC	1.65	
S489	SgrAI K96A	18-1GC	1.89	
S506	SgrAI K96A	18-1GC	2.19	
S501	SgrAI WT	18-1GC	1.80	
S706	SgrAI WT	18-1AT	2.19	
S708	SgrAI WT	18-1AT	3.00	
S634	SgrAI WT	22-1HT	3.50	
S701	SgrAI WT	PC	2.80	No DNA
S669	SgrAI WT	PC-4	3.50	No DNA
S671	SgrAI WT	PC-5	4.00	No DNA
S664	SgrAI WT	16-2B	4.96	

APPENDIX D – LIST OF PUBLICATIONS

SN	Publications
1.	Chad K. Park, Amanda P. Stiteler, Santosh Shah, M. Imran Ghare, Jurate Bitinaite, and Nancy C. Horton (2010) Activation of DNA Cleavage by Oligomerization of DNA-Bound SgrAI; *Biochemistry* 49 (41) 8818–8830.
2.	Xin Ma, Santosh Shah, Mowei Zhou, Chad K. Park, Vicki H. Wysocki, and Nancy C. Horton (2013) Structural Analysis of Activated SgrAI-DNA Oligomers Using Ion Mobility Mass Spectrometry; *Biochemistry* 52, (25) 4373-4381.
3.	Dmitry Lyumkis, Heather Talley, Andrew Stewart, Santosh Shah, Chad K. Park, Florence Tama, Clinton S. Potter, Bridget Carragher, Nancy C. Horton (2013) Allosteric Regulation of DNA Cleavage and Sequence Specificity through Run-On Oligomerization; *Structure* 21, 10, 1848-1858.

APPENDIX E – SUMMARY OF CONTRIBUTIONS

Table/Figure	Contributed By
Figure 1.9	Figure constructed by Dr. Nancy Horton.
Table 2.1	Single turn-over (STO) cleavage rate constants of 1 µM WT SgrAI and 40-1AT, 40-2A, 40-2A-Top, 40-2A-Bottom, 40-2B, 40-2B-Top, 40-2B-Bottom, 18-NCTA, 18-NCCG and 40-NCTA were measured by Santosh Shah. STO cleavage rate constants of 1 µM WT SgrAI and 18-1AT in absence and presence of 1 µM PC was taken from Park *et. al.* 2010b paper.
Figure 2.3	Data of 18-1 and 18-2A plotted in bar graph was taken from Park *et. al.* 2010b paper.
Figure 2.4	Figure A adapted from Dunten *et al*. 2008; Figure B adapted from Park *et al*. 2010b; and Figure C adapted from Deibert *et al*. 2000. All the figures were made by Santosh Shah.
Figure 2.5	Figure adapted from Park *et al*. 2010b. Figure was made by Santosh Shah.
Table 2.2	Dissociation constants of WT and R29A SgrAI measured by Dr. Nancy Horton.
Table 2.3	STO rate constants of WT SgrAI are from Park *et al*. 2010b paper, and of R29A SgrAI were measured by Amanda Stiteler.
Figure 2.7	Bar graph of data from Table 2.3 drawn by Santosh Shah
Table 2.4	STO rate constants of WT SgrAI and R29A SgrAI were measured by Amanda Stiteler.
Figure 2.8	Bar graph of data from Table 2.4 drawn by Santosh Shah.
Table 2.5	Crystal obtained by Santosh Shah. Data collection and refinement of R29A SgrAI crystal structure carried out by Dr. Nancy Horton.
Figures 2.9, 2.10, & 2.11	Figures constructed by Santosh Shah.
Figure 2.12	Purity gel run by Santosh Shah.
Table 2.6	STO rate constants of E301W SgrAI were measured by Santosh Shah.
Figure 2.13	Bar graph of data from Tables 2.1 and 2.6 drawn by Santosh Shah.
Table 2.7	Data collection and refinement of WT SgrAI bound to 22-1HT crystal structure carried out by Dr. Nancy Horton.
Figures 2.14, 2.16, & 2.19	Figures constructed by Dr. Nancy Horton.
Figures 2.15, 2.17, 2.18	Figures constructed by Santosh Shah.
Figures 2.21, 2.22, & 2.23	Plasmid assays carried out by Santosh Shah and figures constructed by Dr. Nancy Horton.
Figure 3.1	Figure adapted from Dunten *et al*. 2008 and figure constructed by Santosh Shah.

Table 3.2	Binding affinities of K96A SgrAI measured by Santosh Shah, and of WT SgrAI adapted from Park *et al*. 2010b paper.
Table 3.3	STO cleavage rate constants of K96A SgrAI measured by Santosh Shah.
Figure 3.2	Crystal of K96A SgrAI obtained by Santosh Shah. Data collection by Santosh Shah.
Table 3.4	Data collection and refinement of all three structures by Santosh Shah with the help of Dr. Sue Roberts and Dr. Andrzej Weichsel.
Figures 3.3, 3.4, & 3.5	Figures constructed by Santosh Shah.
Figure 4.1	Sequence alignment carried out by Dr. Nancy Horton, and the figure constructed by Santosh Shah.
Figure 4.2	Figure adapted from Chapados *et al*. 2004, and figure constructed by Santosh Shah.
Figures 4.3 & 4.4	Figure 4.3 (Left) constructed by Santosh Shah. Figure 4.3 (Right) and 4.4 adapted from the senior thesis of Amanda Stiteler.
Figures 4.6 & 4.7	Gels run by Santosh Shah. Lot 2 Mk0566 used as positive control marker was purified by Amanda Stiteler and Dr. Andrea Babic.
Figures 4.8 & 4.9	AUC experiments performed by and figures constructed by Dr. Chad Park.
Figure 4.10	Crystals obtained by Santosh Shah. Data collection carried out by Santosh Shah and Dr. Nancy Horton.
Table 4.2	Mk0566 crystal structure data collection by Dr. Nancy Horton and Santosh Shah, refinement by Dr. Nancy Horton, Dr. Pete Dunten and Santosh Shah.
Figures 4.11, 4.12, & 4.13	Figures constructed by Santosh Shah.
Figure 5.1	Figure constructed by Dr. Nancy Horton.

APPENDIX F – COPYRIGHT PERMISSION FOR CHAPTER 2 - FIGURE 2.1

PQ www.umi.com/assets/dov × Allosteric Regulation of DI × Rig

https://s100.copyright.com/AppDispatchS

Copyright Clearance Center **RightsLink®**

Home | Account Info | Help

Structure

Title: Allosteric Regulation of DNA Cleavage and Sequence-Specificity through Run-On Oligomerization

Author: Dmitry Lyumkis,Heather Talley,Andrew Stewart,Santosh Shah,Chad K. Park,Florence Tama,Clinton S. Potter,Bridget Carragher,Nancy C. Horton

Publication: Structure

Publisher: Elsevier

Date: 8 October 2013

Copyright © 2013, Elsevier

Logged in as: Santosh Shah

LOGOUT

Order Completed

Thank you very much for your order.

This is a License Agreement between Santosh Shah ("You") and Elsevier ("Elsevier"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

Get the printable license.

License Number	3276190423694
License date	Nov 25, 2013
Licensed content publisher	Elsevier
Licensed content publication	Structure
Licensed content title	Allosteric Regulation of DNA Cleavage and Sequence-Specificity through Run-On Oligomerization
Licensed content author	Dmitry Lyumkis,Heather Talley,Andrew Stewart,Santosh Shah,Chad K. Park,Florence Tama,Clinton S. Potter,Bridget Carragher,Nancy C. Horton
Licensed content date	8 October 2013
Licensed content volume number	21
Licensed content issue number	10
Number of pages	11
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Title of your thesis/dissertation	Structural and Functional Studies of DNA Nucleases: SgrAI and Mk0566
Expected completion date	Dec 2013
Estimated size (number of pages)	200
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	0.00 USD

ORDER MORE... | CLOSE WINDOW

Copyright © 2013 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement.
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

REFERENCES

Abramoff, M. D. (2009) Image Processing with ImageJ. *Biophotonics International* 11, (7) 36-42.

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W. *et al*. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta. Crysta. D Biol. Crysta.* 58 (Pt 11) 1948.

Aggarwal, A. K. (1990) Crystallization of DNA binding proteins with oligodeoxynucleotides. *Methods: A Companion to Methods in Enzymology* 1, 83-90.

Aggarwal, A. K. (1995) Structure and function of restriction endonucleases. *Curr. Opin. Struct. Biol.* 5, 11-19.

Aiyar, A., Xiang, Y., and Leis, J. (1996). Site-directed mutagenesis using overlap extension PCR. *Methods Mol. Biol*. 57, 177-191.

Anderson, W. F., Ohlendorf, D. H., Takeda, Y. *et al.* (1981) Structure of the cro repressor from bacteriophage lambda and its interaction with DNA. *Nature* 290, 754-758.

Anderson, J. E. (1993) Restriction endonucleases and modification methylases. *Curr. Opin. Struct. Biol.* 3, 24-30.

Arber,W., Dussoix,D. (1962) Host specificity of DNA produced by *Escherichia coli*. I. Host controlled modification of bacteriophage lambda. *J. Mol. Biol*. 5, 18–36.

Athanasiadis, A., Vlassi, M., Kotsifaki, D., Tucker, P. A., Wilson, K. S., Kokkinidis, M. (1994) Crystal structure of PvuII endonuclease reveals extensive structural homologies to EcoRV. *Nature Struct. Biol*. 1, 469–475.

Babic, A. C., Little, E. J., Manohar, V. M., Bitinaite, J., Horton, N. C. (2008) DNA distortion and specificity in a sequence-specific endonuclease. *J. Mol. Biol.* 383, 186-204.

Bareket-Samish, A., Cohen, I., Haran, T. E. (1998) Direct versusindirect readout in the interaction of the trp repressor with non-canonical binding sites. *J. Mol. Biol.* 277, 1071-1080.

Beese, L. S., Steitz, T. A. (1991) Structural basis for the 3´-5´ exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J*. 10, 25-33.

Bennett, M. J., Sawaya, M. R., Eisenberg, D. (2006) Deposition diseases and 3D domain swapping. *Structure* 14, 811-824.

Bentley, S. D. *et al.* (2002) Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* 417, 141-147.

Bilcock, D. T., Daniels, L. E., Bath, A. J., Halford, S. E. (1999) Restrictions of type II restriction endonuclease with 8-base pair recognition sites. *J. Biol. Chem.*, 274 (51) 36379-36386.

Bindschadler, M., Osborn, E. A., Dewey, C. F., Jr. McGrath, J. L. (2004) A mechanistic model of the actin cycle. *Biophys. J*. 86, 2720-2739.

Bitinaite, J., Schildkraut, I. (2002) Self-generated DNA termini relax the specificity of SgrAI restriction endonuclease. *Proc. Natl. Acad. Sci.* USA 99 (3) 1164-1169.

Blakaj, D. M., Kattamuri, C., Khrapunov, S., Hegde, R. S., Brenowitz, M. (2006) Indirect readout of DNA sequence by papillomavirus E2 proteins depends upon net cation uptake. *J. Mol. Biol*. 358, 224–240.

Bourniquel, A. A., Bickle, T. A. (2002) Complex restriction enzymes: NTP-driven molecular motors. *Biochimie* 84, 1047-1059.

Boyce, R. P., Howard-Flanders, P. (1964) Release of ultraviolet light-induced thymine dimers from DNA in *E. coli* K-12. *Proc. Natl. Acad. Sci*. USA 51, 293–300.

Bozic, D., Grazulis, S., Siksnys, V., Huber, R. (1996) Crystal structure of *Citrobacter freundii* restriction endonuclease Cfr10I at 2.11 Å resolution. *J. Mol. Biol*. 255 (1) 176-186.

Brownsey, R. W., Boone, A. N., Elliott, J. E., Kulpa, J. E., Lee, W. M. (2006) Regulation of acetyl-CoA carboxylase, *Biochem Soc Trans 34*, 223-227

Buschta-Hedayat, N., Buterin, T., Hess, M. T., Missura, M., Naegeli, H. (1999) Recognition of nonhybridizing base pairs during nucleotide excision repair of DNA. *Proc. Natl. Acad. Sci.* 96, 6090-6095.

Carson, M., Johnson, D. H., McDonald, H., Brouillette, C., DeLucas, L. J. (2007) His-tag impact on structure. *Acta Cryst*. D63, 295-301.

Chant, A., Kraemer-Pecore, C. M., Watkin, R. and Kneale, G. G. (2005) Attachment of a histidines tag to the minimal zinc finger protein of the Aspergillus nidulans gene regulatory protein AreA causes a conformational change at the DNA-binding site. *Protein Expr. Purif*. 41, 61-67.

Chapados, B. R., Hosfield, D. J., Han, S., Qiu, J., Yelent, B., Shen, B., Tainer, J. A. (2004) Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. *Cell*, 116, 39–50.

Chen, S., Gunasekera, A., Zhang, X., Kunkel, T. A., Ebright, R. H., Berman, H. M. (2001) Indirect readout of DNA sequence at the primary-kink site in the CAP–DNA complex: alteration of DNA binding specificity through alteration of DNA kinking. *J. Mol. Biol.* 314, 75–82.

Chen, Z., Yang, H., Pavletich, N. P. (2008) Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* 453, 489-484.

Cleaver, J. E., Kraemer, K. H. (1989) The Metabolic Basis of Inherited Disease. Ed. Scriver, C. R., Beaudet, A. L., Sly, W. S., Valle, D. New York McGraw-Hill 2, 2949-2971.

Cloud, K. G., Shen, B., Strniste, G. F., Park, M. S. (1995) XPG protein has a structure-specific endonuclease activity. *Mutat. Res.*, 347, 55–60.

Collaborative Computational Project, Number 4. 1994. "The CCP4 Suite: Programs for Protein Crystallography". *Acta Cryst.* D50, 760-763.

Connolly, B. A. Eckstein, F., Pingoud, A. (1984) The stereochemical course of the restriction endonuclease EcoRI-catalyzed reaction. *J. Biol. Chem*. 259, 10760-10763.

Cowan, J. A. (1998) Metal activation of enzymes in nucleic acid biochemistry. *Chem. Rev*. 98, 1067-1088.

Crestfield, A. M., Stein, W. H., Moore, S., (1962) On the aggregation of bovine pancreatic ribonuclease. *Arch. Biochem. Biophys*. 1, 217-222.

Daniels, L. E., Wood, K. M., Scott, D. J., Halford, S. E. (2003) Subunit assembly for DNA cleavage by restriction endonuclease SgrAI. *J. Mol. Biol.* 327, 579-591.

de Laat, W. L., Jaspers, N. G., Hoeijmakers, J. H. (1999) Molecular mechanism of nucleotide excision repair. *Genes Dev.* 13, 768-785.

Deibert, M., Grazulis, S., Sasnauskas, G., Siksnys, V., and Huber, R. (2000) Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nat. Struct. Biol.* 7, 792-799.

DeLano, W.L. (2002) The PyMOL Molecular Graphics System.

Derewenda, Z. S. (2004). The use of recombinant methods and molecular engineering in protein crystallization. *Methods*, 34, 354-363.

Duderstadt, K. E., Chuang, K. Berger, J. M. (2011) DNA stretching by bacterial initiators promotes replication origin opening. *Nature* 478, 209-213.

Dulbecco, R. (1949) Reactivation of ultra-violet-inactivated bacteriophage by visible light. *Nature* 162, 949-950.

Dunten, P. W., Little, E. J., Gregory, M. T., Manohar, V. M., Dalton, M., Hough, D., Bitinaite, J., Horton. N. C. (2008) The structure of SgrAI bound to DNA; recognition of an 8 base pair target. *Nucleic Acids Research* 36, 16, 5405-5416.

Elrod-Erickson, M., Rould, M. A., Nekludova, L., Pabo, C. O. (1996) Zif268 protein–DNA complex refined at 1.6 Å: a model system for understanding zinc finger–DNA interactions *Structure* 4, 1171–1180.

Elrod-Erickson, M., Benson, T. E., Pabo, C. O. (1998) High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure* 6(4), 451-464.

Emsley, P., Cowtan, K., (2004) Coot: model-building tools for molecular graphics. *Acta. Crysta.* D60, 2126-2132.

Etzkorn, C., Horton, N. C. (2004a) Ca^{2+} binding in the active site of HincII: Implications for the catalytic mechanism. *Biochemistry* 43, 13256-13270.

Etzkorn, C., Horton, N. C. (2004b) Mechanistic insights from the structures of HincII bound to cognate DNA cleavage from addition of Mg^{2+} and Mn^{2+}. *J. Mol. Biol*. 343, 833-849.

Evans, P. R. (2006) Scaling and assessment of data quality. *Acta. Crysta*. D62, 72-82.

Fasman, G. D. (1975) *CRC Handbook of Biochem. and Mol. Biol*. 3rd ed. Cleveland, OH:CRC.

Finger, L. D., Shen, B., (2010) FEN1 (flap structure-specific endonuclease 1). *Atlas Genet. Cytogenet. Oncol. Haematol.* (http://AtlasGeneticsOncology.org/Genes/FEN1 ID40543ch11q12.html).

Fita, I., Rossmann, M. G. (1985) The NADPH binding site on beef liver catalase. *Proc. Natl. Acad. Sci*. 82, 1604-1608.

Friedberg, E. C., Walker, G. C., Siede, W., Wood, R. D., Schultz, R. A., Ellenberger, T. (2005) DNA Repair and Mutagenesis, 2nd ed. *American Society of Microbiology Press, Washington, DC.*

Friedman, R. A., Honig, B. (1995) A free energy analysis of nucliec acid base stacking in aqueous solution. *Biophys. J.* 69 (4), 1528-1535.

Friedrich-Heineken, E., Hubscher, U. (2004) The FEN-1 extrahelical 30-flap pocket is conserved from archaea to human and regulates DNA substrate specificity. *Nucleic Acids Res.*, 32, 2520–2528.

Fry, M., Loeb, L. A. (1999) Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n. *J. Biol. Chem.* 274, 12797-12802.

Gabrielsen, O. S., Sentenac, A., Fromageot, P. (1991) Specific DNA binding by c-Myb: Evidence for a double helix- turn-helix-related motif. *Science* 253, 1140–1143.

Garvie, C. W., Wolberger, C. (2001) Recognition of specific DNA sequences. *Mol. Cell*, 8, 937-946.

Gillet, L. C., Scharer, O. D. (2006) Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem. Rev.*, 106, 253–276.

Hatfield, G. W., Benham, C. J. (2002) DNA topology-mediated control of global gene expression in *E. coli*. *Ann. Rev. Gent*. 36, 175-203.

Hegde, R. S., Grossman, S. R., Laimins, L. A., Sigler, P. B. (1992) Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature* 359, 505–512.

Henneke, G., Friedrich-Heineken, E., Hubscher, U. (2003) Flap endonuclease 1: a novel tumor suppressor protein. *Trends Biochem. Sci.*, 28, 384-390.

Hiller, D. A., Fogg, J. M., Martin, A. M., Beechem, J. M., Reich, N. O., Perona, J. J. (2003). Simultaneous DNA binding and bending by EcoRV endonuclease observed by real-time fluorescence. *Biochemistry*, 42, 14375–14385.

Hingorani-Varma, K., Bitinaite, J. (2003) Kinetic analysis of the coordinated internaction of SgrAI restriction endonuclease with different DNA targets. *Journal of Biological Chemistry*, 278, 40392-40399.

Hodgson, D. A. (2000) Primary metabolism and its control in streptomycetes: a most unusual group of bacteria. *Adv. Microb. Physiol*. 42, 47-238.

Horton, N. C., Connolly, B. A., and Perona, J. J. (2000) Inhibition of EcoRV endonuclease by deoxyribo-3´-S-phosphorothiolates: a high resolution x-ray crystallographic study. *J. Am. Chem. Soc. 122*, 3314-3324.

Horton, N. C., Perona, J. J. (2001) Making the most of metal ions. *Nature Struct. Biol*. 8, 290-293.

Horton, N. C., Dorner, L. F., Perona, J. J. (2002) Sequence selectivity and degeneracy of a restriction endonuclease mediated by DNA intercalation. *Nature Struct. Biol*. 9, 42-47.

Hosfield, D. J., Mol, C. D., Shen, B., Tainer, J. A. (1998) Strucutre of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA andPCNA binding to FEN-1 activity. *Cell* 95. 135-146.

Huber, R., Kurr, M., Jannasch, H. W., Stetter, K. O. (1989) A novel group of abyssal methanogenic archaebacteria (*Methanopyrus*) growing at 110°C. *Nature* 342, 833-834.

Huntington J. A., Pannu, N. S., Hazes, B., Read, R. J., Lomas, D. A., Carrell, R. W. (1999) A 2.6 Å structure of a serpin polymer and implications for conformational disease. *J. Mol. Biol*. 293, 449-455.

Hwang, K. Y., Baek, K., Kim, H. Y., Cho, Y. (1998) The crystal structure of flap endonuclease-1 from *Methanococcus jannaschii*. *Nat. Struct. Biol*. 5, 707-713.

Ingerson-Mahar, M., Briegel, A., Werner, J. N., Jensen, G. J., Gitai, Z. (2010) The metabolic enzyme CTP synthase forms cytoskeletal filaments, *Nat. Cell Biol. 12*, 739-746.

Johnson, K. A. (1992) Transient state kinetic analysis of enzyme reaction pathways. *The Enzymes* 20, 1-61.

Jackson, D. A., Yanofsky, C. (1969) Restoration of enzymatic activity by complementation in vitro between mutant alpha subunits of tryptophan synthetase and between mutant subunits and fragments of the alpha subunit. *J. Biol. Chem*. 244, 4539-4546.

Jeltsch, A., Wenz, C., Stahl, F., Pingoud, A. (1996) Linear diffusion of the restriction endonuclease EcoRV on DNA is essential for the in vivo function of the enzyme. *EMBO J*. 15, 5104–5111.

Joshi, H. K., Etzkorn, C., Chatwell, L., Bitinaite, J., Horton, N. C. (2006) Alteration of sequence specificity of the Type II restriction endonuclease HincII through an indirect readout mechanism. *J. Biol. Chem.* 281, 33, 23852-23869.

Kabsch, W. (2010). *XDS. Acta Cryst.* D66, 125-132.

Kalifa, L. Beutner, G., Phadnis, N., Sheu, S. S., and Sia, E. A. (2009) Evidence for a role of FEN1 in maintaining mitochondrial DNA integrity. *DNA repair (Amst.)* 8, 1242-1249.

Kelly, T. J. Jr., Smith, H. O. (1970) A restriction enzyme from *Hemophilus influenzae*. II. *J. Mol. Biol.* 51, 393-409.

Kelman, Z., White, M. F. (2005) Archaeal DNA replication and repair. *Current opinion in microbiology* 8, 669-676.

Kelner, A. (1949) Effect of visible light on the recovery of *Streptomyces griseus conidia* from ultraviolet irradiation injury. *Proc. Natl. Acad. Sci.* 35, 73-79.

Kielkof, C. L., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., Rees, D. C. (1998) A structural basis for recognition of A.T and T.A base pairs in the minor groove of B-DNA. *Science* 282 (5386), 111-115.

Kim, Y., Geiger, J. H., Hahn, S., Sigler, P. B. (1993) Crystal structure of a yeast TBP/ TATA-box complex. *Nature* 365, 512–520.

Kim, C. W., Moon, Y. A., Park, S. W., Cheng, D., Kwon, H. J., Horton, J. D. (2010) Induced polymerization of mammalian acetyl-CoA carboxylase by MIG12 provides a tertiary level of regulation of fatty acid synthesis. *Proc. Natl. Acad. Sci. U S A 107*, 9626-9631.

Klose, J., Wendt, N., Kubald, S., Krause, E., Fechner, K., Beyermann, M., Bienert, M., Rudolph, R. and Rothemund, S. (2004). Hexa-histidin tag position influences disulfide structure but not binding behavior of in vitro folded N-terminal domain of rat corticotropin-releasing factor receptor type 2a. *Protein Sci.*13, 2470–2475.

Kong, H., Higgins, L. S., Dalton, M. A. (2000) Method for cloning and producing the SgrAI restriction endonuclease. US Patent 6048731 A. Filed September 3, 1998, Published April 11, 2000.

Kono, H., Sarai, A. (1999) Structure-based prediction of DNA target sites by regulatory proteins. *Proteins: Structure, Function, and Genetics* 35, 114-131.

Korennykh, A. V., Egea, P. F., Korostelev, A. A., Finer-Moore, J., Zhang, C., Shokat, K. M., Stroud, R. M., Walter, P. (2009) The unfolded protein response signals through high-order assembly of Ire1. *Nature 457*, 687-693.

Koudelka, G. B., Harrison, S. C., Ptashne, M. (1987) Effect of non-contacted bases on the affinity of 434 operator for 434 repressor and Cro. *Nature*, 326, 886–888.

Koudelka, G. B., Mauro, S. A., Ciubotaru, M. (2006) Indirect readout of DNA sequence by proteins: the roles of DNA sequence-dependent intrinsic and extrinsic forces. *Prog. Nucl. Acid Res Mol. Biol*. 81, 143-177.

Kovall, R. A., Matthews, B. W. (1999) Type II restriction endonuclease: structural, functional and evolutionary relationships. *Curr. Opin. Chem. Biol*. 3, 578-583.

Kozlov, A. G., Galletto, R., Lohman, T. M. (2012) SSB-DNA binding monitored by fluorescence intensity and anisotropy. *Methods Mol. Biol.* 922, 55-83.

Kueh, H. Y., Mitchison, T. J. (2009) Structural plasticity in actin and tubulin polymer dynamics. *Science* 325, 960-963.

Kurr, M., Huber, R., Konig, H., Jannasch, H. W., Fricke, H., Trincone, A., Kristjansson, J. K., Stetter, K. O. (1991) *Methanopyrus kandleri*, gen. and sp. Nov. represents a novel group of hyperthermophilic methanogens, growing at 110˚C. *Arch. Microbiol*. 156, 239-247.

Leslie, A. G. (1999) Integration of macromolecular diffraction data. *Acta Crystallogr. D Biol. Crystallogr.* 55, 1696-1702.

Li, J., McQuade, T., Siemer, A. B., Napetschnig, J., Moriwaki, K., Hsiao, Y. S., Damko, E., Moquin, D., Walz, T., McDermott, A., Chan, F. K., and Wu, H. (2012) The RIP1/RIP3 necrosome forms a functional amyloid signaling complex required for programmed necrosis, *Cell 150*, 339-350.

Lieber, M. R. (1997) The FEN-1 family of structure-specific nucleases in eukaryotic DNA replication, recombination and repair. *Bioessays*, 19, 233-240.

Little, E. J., Horton, N. C. (2005) DNA-induced conformational changes in Type II restriction endonucleases: The structure of unliganded HincII. *J. Mol. Biol*. 351, 76-88.

Little, E. J., Babic, A. C., Horton, N. C. (2008) Early interrogation and recognition of DNA sequence by indirect readout. *Structure* 16, 1828-1837.

Little, E. J., Dunten, P. W., Bitinaite, J., Horton, N. C. (2011) New clues in the allosteric activation of DNA cleavage by SgrAI: structures of SgrAI bound to cleaved primary-site DNA and uncleaved secondary-site DNA. *Acta Cryst. D* 67, 67-74.

Liu, Y., Kao, H. I., Bambara, R. A. (2004) Flap endonuclease 1: a central component of DNA metabolism. *Annu. Rev. Biochem*., 73, 589-615.

London, J., Skrzynia, C., Goldberg, M. E. (1974) Renaturation of Escherichia coli tryptophanase after exposure to 8M urea. Evidence for the existence of nucleation centers. *Eur. J. Biochem*. 47, 409-415.

Lukacs, C. M., Kucera, R., Schildkraut, I., Aggarwal, A. K. (2001). Structure of free BglII reveals an unprecedented scissor-like motion for opening an endonuclease. *Nature Struct. Biol.* 8, 126–130.

Lyumkis, D., Talley, H., Stewart, A., Shah, S., Park, C. K., Tama, F., Potter, C. S., Carragher, B., and Horton, N. C. (2013) Allosteric Regulation of DNA Cleavage and Sequence-Specificity through Run-On Oligomerization. *Structure* 21 (10) 1848-1858.

Ma, X., Shah, S., Zhou, M., Park, C. K., Wysocki, V. H., and Horton, N. C. (2013) Structural Analysis of Activated SgrAI-DNA Oligomers Using Ion Mobility Mass Spectrometry. *Biochemistry* 52, (25) 4373-4381.

Mast, A. E., Enghild, J. J., and Salvensen, G. (1992) Conformation of the reactive site loop of alpha 1-proteinase inhibitor probed by limited proteolysis. *Biochemistry* 31, 2720-2728.

Matsui, E., Musti, K. V., Abe, J., Yamasaki, K., Matsui, I., Harata, K. (2002) Molecular structure and novel DNA binding sites located in loops of flap endonuclease-1 from *Pyrococcus horikoshii*. *J. Biol. Chem.* 277, 37840-37847.

McClealand, S. E., Sczcelkun, M. D. (2004) The type I and II restriction endonucleases: structural elements in molecular motors that process DNA. In: Restriction Endonucleases 111-135, Pingoud A. (Ed), Springer, Berlin.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni L.C., and Read. R.J. (2007) *Phaser* crystallographic software. *J. Appl. Cryst.* **40**, 658-674.

McPherson A (1976) *Methods of Biochemical Analysis*. 23, 249-345, Academic Press, New York.

McRee, D. E. (1999) XtalView/Xfit-A versatile program for manipulating atomic coordinates and electron density. *Journal of Structural Biol.* 125 (2-3), 156.

Min, J. H., Pavletich, N. P. (2007) Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* 449, 570-575.

Mitchell, D. L., Nairn, R. S. (1989) The biology of the (6–4) photoproduct. *Photochem. Photobiol*. 49, 805–819.

Mu, D., Wakasugi, M., Hsu, D. S., Sancar,A. (1997) Characterization of reaction intermediates of human excision repair nuclease. *J. Biol. Chem.*, 272, 28971–28979.

Mucke, M., Kruger, D. H., Reuter, M. (2003) Diversity of Type II restriction endonucleases that require two DNA recognition sites. *Nuc. Acids Res*. 31 (21) 6079-6084.

Murante, R. S., Rust, L., Bambara, R. A. (1995) Calf 50 to 30 exo/endonuclease must slide from a 50 end of the substrate to perform structure-specific cleavage. *J. Biol. Chem.*, 270, 30377–30383.

Murray, N. E. (2000) Type I restriction systems: sophisticated moleculat machines (a legacy of Bertani and Weigle). *Microbiol. Mol. Biol. Rev*. 64, 412-434.

Murshudov, G. N., Vagin, A. A., Lebedev, A., *et al.* (1999) Efficient anisotropic refinement of macromolecular structures using FFT. *Acta. Crysta. D Biol. Crysta.* 55 (PT 1), 247.

Nagarkar, R. P., Hule, R. A., Pochan, D. J., Schneider, J. P. (2010) Domain swapping in materials design, *Biopolymers 94*, 141-155.

Nagata, S. (2007) Autoimmune diseases caused by defects in clearing dead cells and nuclei expelled from erythroid precursors. *Immunol. Rev.* 220, 237-250.

Nance, M. A., Berry, S. A. (1992) Cockayne syndrome: review of 140 cases. *Am. J. Med. Genet.* 42, 68-84.

Narayanaswamy, R., Levy, M., Tsechansky, M., Stovall, G. M., O'Connell, J. D., Mirrielees, J., Ellington, A. D., Marcotte, E. M. (2009) Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation, *Proc. Natl. Acad. Sci. U S A 106*, 10147-10152

Newman, M., Strzelecka, T., Dorner, L. F., Schildkraut, I., Aggarwal, A. K. (1995) Structure of BamHI endonuclease bound to DNA: partial folding and unfolding on DNA binding. *Science*, 269, 656–663.

Newman, M., Murray-Rust, J., Lally, J., Rudolf, J., Fadden, A., Knowles, P. P., White, M. F., McDonald, N. Q. (2005) Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition. *EMBO J.* 24, 895-905.

Noree, C., Sato, B. K., Broyer, R. M., Wilhelm, J. E. (2010) Identification of novel filament-forming proteins in Saccharomyces cerevisiae and Drosophila melanogaster. *The Journal of cell biology 190*, 541-551.

O'Donovan, A., Davies, A. A., Moggs, J. G., West, S. C., Wood, R. D. (1994) XPG endonuclease makes the 30 incision in human DNA nucleotide excision repair. *Nature,* 371, 432–435.

Ogrunc, M., Becker, D. F., Ragsdale, S.W., Sancar, A. (1998) Nucleotide excision repair in the third kingdom. *J. Bacteriol*. 180, 5796-5798.

Ohnishi, Y., Ishikawa, J., Hara, H., Suzuki, H., Ikenoya, M., Ikeda, H., Yamashita, A., Hattori, M., Horinouchi, S. (2008) Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J. Bact*. 190 (11) 4050-4060.

Otwinowski, Z., Schevitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., Sigler, P. B. (1998) Crystal structure of trp repressor/operator complex at atomic resolution. *Nature*, 335, 321-329.

Parge, H. E., Arvai, A. S., Murtari, D. J. *et al.* (1993) Human CksHs2 atomic structure: a role for its hexameric assembly in cell cycle control. *Science* 262, 387-95.

Park, C. K., Joshi, H. K., Agrawal, A., Ghare, M. I., Little, E. J., Dunten, P. W., Bitinaite, J., Horton, N. C. (2010a) Domain swapping in Allosteric modulation of DNA specificity. *PLoS Biol.* 8 (12): e1000554.

Park, C. K., Stiteler, A., Shah, S., Ghare, M. I., Bitinaite, J., and Horton, N. C. (2010b) Activation of DNA Cleavage by Oligomerization of DNA Bound SgrAI. *Biochemistry* 49 (41) 8818-8830.

Parkhurst, K.M., Brenowitz,M., Parkhurst, L. J. (1996). Simultaneous binding and bending of promoter DNA by the TATA binding protein: real time kinetic measurements. *Biochemistry*, 35, 7459–7465.

Parrish, J. Z., Xue, D. (2006) Cuts can kill: the roles of apoptotic nucleases in cell death and animal development. *Chromosoma* 115, 89-97.

Perona, J. J., Martin, A. M. (1997) Conformational transitions and structural deformability of EcoRV endonuclease revealed by crystallographic analysis. *J. Mol. Biol*. 273, 207–225.

Petit, C., Sancar, A. (1999) Nucleotide excision repair: from *E. coli* to man. *Biochimie*. 81, 15–25.

Pettijohn, D. E., Hanawalt, P. C. (1964) Evidence for repair-replication of ultraviolet damaged DNA in bacteria. *J. Mol. Biol*. 9, 395–410.

Pingoud, A., Jeltsch, A. (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res.* 29 (18) 3705-3727.

Pingoud, A., Fuxreiter, M., Pingoud, V., Wende, W. (2005) Type II restriction endonuclease: structure and mechanism. *CMLS, Cell. Mol. Life Sci*. 62, 685-707.

Reid, S. L., Parry, D., Liu, H. H., Connolly, B. A. (2001) Binding and recognition of GATATC target sequences by the EcoRV restriction endonuclease: a study using fluorescent oligonucleotides and fluorescence polarization. *Biochemistry* 40, 2484-2494.

Remington, S., Wiegand, G., Huber, R. (1982) Crystallographic refinement and atomic models of two different forms of citrate synthase at 2.7 and 1.7 Å resolution. *J. Mol. Biol*. 158, 111-52.

Rhodes, D., Schwabe, J. W. R., Chapman, L., Fairall, L. (1996) Towards an understanding of protein-DNA recognition. *Phil. Trans. R. Soc. Lond. B* 351, 501-509.

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

Rice, P. A. (1997) Making DNA do a U-trun: IHF and related proteins. *Curr. Op. Struct. Biol*. 7, 86-93.

Roberts, R. J., Halford, S. E. (1993) Type II restriction enzymes. In: *Nucleases,* 35-88, Linn, S. M., Lloyd, R. S., Roberts, R. J. (Eds), Cold Spring Harbor Laboratory Press, Cold Srping Harbor, NY.

Roberts, R. J., Macelis, D. (2001) REBASE-restriction enzymes and methylases. *Nuc. Acids Res.* 29, 268-269.

Roberts, R. J., Belfort, M., Bestor, T. *et al.* (2003) A nomenclature for restriction enzymes, DNA methyltransferase, homing endonucleases and their genes. *Nuc. Acids Res*. 31(7) 1805-1812.

Robins, P., Pappin, D. J., Wood, R. D., Lindahl, T. (1994) Structural and functional homology between mammalian DNase IV and the 50-nuclease domain of Escherichia coli DNA polymerase I. *J. Biol. Chem.*, 269, 28535–28538.

Rodgers, D. W., Harrison, S. C. (1993) The complex between phage 434 repressor DNA-binding domain and operator site OR3: Structural differences between consensus and non-consensus half-sites. *Structure* 1, 227–240.

Sakurai, S., Kitano, K., Yamaguchi, H., Hamada, K., Okada, K., Fukuda, K., Uchida, M., Ohtsuka, E., Morioka, H., Hakoshima, T. (2005) Structural basis for recruitment of human flap endonuclease 1 to PCNA. *EMBO J.* 24, 683-693.

Sancar, A. (1996) DNA excision repair. *Annu. Rev. Biochem*. 65, 43–81.

Sapranausks, R., Sasnauskas, G., Lagunavicius, A., Vilkaitis, G., Lubys, A., Siksnys, V. (2000) Novel subtype of type IIs restriction enzymes. *Bfi*I endonuclease exhibits similarities to the EDTA-resistant nuclease Nuc of *Salmonella typhimurium. J. Biol. Chem.* 275, 30878-30885.

Sarai, A., Selvaraj, S., Gromiha, M. M., Siebers, J. G., Prabakan, P., Kono, H. (2001) Target prediction of transcription factors: Refinement of structure-based method. *Genome Infor*. 12, 384-385.

Sawaya, M. R., Guo, S., Tabor, S., Richardson, C. C., Ellenberger, T. (1999) Crystal structure of the helicase domain from the replicative helicase-primase of bacteriophage T7. *Cell* 99, 167-177.

Scherly, D., Nouspikel, T., Corlet, J., Ucla, C., Bairoch, A. and Clarkson, S. G. (1993) Complementation of the DNA repair defect in xeroderma pigmentosum group G cells by a human cDNA related to yeast RAD2. *Nature*, 363, 182–185.

Scheviz, R. W., Otwinowski, Z., Joachimiak, A., Lawson, C. L., Sigler, P. B. (1985) The three dimensional structure of trp repressor. *Nature* 317, 782–786.

Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* 78 (3), 1606-1619.

Schwabe, J. W. R., Chapman, L., Finch, J. T., Rhodes, D. (1993) The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: How receptors discriminate between their response elements. *Cell* 75, 567–578.

Segal, D. J., Crotty, J.W., Bhakta, M. S., Barbas, C. F., III, Horton, N. C. (2006) Structure of Aart, a designed six-finger zinc finger peptide, bound to DNA. *J. Mol. Biol*. 363, 405–421.

Setlow, R. B., Carrier, W. L. (1964) The disappearance of thymine dimers from DNA: an error-correcting mechanism. *Proc. Natl. Acad. Sci.* USA 51, 226–231.

Siksnys, V., Skirgaila, R., Sasnauskas, G., Urbanke, C., Cherny, D., Grazulis, S., and Huber, R. (1999) The Cfr10I restriction enzyme is functional as a tetramer. *J. Mol. Biol.* 291, 1105-1118.

Smith, M. C., Furman, T. C., Ingolia, T. D., and Pidgeon, C. (1988) Chelating peptide-immobilized metal ion affinity chromatography. A new concept in affinity chromatography for recombinant proteins. *J. Biol. Chem*. 263, 7211-7215.

Somers, W. S., Phillips, S. E. V. (1992) Crystal structure of the met repressor-operator complex at 2.8 Å resolution reveals DNA recognition by b-strands. *Nature* 359, 387–393.

Spolar, R. S., Record, M. T., Jr. (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science*, 263, 777–784.

Staniforth, R. A., Giannini, S., Higgins, L. D., Conroy, M. J., Hounslow, A. M., Jerala, R., Craven, C. J., Waltho, J. P. (2001) Three-dimensional domain swapping in the folded and molten-globule states of cystatins, an amyloid-forming structural superfamily, *EMBO J 20*, 4774-4781.

Stefanini, M., Vermeulen, W., Weeda, G., Giliani, S., Nardo, T., *et al*. (1993) A new nucleotide-excision-repair gene associated with the disorder trichothiodystrophy. *Am. J. Hum. Genet.* 53, 817-21.

Steffen, N. R., Murphy, S. D., Lathrop, R. H., Opel, M. L., Tolleri, L., Hatfield, G. W. (2002) The role of DNA deformation energy at individual base steps for the identification of DNA-protein binding sites. *Gen. Infor*. 13, 153-162.

Storici, F., Henneke, G., Ferrari, E., Gordenin, D. A., Hubscher, U., Resnick, M. A. (2002) The flexible loop of human FEN1 endonuclease is required for flap cleavage during DNA replication and repair. *EMBO J.*, 21, 5930–5942.

Storoni, L. C., McCoy, A. J., Read, R. J. (2001) Likelihood-enhanced fast rotation functions. *Acta Crysta D Biol Crysta* (2004) 60 (pt 3), 432; Read, R. J.; Pushing the boundaries of molecular replacement with maximum likelihood. *Acta. Crysta. D Biol. Crysta.* 57 (Pt 10) 1373.

Story, R. M., Weber, I. T., Steitz, T. A. (1992) The structure of the *E. coli* recA protein monomer and polymer. *Nature* 355, 318-25.

Strop, P., Smith, K. S., Iverson, T. M., Ferry, J. G., Rees, D. C. (2001) Crystal structure of the "cab"-type β class carbonic anhydrase from the archaeon *Methanobacterium thermoautotrophicum. J. Biol. Chem*. 276, 10299-10305.

Stroud, J. C., Liu, C., Teng, P. K., Eisenberg, D. (2012) Toxic fibrillar oligomers of amyloid-beta have cross-beta structure, *Proc Natl Acad Sci U S A 109*, 7717-7722.

Tajika, Y., Sakai, N., Tamura, T., Yao, M., Watanabe, N. and Tanaka, I. (2004). Crystal structure of hypothetical protein PH0828 from *Pyrococcus horikoshii*. *Proteins*, 57, 862–865.

Tautz, N. Kaluza, K., Frey, B., Jatsch, M., Schmitz, G. G., Kessler, C. (1990a) SgrAI, a novel class-II restriction endonuclease from Streptomyces griseus recognizing the octanucleotide sequence 5´-CR/CCGGYG-3´ [corrected]. *Nucleic Acids Res*. 18, 3087.

Tautz, N., Kaluza, G., Laue, F., Frey, B., Schmitz, G., Jarsch, M., Ankenbauer, W., Kessler, C. (1990b) SgrAI, a unique restriction endonuclease from *Streptomyces griseus* which recognizes 5´CrCCGGyG 3´. *Fresenius J. Anal. Chem*. 337, 110.

Thoms, K. M., Kuschal, C., Emmert, S. (2007) Lessons learned from DNA repair defective syndromes. *Experimental Dermatology* 16, 532-544.

Tomlinson, C. G., Atack, J. M., Chapados, B. R., Tainer, J. A., Grasby, J. A. (2010) Substrate recognition and catalysis by flap endonucleases and related enzymes. *Biochem. Soc. Trans*. 38, 433-437.

Tsutakawa, S. E., Classen, S., Chapados, B. R., Arvai, A. S., Finger, L. D., *et al.* (2011) Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell* 145, 198-211.

Tumey, L. N., Bom, D., Huck, B., Gleason, E., Wang, J., Silver, D., Brunden, K., Boozer, S., Rundlett, S., Sherf, B., *et al.* (2005) The identification and optimization of a N-hydroxy urea series of flap endonuclease 1 inhibitors. *Biorog. Med. Chem. Lett.* 15, 277-281.

Vanamee, E. S., Viadiu, H., Kucera, R., Dorner, L., Picone, S., Schildkraut, I., and Aggarwal, A. K. (2005) A view of consecutive binding events from structures of tetrameric endonuclease SfiI bound to DNA. *EMBO J*. 24, 4198-4208.

Vencloves, C., Timinskas, A., Sisksnys, V. (1994) Five-stranded beta-sheet sandwiched with two alphs-helices: a structural link between restriction endonucleases EcoRI and EcoRV. *Proteins* 20, 279-282.

Viadiu, H., Aggarwal, A. K. (2000) Structure of BamHI bound to nonspecific DNA: a model for DNA sliding. *Mol. Cell*, 5, 889–895.

Vistica, J. *et al.* (2004) Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. *Anal. Biochem.* 326, 234-256.

Wainwright, M., Skiba, U. and Betts, R.E (1984) Sulphur oxidation by a *Streptomyces* sp.

growing in a carbon-deficient medium and autoclaved soil. *Arch. Microbiol*. 139, 272-276.

Wakasugi, M., Reardon, J. T., Sancar,A. (1997) The non-catalytic function of XPG protein during dual incision in human nucleotide excision repair. *J. Biol. Chem.,* 272, 16030–16034.

Waksman, S. A. (1953) Streptomycin: background, isolation, properties, and utilization. *Science* 118, 259-266.

Waugh, D. S. (2005). Making the most of affinity tags. *Trends Biotechnol*. 23, 316-320.

Wegner, A. (1976) Head to tail polymerization of actin. *J. Mol. Biol*. 108, 139-150.

Welsh, A. J., Halford, S. E., Scott, D. J. (2004) Analysis of type II restriction endonucleases that interact with two recognition sites. In: *Restriction Endonucleases,* 297-317, Pingoud, A. (ed.), Springer, Berlin.

Whitaker, R. D., Dorner, L. F., Schildkraut, I. (1999) A mutant of BamHI restriction endonuclease which requires N6-methyladenine for cleavage. *J. Mol. Biol.* 285, 1525-1536.

Wishner, B. C., Ward, K. B., Lattman, E. E., and Love, W. E. (1975) Crystal structure of sickle-cell deoxyhemoglobin at 5 Å resolution. *J. Mol. Biol*. 98, 179-194.

Winter, R. B., von Hippel, P. H. (1981) Diffusion driven mechanisms of protein translocation on nucleic acids. 2. The Escherichia coli repressor–operator interaction: equilibrium measurements. *Biochemistry,* 20, 6948–6960.

Woese, C. R., Kandler, O., Wheelis, M. L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci.* USA 87 (12) 4576-4579.

Wolfe, S. A., Nekludova, L., Pabo, C. O. (1999) DNA Recognition by Cys_2His_2 Zinc Finger Proteins. *Annu. Rev. Biophys. Biomol. Struct.* 3, 183–212.

Wood, R. D. (1999) DNA damage recognition during nucleotide excision repair in mammalian cells. *Biochimie*. 81, 39-44.

Wu, L., Vertino, A., Koudelka, G. B. (1992) Non-contacted bases affect the affinity of synthetic P22 operators for P22 repressor. *J. Biol. Chem*. 267, 9134–9135.

Xing, X., Bell, C. E. (2004) Crystal structures of *Escherichia coli* RecA in a compressed helical filament. *J. Mol. Biol*. 342, 1471-1485.

Yang, C. C., Nash, H. A. (1989) The interaction of *E. coli* IHF protein with its specific binding sites. *Cell* 57, 869–880.

Yoshimori, R., Toulland-Dussoix, D., Boyer, H. W. (1972) R factor-controlled restriction and modification of deoxyribonucleic acid: restriction mutants, *J. Bacteriol*. 112, 1275-1279.

Zheng, L. Zhou, M., Chai, Q., Parrish, J., Xue, D., Patrick, S. M., Turchi, J. J., Yannone, S. M., Chen, D., Shen, B. (2005) Novel function of the flap endonuclease I complex in processing stalled DNA replication forks. *EMBO reports* 6, 83-89.