

# REQUIREMENTS FOR CATALYSIS IN CRE RECOMBINASE

**BRYAN PAUL STEWART GIBB**

A Dissertation in

Biochemistry and Molecular Biophysics

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the  
Requirements for the Degree of Doctor of Philosophy

2010



Dr. Gregory Van Duyne, Professor of Biochemistry and Biophysics  
Supervisor of Dissertation



Dr. Kathryn M. Ferguson, Associate Professor of Physiology  
Graduate Group Chairperson

## Dissertation Committee

Dr. Mitch Lewis, Professor of Biochemistry and Biophysics

Dr. Ronen Marmorstein, Wistar Professor of Biochemistry and Biophysics

Dr. Kim Sharp, Associate Professor of Biochemistry and Biophysics

Dr. Kristen Lynch, Associate Professor of Biochemistry and Biophysics

Dr. James Shorter, Assistant Professor of Biochemistry and Biophysics

Dr. Kathryn M. Ferguson, Associate Professor of Physiology

## ACKNOWLEDGMENTS

I owe a great deal of gratitude to my family, friends, co-workers, and Dr. Greg Van Duyne. This dissertation would not have been possible without their help, advice and support.

Greg has been a fantastic mentor whom I will always look up to (figuratively speaking). While in the lab I have compiled a power list of skills at the bench and learned valuable lessons in critical thinking and reasoning. Greg allowed me freedom to explore all of my crazy ideas even though I'm sure he knew that many of them would fail. I have been fortunate to work in his lab and I greatly appreciate the opportunity as I am a better scientist for it.

I would like to thank the past and present members of the Van Duyne lab for their constant support and discussions, science related or not. I would specifically like to thank Dr. Kaushik Ghosh and Dr. Kushol Gupta for teaching me 'the ropes' of Cre biochemistry and crystallization. I had many helpful scientific discussions with Dr. Kay Perry, Dr. Peng Yuan, Dr. Karen Rutherford, and Renne Carrington. James Chen helped with Cre assays.

I have made many friends while at Penn, but would especially like to acknowledge Dr. David Hokanson, Dr. Bob Daber, and Tammer Farid. Without Dave, I would not have chosen to attend graduate school at UPenn, or taken up brewing beer as a hobby. Bob and I commuted from New Jersey for several years and was a constant source of support. Tammer, aside from providing a place for me to live for a year has been a great friend and instigator in acquiring a hobby for working on cars.

To my family, thank you for always being supportive. To my lovely wife Jennifer: You have been with me through all of this. Without your unwavering support I would not be here today. Thank you so much for your patience and understanding. To my parents, thank you for being a constant source of motivation by never failing to ask me 'so when do you think you are going to finish?'

I must also thank my thesis committee, Dr. Mitch Lewis, Dr. Ronen Marmorstein, Dr. Kate Ferguson, and Dr. Kim Sharp for their input.

## ABSTRACT

### REQUIREMENTS FOR CATALYSIS IN CRE RECOMBINASE

Bryan P.S. Gibb

Gregory Van Duyne, Ph.D., Thesis Advisor

Cre recombinase, a member of the tyrosine recombinase (YR) family of site-specific recombinases catalyzes DNA rearrangements using phosphoryl transfer chemistry that is identical to that used by the type IB topoisomerases (TopIBs). In this dissertation, the requirements for YR catalysis and the relationship between the YRs and the TopIBs are explored. I have analyzed the *in vivo* and *in vitro* recombination activities of all possible substitutions of the seven active site residues in Cre recombinase. To facilitate the interpretation mutant activities, I also determined the structure of a vanadate transition state mimic for the Cre-loxP reaction that allows for a comparison with similar structures from the related TopIBs. The results demonstrate that active site residues shared by the TopIBs are most sensitive to substitution. Two of the conserved active site residues in YRs have no equivalent in TopIBs. I have concluded that Glu176 and His289 in Cre evolved to have functional roles in site-specific recombination, that are unnecessary for relaxation by TopIB. His289 is not essential for cleavage of DNA, but accelerates water mediated hydrolysis of the 3'-phosphotyrosine covalent intermediate. Glu176 serves a structural role in facilitating the activation of Cre for cleavage by helping position the general acid, K201 in the active site. These residues may compensate for activity lost during the evolution of allosteric regulation required for recombination. Examination of two regions involved in protein:protein interactions in the Cre-DNA tetramer complex has led to the characterization of two modes of allosteric regulation. The first involves a C-terminal helix (hN), which binds in the pocket of a neighboring Cre monomer to serve as a regulatory switch by acting as a tether to

position the tyrosine nucleophile (Y324). The second regulatory module involves the mobile  $\beta$ 2- $\beta$ 3 hairpin carrying the general acid, K201. This hairpin must form synapsis dependent contacts to efficiently position K201 near O5' of the scissile phosphate. Cre has evolved to modulate the positions of critical residues for catalysis, which is an effective and highly sensitive mechanism of regulation of catalysis not shared by TopIBs.

# Table of Contents

<b>Chapter 1. Introduction.....</b>	<b>1</b>
1.1 Site specific recombination.....	1
1.2 Tyrosine recombinase family members.....	5
1.2.1 Type Ib topoisomerases.....	6
1.2.2 Cre and Flp: Simple YRs.....	7
1.2.3 $\lambda$ -int and XerCD: Complex YRs.....	8
1.3 Recombinases as tools for DNA manipulation.....	11
1.4 Mechanism of recombination.....	12
1.4.1 Phosphoryl-transfer chemistry.....	12
1.4.2 Architecture.....	19
1.5 Thesis objectives.....	24
<b>Chapter 2. Requirements for catalysis in the Cre active site (adapted from Gibb, B. et al. Nucleic Acids Research. 2010.).....</b>	<b>26</b>
2.1 Introduction.....	26
2.2 Structure of a Cre-loxP transition state mimic.....	29
2.3 Comprehensive analysis of Cre active site substitutions.....	35
2.4 In vitro recombination of Cre mutants.....	39
2.5 Results and discussion of the active site.....	45
2.5.1 Tyr324 substitutions.....	45
2.5.2 Arg173 substitutions.....	47
2.5.3 Lys201 substitutions.....	48
2.5.4 Substitutions at Arg292.....	49
2.5.5 Substitutions at Trp315.....	51

2.5.6 His289 is unique to tyrosine recombinases.....	53
2.5.7 Glu176 organizes the active site.....	55
2.6 Discussion.....	56
2.7 Methods.....	60
<b>Chapter 3. A Dual Role in Catalysis for H289 in Cre.....</b>	<b>68</b>
3.1 Introduction.....	68
3.2 Studies of all 19 H289 mutants.....	69
3.2.1 H289 mutant binding and synapsis.....	71
3.2.2 H289 mutants in suicide cleavage.....	73
3.2.3 H289 mutants in equilibrium cleavage.....	76
3.2.4 H289 mutant summary.....	78
3.3 Detailed investigation of H289Q and H289A.....	78
3.3.1 Synapsis of H289A and H289Q .....	78
3.3.2 Rate of in vitro recombination for H289Q.....	79
3.3.3 pH rate profile of Cleavage.....	83
3.3.4 H289Q strand cleavage preference.....	86
3.3.5 pH profile of equilibrium cleavage.....	88
3.3.6 pH Dependence of Hydrolysis.....	90
3.3.7 Activities of H289Q and H289A in ligation.....	92
3.4 Discussion.....	96
3.4.1 Top Strand Bias.....	100
3.4.2 Hydrolysis.....	100
3.4.3 H289 in ligation.....	102
3.4.4 Summary.....	102
3.5 Methods.....	103
3.5.1 Cloning and purification of Cre mutants.....	103

3.5.2	LoxP assay substrate purification .....	104
3.5.3	Cre Binding assay.....	107
3.5.4	Cre Synapsis Assay.....	107
3.5.5	Cleavage Assays.....	108
3.5.6	Ligation assay with pNP substrate.....	111
3.5.7	In vitro Recombination assay.....	111
<b>Chapter 4. E176 is part of a regulatory switch involving the general acid</b>		
<b>K201: characterization of the residues involved in general acid catalysis 112</b>		
4.1	Introduction.....	112
4.2	Characterization of E176 mutants.....	113
4.3	Characterizing the role of R173.....	116
4.4	Thiol rescue of Cre mutants at R173, K201 and E176.....	122
4.5	Discussion.....	134
4.5.1	E176 in general acid catalysis.....	134
4.6	Methods.....	145
4.6.1	Cloning and Cre purification.....	145
4.6.2	EMSA assays.....	145
4.6.3	In vitro recombination.....	145
4.6.4	Concentration dependent cleavage with 5'-bridging phosphorothiolate substrate...	146
4.6.5	Cleavage assays for thio-effect.....	146
<b>Chapter 5. Allosteric Regulation of Cleavage in Cre.....148</b>		
5.1	Introduction.....	148
5.2	Minimal complex for cleavage of loxP.....	154
5.3	Characterization of the Cre C-terminal deletion mutant.....	159
5.4	Mechanism of regulation by hN.....	163
5.4.1	Helix-N binding pocket occupancy is insufficient for Cre activity.....	163

5.4.2 Helix-N as a tether for Y324.....	166
5.5 Discussion.....	172
5.5.1 Assembly of an active Cre-loxP complex.....	172
5.5.2 Regulation of Cre activity by the C-terminal helix.....	176
5.6 Chapter specific methods.....	181
5.6.1 Cloning and Cre purification.....	181
5.6.2 EMSA assays.....	181
5.6.3 Concentration dependent cleavage as dimer.....	181
5.6.4 Complementation cleavage assay.....	181
5.6.5 hN peptide cleavage.....	182
<b>Chapter 6. Conclusions and Perspectives.....</b>	<b>183</b>
6.1 Summary of conclusions.....	183
6.2 Tolerance to change in the active site.....	184
6.3 H289 accelerates water mediated hydrolysis.....	184
6.4 E176 plays a structural role.....	186
6.5 K201 as a regulatory catalytic switch.....	187
6.6 Cre cleaves loxP as a dimer.....	188
6.7 The C-terminus of Cre regulates activity through Y324.....	188
<b>Appendix A.Cre Mutant Reference Table.....</b>	<b>190</b>
<b>Appendix B.Oligonucleotides and annealed substrates used in Cre assays</b> .....	<b>193</b>

## Illustration Index

Figure 1.1	Examples of site-specific recombination.....	3
Figure 1.2	Reaction mechanisms.....	4
Figure 1.3	Site specific recombination by $\lambda$ -integrase.....	9
Figure 1.4	Proposed mechanism of phosphoryl transfer in YRs and TopIB.....	16
Figure 1.5	Pre-cleavage and covalent intermediate active sites of Cre and TopIB.....	17
Figure 1.6	Active site of Variola topoisomerase transition state intermediate.....	18
Figure 1.7	Crystal structures of higher-order Cre-DNA complexes. ....	21
Figure 1.8	HJ intermediate structures of YRs.....	23
Figure 2.1	Structure of Cre-DNA-vanadate transition state mimic.....	30
Figure 2.2	The Cre-DNA-vanadate transition state mimic active site. ....	33
Figure 2.3	Cre and TopIB active sites in the transition state.....	34
Figure 2.4	In vivo recombination activities of Cre active site mutants.....	37
Figure 2.5	Cre concentration in vivo.....	38
Figure 2.6	In vitro loxP-binding for Cre H289 substitutions.....	40
Figure 2.7	In vitro recombination activities of Cre active site mutants.....	41
Figure 2.8	Identification of intermediates formed during in vitro recombination. ....	43
Figure 2.9	Y324T defective in binding, but soluble.....	46
Figure 3.1	Binding and synapsis of purified H289 mutants.....	72
Figure 3.2	Suicide cleavage of purified H289 mutants at 1 hour and 24 hours.....	75
Figure 3.3	Equilibrium cleavage of intact loxP at 1 hour and 24 hours.....	77
Figure 3.4	Synapsis of WT, H289Q and H289A.....	80
Figure 3.5	In vitro recombination time course.....	81
Figure 3.6	Representative cleavage on BS-nicked substrate at pH 7.5.....	84
Figure 3.7	Initial Velocity of Cre Cleavage vs. pH.....	85
Figure 3.8	Representative data for strand cleavage preference.....	87

Figure 3.9	Equilibrium cleavage as a function of pH.....	89
Figure 3.10	pH profile of cleavage and hydrolysis timecourse.....	91
Figure 3.11	Ligation of pNP modified TS and BS loxP DNA substrates by Cre.....	95
Figure 3.12	Comparison of structure at H289.....	98
Figure 4.1	Catalytic activities of E176 mutants.....	114
Figure 4.2	In vitro recombination assay of the Cre R173 mutants.....	119
Figure 4.3	Binding and synapsis of R173 mutants.....	120
Figure 4.4	Cleavage by R173 mutants.....	121
Figure 4.5	Synapsis of R173 and K201A mutants.....	124
Figure 4.6	Concentration dependence in cleavage of 5'-bridging phosphorothiolate substrates .....	125
Figure 4.7	Rescue of cleavage activity with 5'-bridging phosphorothiolates.....	126
Figure 4.8	Cleavage time course with nicked and phosphorothiolate substrates.....	128
Figure 4.9	Cleavage by Cre mutants on nicked and phosphorothiolate substrates.....	132
Figure 4.10	Disruption of K201 by R173K.....	136
Figure 4.11	Variation in the position of catalytic arginine and lysine in Cre and TopIB.....	139
Figure 4.12	K201 is on a mobile structural element in Cre.....	140
Figure 4.13	Model for regulation of Cleavage in Cre by a K201 switch.....	144
Figure 5.1	Intersubunit interactions formed by Cre in the tetrameric-DNA complex.....	149
Figure 5.2	Sequence alignment of tyrosine recombinase C-terminal tails.....	150
Figure 5.3	C-terminal elements and binding compared in Cre and $\lambda$ -integrase.....	153
Figure 5.4	WT Cre cleaves the TS of loxP preferentially as a dimer.....	157
Figure 5.5	WT Cre cleaves nicked substrates as a dimer bound to loxP.....	158
Figure 5.6	Binding and Synapsis of Cre $\Delta$ C.....	161
Figure 5.7	Cre $\Delta$ C cleaves phosphorothiolate substrates.....	162
Figure 5.8	$\Delta$ C complementation assay.....	165
Figure 5.9	Binding and Synapsis of Glycine Insertion Mutants.....	167

Figure 5.10 Cleavage activity of glycine insertion mutants.....	169
Figure 5.11 The hM-hN linker conformation in active and inactive Cre subunits.....	170

## Index of Tables

Table 1.1: Tyrosine recombinase and topoisomerase IB active site residues.....	15
Table 2.1: Tyrosine recombinase and topoisomerase IB active site residues.....	27
Table 2.2: Crystallographic data for the Cre-DNA-vanadate transition state mimic.....	31
Table 2.3: Accumulation of Holliday junction and covalent intermediates during in vitro recombination.....	42
Table 2.4: Oligonucleotide sequences used in construction of in vivo recombination reporter .....	63
Table 3.1: H289 mutant activity summary.....	70
Table 3.2: In vitro recombination rates for WT Cre and H289Q.....	82
Table 3.3 Crystal statistics for H289Q and H289A loxP DNA complexes.....	99
Table 4.1: Initial velocities of E176 mutants in in vitro recombination.....	115
Table 4.2: Thio-effects for Cre and vaccinia virus topoisomerase.....	129
Table 4.3: Summary of activities of mutants at E176, R173 and K201.....	133
Table 5.1: Summary of activities for mutants characterized in study of hN.....	171
Table 6.1: Single stranded oligos that are annealed to form assay substrates.....	193
Table 6.2: Annealed DNA substrates for Cre assays constructed from single stranded oligonucleotides.....	194

# Chapter 1. Introduction

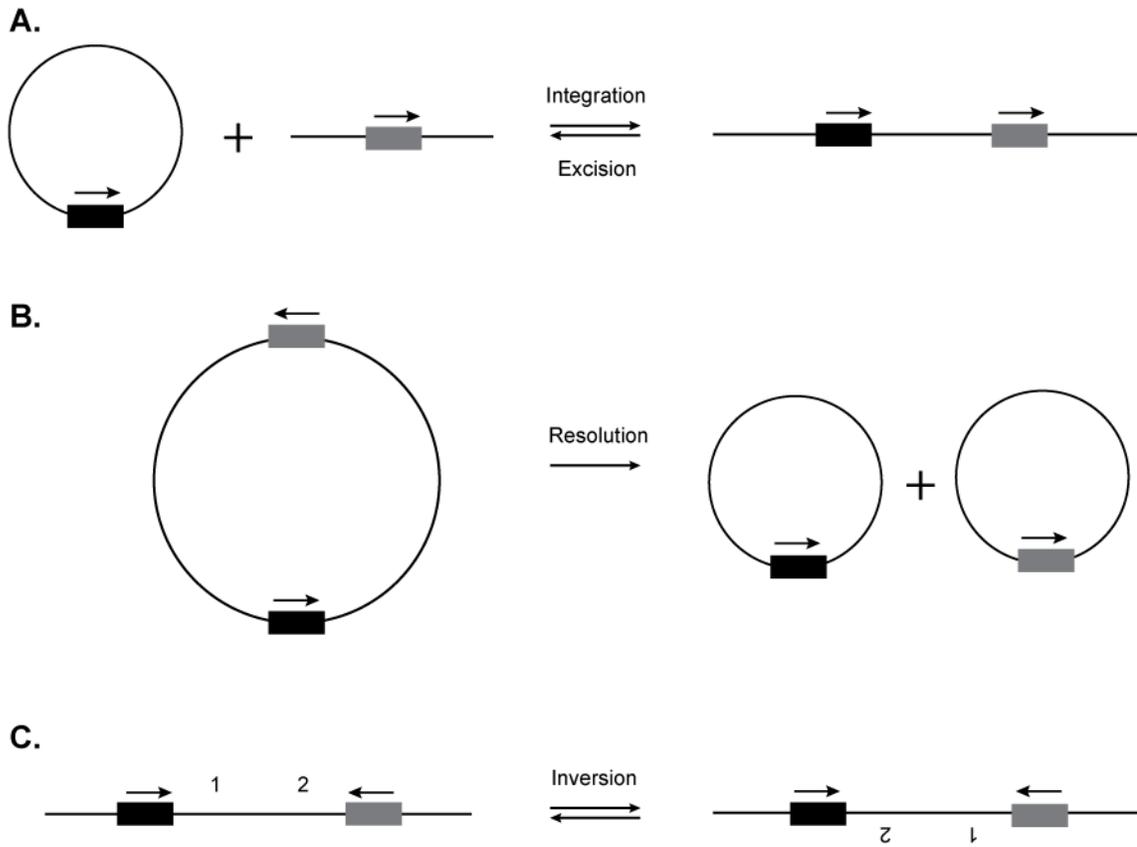
## 1.1 Site specific recombination

Site-specific recombination is a form of DNA rearrangement that involves reciprocal exchange between defined DNA sites. The products are rearranged versions of the starting DNA substrates that may have altered topology and sequences. Site-specific recombination enzymes that catalyze these DNA rearrangements are commonly found in prokaryotic systems and in some cases in yeast, where they perform various functions such as mediating the integration and excision of viral genomes into and out of host chromosomes, resolving multimeric plasmids to monomers in order to ensure faithful partitioning at cell division, regulating the copy number of yeast 2  $\mu$ m circles, and switching on and off expression of genes (Craig, Gragie, Gellert & Lambowitz, 2002).

Site-specific recombination involves two DNA sites that are recognized by a specialized recombinase that mediates DNA breakage and reunion of the phosphodiester backbone through a mechanism that does not require high-energy cofactors such as ATP. The orientation and topology of the recombination sites in the originating DNA molecules dictate one of three possible DNA rearrangements that are depicted in Fig. 1.1. Intramolecular recombination where the sites are in a direct (head-to-tail) orientation will result in the excision of the region flanked between the two sites. Resolution is a special form of excision where the substrate is a circular DNA molecule which is resolved to form two individual circular products. The reverse reaction, intermolecular recombination of sites located on two separate DNA molecules, will result in the integration of the two segments. Inversion is a form of intramolecular recombination where the two recombinase sites are oriented as inverted repeats (head-to-head) (for reviews see (Nash, 1996), (Van Duyne, 2008)).

Many site-specific recombination systems have been identified that perform these DNA rearrangements in a variety of biological roles, but nearly all can be grouped into one of two families; tyrosine recombinases (YR) and serine recombinases. These two families are capable of all three types of DNA rearrangements, but are unrelated by sequence or structure and have very different functional mechanisms. The serine recombinases are also known as the resolvase/invertase family and are defined by a catalytic serine nucleophile that is responsible for attacking the scissile phosphate to form a 5'-phosphoserine linkage. The proposed mechanism for serine recombinases involves concomitant cleavage of all four DNA strands by a tetramer of recombinase protomers to introduce double strand breaks at crossover in both recombination sites before strand exchange. Although the exact mechanistic details are not completely understood, a large rotation is thought to swap the strands for ligation to generate products (reviewed in (Grindley, Whiteson & Rice, 2006), (Van Duyne, 2008)). The tyrosine recombinases utilize a different mechanism. This family is also known as the lambda integrase family ( $\lambda$ -int) for the recombinase encoded by bacteriophage  $\lambda$  that controls integration and excision of phage DNA into and out of the bacterial genome. This family is defined by a conserved tyrosine nucleophile that attacks the scissile phosphate to form a 3'-phosphotyrosine linkage to the DNA.

YRs bind to a core site where cleavage and strand exchange reactions are catalyzed. The core sites contain two recombinase binding elements (RBEs) that vary in size and are usually but not always, inverted repeats that flank a short 6-8 bp central crossover region where cleavage and strand exchange occur (Fig. 1.2B). Like the serine recombinases, strand exchange occurs in a tetrameric complex of recombinase proteins containing two core DNA sites. However, unlike the serine recombinases, YRs perform cleavage in a stepwise manner with only two of the four recombinase subunits active at a time. Cleavage results in a 3'-phosphotyrosine covalent intermediate and release of a 5' hydroxyl group as shown schematically in Fig. 1.2A. Exchange of the cleaved strands and subsequent ligation forms a Holiday junction (HJ) intermediate. Isomerization of the HJ activates the second pair of recombinases for cleavage, strand exchange and ligation on the complementary set of strands to form products (Van Duyne, 2002).



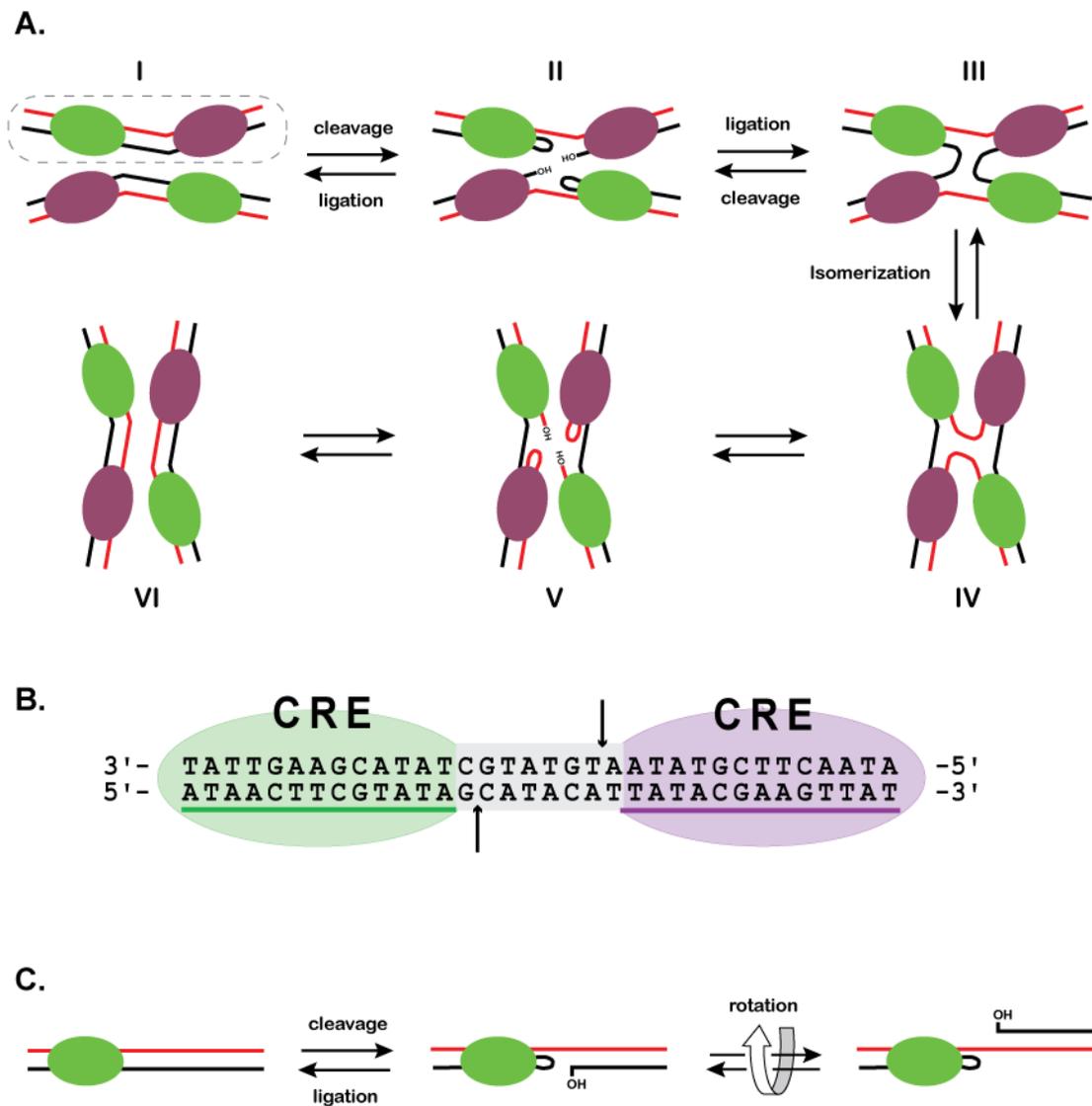
**Figure 1.1 Examples of site-specific recombination**

A. Intermolecular recombination as mediated by  $\lambda$ -int

B. Resolution, a form of intramolecular recombination occurs between two core sites in direct repeat (head-to-tail orientation), as observed for Cre, XerC/D, and Flp

C. Inversion is another form of intramolecular recombination that occurs between core sites oriented as inverted repeats (head-to-head orientation) as observed in Flp recombination.

Adapted from: (Van Duyne, 2008)



**Figure 1.2 Reaction mechanisms**

A. Recombination pathway for YRs as modeled after structures of Cre-DNA reaction intermediates.

B. The *loxP* site contains two 13 bp inverted repeat recombinase binding sites that flank an 8 bp central crossover region where cleavage and strand exchange take place. This central region contains asymmetry that confers directionality to the site by defining a top strand (TS) where cleavage is between an A-T base step AT, and a bottom strand (BS) where cleavage take place between G-C base step.

C. Schematic of topoisomerase reaction. The cleavage and ligation steps are mechanistically related to the equivalent step in the YR recombination pathway, except topoisomerase achieves relaxation as a monomer.

The mechanism of cleavage and ligation in YRs is shared by the type IB topoisomerases (TopIB), and the catalytic domains of YRs are also closely related to those found in the type IB topoisomerases. However, unlike YRs, TopIBs act as monomers to relax supercoiled DNA through the formation of a covalent 3' phosphotyrosine intermediate (Fig. 1.2C). The similarity of mechanism and active site residues in the catalytic domain led to speculation that topoisomerases and YRs are evolutionarily related, possibly sharing a common ancestor (reviewed in (Dong & Berger, 2008)).

## 1.2 Tyrosine recombinase family members

The YR family has been thoroughly studied through molecular genetics, biochemistry and structural biology. PsiBlast (Altschul *et al.*, 1997) searches indicate that the family consists of over 10,000 members (this work), however most of the knowledge regarding mechanism has been garnered from studies of  $\lambda$ -int, XerC, XerD, Flp and Cre recombinases (for reviews see, (Azaro & Landy, 2002),(Barre & Sherratt, 2002),(Jayaram, Grainge & Tribble, 2002),(Van Duyne, 2002)). Extensive mutagenesis and biochemistry performed on the type IB topoisomerase enzymes has contributed significantly to the understanding of the YRs. Consequently, the YRs can be broken down into three groups that correlate with the reaction complexity of the system.

The type IB topoisomerases, acting as monomers, make up the most basic group. The second group consists of recombinases that are often referred to as 'simple' members because they carry out site-specific recombination functions on simple substrates without the need for accessory factors or auxiliary DNA sequences. This group is best exemplified by the Cre and Flp recombinases. The third group consists of recombinases that require or exploit accessory proteins and auxiliary DNA sites for regulatory control of site-specific recombination.  $\lambda$ -int and XerCD are two well studied examples of 'complex' recombination systems.

### **1.2.1 Type Ib topoisomerases**

Type IB topoisomerases (TopIB) (for reviews see, (Dong & Berger, 2008) and (Champoux, 2001)) function as monomers to relax both positive and negative supercoils. This is done by introduction of transient breaks in one strand of the DNA duplex, allowing rotation about the intact strand. These enzymes have critical roles in cellular processes by relieving stress built up by overwinding or underwinding the DNA duplex that occurs during transcription, replication, and DNA repair. The type IB subfamily shows no sequence or structural homology to other known topoisomerases and is functionally distinct from the type IA subfamily. The TopIB family consists of eukaryotic cellular TopIBs, topoisomerases encoded by poxviruses, and a small number of bacterial TopIBs.

The eukaryotic TopIB enzymes are quite large, varying in size from 765 to 1019 amino acids. These enzymes contain an N-terminal domain which is dispensable for relaxation activity, but contains signals for nuclear localization and sites for regulatory interactions with other cellular proteins such as nucleolin, SV40, T-antigen, and a number of transcription factors. A central domain called the capping lobe is involved with mediating contacts with the DNA. The C-terminal catalytic domain contains the active site residues responsible for phosphoryl-transfer chemistry, but is split in half by a large helical domain.

The poxvirus topoisomerases are much smaller and more highly conserved, typically 314 to 333 amino acids in length. These topoisomerases share a similar domain architecture by retaining the capping lobe and a highly conserved catalytic domain, however they lack the N-terminal regulatory domain as well as the helical insertion in the catalytic domain found in the eukaryotic TopIBs. This has led to the suggestion that the viral TopIBs are minimal functional topoisomerases and have been the focus of extensive study.

Another interesting difference between eukaryotic and viral TopIB enzymes is that the poxvirus family exhibits a high degree of sequence specificity that has made the system amenable to detailed kinetic analysis. Recent structural studies of both eukaryotic and viral topoisomerases have explained the observed differences in binding specificity as well as similarities in the mechanism of relaxation with detailed views of the active site.

### **1.2.2 Cre and Flp: Simple YRs**

Cre recombinase (causes recombination) is a 38 kDa protein encoded by bacteriophage P1, and represents one of the simplest members of the YR family (Sternberg, Sauer, Hoess & Abremski, 1986). Cre recognizes and mediates recombination between two 34 bp sites called *loxP* (locus of X-over, phage; Fig. 1.2B) (Hoess & Abremski, 1984), (Hoess, Ziese & Sternberg, 1982). *LoxP* contains two 13 bp inverted repeats that serve as recombinase binding elements that flank an 8 bp asymmetric central cross over region that imparts directionality to the site. The protein consists of two domains, an N-terminal domain responsible for binding DNA and a C-terminal domain that contains all of the active site residues (Guo, Gopaul & van Duyne, 1997).

Cre has two roles in the biology of phage P1 (Sternberg *et al.*, 1981). First, Cre ensures that the 100-kb cyclically permuted and linear redundant genome is circularized after infection into *E.coli*. Generally, the host homologous recombination system fulfills this role, but should this system fail, Cre will efficiently circularize the infecting DNA at *loxP* sites located in the terminal redundancies.

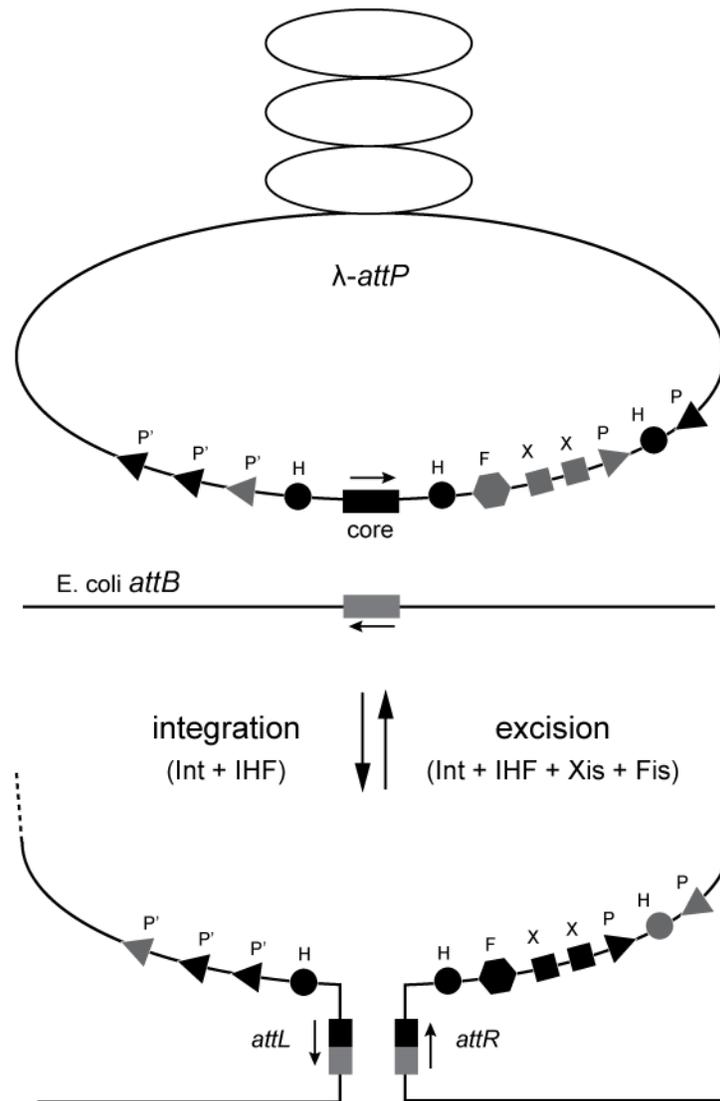
The second role for Cre in P1 is to ensure maintenance of the P1 DNA in lysogens (Sternberg *et al.*, 1981). The prophage genomes of many varieties of bacteriophages, including  $\lambda$  and P2 are normally integrated into the host chromosome. The bacteriophage P1 genome resides primarily as a unit copy plasmid as a lysogen. Homologous recombination can generate

dimers of the P1 genome after replication. To ensure fidelity in the progeny, the dimers are resolved efficiently by Cre mediated excision at *loxP* sites to provide monomers that are able to properly segregate.

The recombinase encoded by *Saccharomyces cerevisiae*, Flp, also functions without a requirement for accessory cofactors and binding sites (reviewed in, (Sadowski, 1995) and (Jayaram *et al.*, 2002)). Flp recognizes two 34 bp sites called *frt* sites, that are arranged in head-to-head (inverted) orientation on the multicopy 2  $\mu$ m plasmid. If missegregation during cell division causes the plasmid copy number to drop, Flp mediated inversion of the intervening region of DNA between *frt* sites alters the direction of replication forks, resulting in the increased plasmid replication of the copy number amplification of the *S. cerevisiae* 2  $\mu$ m circle. Flp then resolves the resulting tandemly repeated products that have *frt* sites oriented in direct repeats to monomers. Cre acts primarily as a resolvase, while Flp can function both as a resolvase and invertase, but neither functions well as an integrase since intramolecular recombination is the natural reaction.

### **1.2.3 $\lambda$ -int and XerCD: Complex YRs**

$\lambda$ -int from bacteriophage lambda serves as the paradigm for the family (reviewed in (Azaro & Landy, 2002)). The recombinase is responsible for integration and excision of the phage  $\lambda$  DNA into and out of the infected cells' genome. During integration, recombination takes place between a site on the viral genome called *attP* and a site on the bacterial genome called *attB*. The *attB* site is a minimal recombination site similar to those found in simple recombinases, however *attP* is far more complex. In addition to containing a simple core recombination site, *attP* contains more than 200 bp of additional DNA that contains accessory sites for proteins that regulate the activity of  $\lambda$ -int. Integration of phage DNA between *attB* and *attP* generates hybrid sites, *attL* and *attR*, which are substrates for excision. Included on *attP* are additional  $\lambda$ -int binding modules called ARM sites (labeled as P and P' in Fig. 1.3).



**Figure 1.3 Site specific recombination by  $\lambda$ -integrase**

During infection  $\lambda$ -int mediates recombination between core sites, *attP* on the phage lambda genome and *E. coli* site *attB* generating hybrid sites, *attL* and *attR* to integrate into the host genome. Excision of the lambda genome requires recombination between *attL* and *attR* to regenerate *attP* and *attB*. The integration and excision reactions are highly regulated requiring proteins IHF in integration and excision, Xis in excision, which is further stimulated by Fis. ARM binding sites are labeled P and P', IHF binding sites H, Xis binding as X and Fis binding site F. Black symbols indicate sites that are used in integration (top) or excision (bottom).

Adapted from: (Van Duyne, 2008)

Unlike Cre and many other YRs,  $\lambda$ -int contains an additional domain found at the N-terminus that binds to the regulatory ARM sites found in *attP*. The central core binding (CB) domain and catalytic domain of  $\lambda$ -int are homologous to those in other YR homologs, which are responsible for binding the core DNA sites where recombination takes place. This makes  $\lambda$ -int heterobivalent in that it independently binds two distal regions of DNA using the same protein subunit. The binding of ARM DNA by  $\lambda$ -int is required for efficient synapsis and regulating recombination. The directionality of recombination of  $\lambda$ -int is highly regulated by a number of accessory DNA binding proteins that alter the topology of the synaptic complex, either activating or repressing integration or excision. Bacterially encoded integration host factor (IHF) binds *attP* at several sites and induces a sharp bend in the DNA to alter the topology of the site, possibly to allow  $\lambda$ -int to physically interact with both core site and P site DNA, simultaneously. The phage encodes the accessory protein Xis, which binds to sites on *attP* and alters the topology of the synaptic complex such that it promotes excision of the viral genome and inhibits integration. The bacterial host encoded protein Fis also binds to *attP* and stimulates excision of lambda from the host genome. All of these cofactors combine to make  $\lambda$ -int one of the more complex YR systems.

The *E.coli* XerC and XerD recombinases have homologs that are found in most eubacteria and have an important role in chromosome segregation at cell division (reviewed in (Barre & Sherratt, 2002)). The system is unique in that two separate recombinase proteins are used to complete a full recombination reaction. Of the commonly studied YRs, XerC and XerD are the most closely related with 37% sequence identity (Barre & Sherratt, 2002). These two recombinases recognize a 28 bp site on the bacterial genome called *dif*, which contains specific RBEs for each recombinase. During replication, homologous recombination generates chromosome dimers for approximately 15% of cell divisions (Barre & Sherratt, 2002). XerC and XerD interact with the protein FtsK, a membrane associated DNA translocase located at the septum during cell division, to resolve chromosome dimers to monomers at the *dif* recombination sites. The XerC/D system is also involved in resolving multimers of multicopy plasmids, pColE1 and pSC101, at two *dif*-like recombination sites called, *cer* and *psi*, respectively. Both of these

sites contain more than 100 bp of accessory DNA that include binding sites for host encoded arginine repressor (ArgR) and leucine aminopeptidase (PepA) at *cer* and aerobic respiration control protein (ArcA) and PepA at *psi*. These accessory proteins bind and bend DNA specifically, creating a 'topological filter' that ensures that only intermolecular recombination will occur.

### 1.3 Recombinases as tools for DNA manipulation

Simple YRs, such as Cre and Flp recombinases were found to have robust activity *in vitro* as well as in bacterial and eukaryotic systems (Dale & Ow, 1990; Sauer, 1987). The minimal requirements for activity of Cre combined with the high specificity for *loxP* allowed for development of Cre as a tool used in DNA manipulations and genome engineering (Kilby, Snaith & Murray, 1993).

In a cell-free system, purified Cre can simplify the introduction and manipulation of genes into and out of plasmids and large viral genomes (Gage, Sauer, Levine & Glorioso, 1992; Peakman, Harris & Gewert, 1992). The use of specific recombination sites eliminates steps involved in complicated subcloning, bypassing the need for restriction enzymes and appropriately placed restriction sites (Liu *et al.*, 1998; Liu, Li, Liu & Elledge, 2000).

In addition to applications *in vitro*, Cre and Flp are extraordinarily useful as tools *in vivo* (reviewed in (Sauer, 2002)). Cre is commonly used to perform recombination events such as gene knock-outs, knock-ins or inversions in eukaryotic systems such as mice. A gene or genes of interest can be bracketed by recombination sites. Depending on the orientation, expression of Cre will either excise or invert the intervening region containing either a gene, promoter or terminator. Since the expression of Cre can be controlled in a time and tissue specific manner, this offers a unique and effective way to study the effects of transgenes in ways not possible by other traditional methods.

Combining multiple recombinases, each with a unique recombination site or mutated derivatives of the same recombinase that recognize novel sites have both been used successfully to expand the number of genome manipulations that can be performed in a single experiment (Birling, Gofflot & Warot, 2009; Buchholz & Stewart, 2001; Livet *et al.*, 2007).

## 1.4 Mechanism of recombination

Decades of biochemistry in YRs and TopIBs (primarily vaccinia virus topoisomerase), have contributed greatly to understanding the mechanism of recombination. More recently, crystal structures of YRs and TopIBs have greatly advanced the understanding of tyrosine recombinases (reviewed in (Chen & Rice, 2003c) and (Van Duyne, 2008) and (Dong & Berger, 2008)). Among the most revealing structures are those that capture recombinase-DNA complexes as 'snapshots' of mechanistic steps in the pathway. In particular, multiple Cre-*loxP* structures have captured views of the pre-cleavage synaptic complex, the post-cleavage covalent intermediate bearing a 3'-phosphotyrosine linkage, and the Holiday Junction intermediate (Fig. 1.7). Structures at recombination reaction intermediates have also been captured for Flp and  $\lambda$ -int (Biswas *et al.*, 2005; Chen *et al.*, 2000) (Fig. 1.8). In addition to revealing the overall architecture, the structures also have made important contributions to understanding phosphoryl transfer chemistry both at the level of catalysis within the active site and with respect to the mechanism of strand exchange within the assembled recombination complexes.

### 1.4.1 Phosphoryl-transfer chemistry

The mechanism of cleavage and ligation with the formation of a 3'-phosphotyrosine intermediate is conserved within the YRs and shared with TopIBs. The reaction has been shown to be  $SN_2$ -like, proceeding through a pentacoordinate phosphorane transition state intermediate in which the tyrosine nucleophile and the O5' leaving group form axial ligands (Grindley *et al.*,

2006; Shuman, 1998). The reaction has been proposed to utilize an acid/base mechanism where in the cleavage reaction, a base is responsible for accepting a proton from the tyrosine, and an acid donates a proton to the O5' leaving group. In the ligation reaction the roles are reversed where the acid accepts a proton from the incoming 5' OH and the base donates a proton to the tyrosine leaving group (Fig. 1.4).

Structural and biochemical work in YRs and TopIBs has helped identify residues that aid in activation of the scissile phosphate and stabilize the transition state in the phosphoryl transfer reaction (Fig. 1.5). In addition to non-covalent and covalent intermediate structures that revealed the active sites of YRs and TopIBs, recent structures of TopIB captured a covalent complex with vanadate and DNA (Davies *et al.*, 2006; Perry *et al.*, 2010). Vanadate forms a trigonal bipyramidal geometry that mimics the transition state of the phosphodiester backbone as it is attacked by the catalytic tyrosine. These structures have aided the understanding and function of residues in the active site of TopIB, however no equivalent structure exists for YRs (Fig. 1.6).

The YR active site is composed of seven conserved residues (Table 1.1). In addition to the tyrosine nucleophile, one lysine and two arginine residues are thought to be essential for efficient recombination. Two histidine residues are also found in most YR active sites; one of these (His/Trp III) is tryptophan in the Cre and Flp recombinases. The seventh active site residue is either Glu or Asp in the YR enzymes. A comparison of ~100 tyrosine recombinases in 1998 revealed aspects of the protein sequences that were strongly conserved, that were similar, and that differed among family members (Nunes-Düby *et al.*, 1998). To obtain an updated estimate of the degree of conservation of the active site residues that includes data from microbial genomes sequenced over the past decade, we analyzed over 12,000 PsiBlast (Altschul *et al.*, 1997) hits from the NCBI RefSeq protein database (Pruitt, Tatusova & Maglott, 2007) using the Cre recombinase catalytic domain as a query sequence (method details described in Chapter 2). As shown in Table 1.1, conservation ranges from the more variable box I histidine (87%) to the highly conserved catalytic tyrosine (99.6%) and box II arginine (99.1%). Two residues (His/Trp III and

Glu/Asp I) are present as one of two alternative amino acid types in the YR enzymes. TopIBs have a similar active site, sharing five of the conserved residues with YRs. TopIBs lack the strongly conserved histidine and the acidic residue (glutamic acid or aspartic acid).

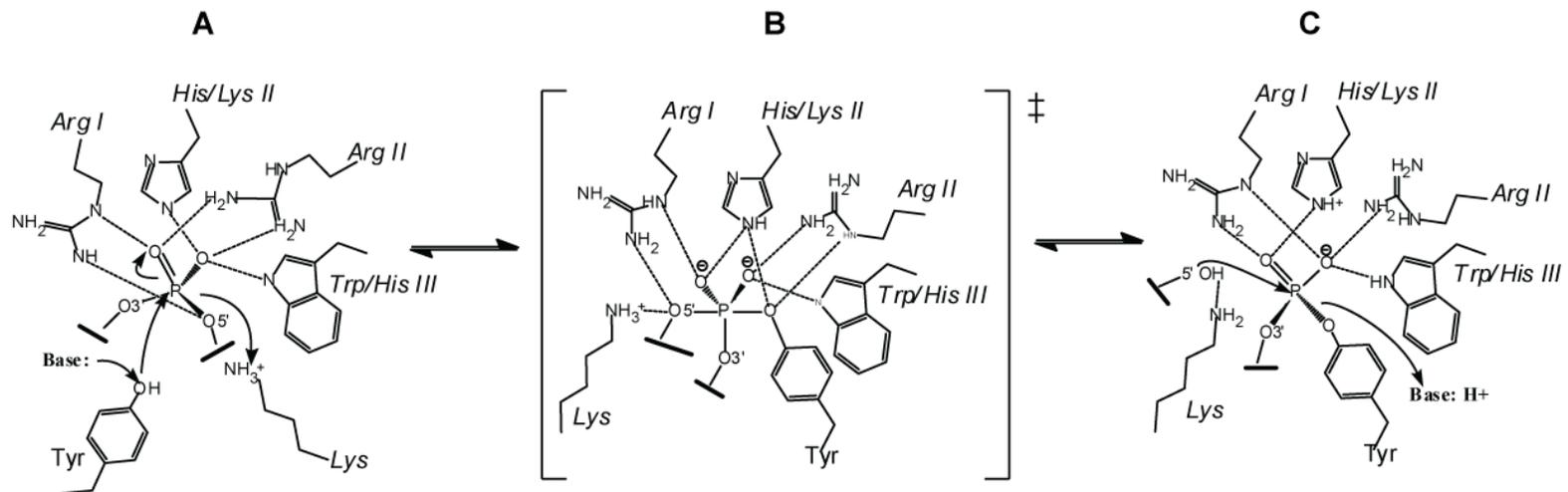
The properties of a number of YR (and TopIB) active site mutants have been described in the literature, including their ability to catalyze recombination (relaxation) *in vivo* or *in vitro* and to carry out specific catalytic activities such as binding, cleavage, and ligation. We have compiled a summary of these results in a table found in Appendix A, that augments a similar compilation reported by Nunes-Duby et al. (Nunes-Duby *et al.*, 1998).

Studies performed in vaccinia virus topoisomerase and Cre have demonstrated that the invariant lysine (K201 in Cre), acts as a general acid to protonate the leaving 5' hydroxyl during cleavage (Krogh & Shuman, 2000). Two arginine residues interact with the scissile phosphate, likely having a role in charge stabilization. Work with TopIB demonstrated that R130 (R173 in Cre) may assist in general acid catalysis, acting in a proton shuttle with the catalytic lysine, lowering the pKa of the lysine, or possibly participating as the sole proton donor to 5'OH (Krogh & Shuman, 2002; Nagarajan *et al.*, 2005; Stivers *et al.*, 2000).

Table 1.1: Tyrosine recombinase and topoisomerase IB active site residues

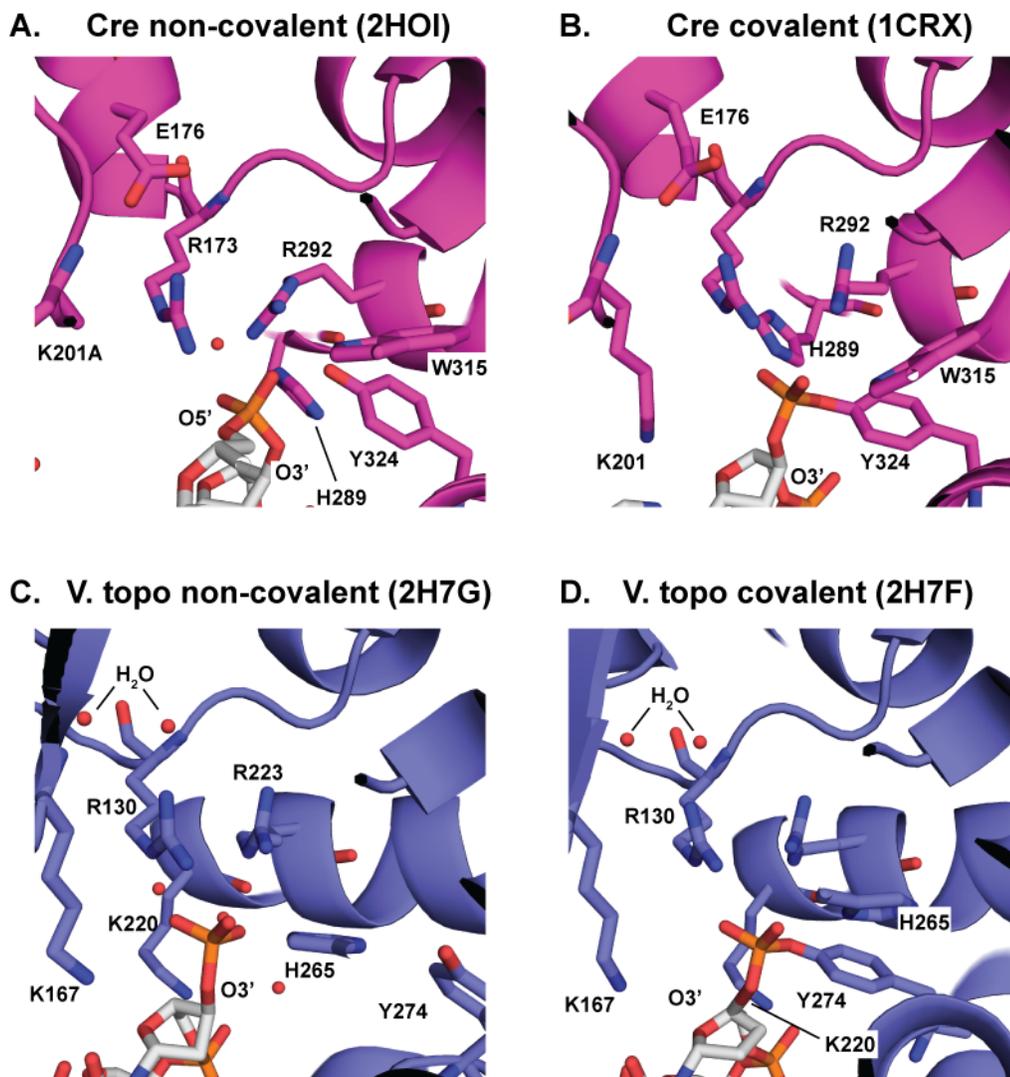
	Arg I	Glu/Asp I	Lys	His II	Arg II	His/Trp III	Tyr
P1 Cre	R173	E176	K201	H289	R292	W315	Y324
$\lambda$ Int	R212	D215	K235	H308	R311	H333	Y342
<i>S. cerevisiae</i> Flp	R191	D194	K223	H305	R308	W330	Y343
<i>E. coli</i> XerC/D	R148	D151	K172	H240/244	R243/247	H243/247	Y275/279
Poxvirus TopIB <sup>a</sup>	R130	None	K167	None <sup>b</sup>	R223	H265	Y274
Conservation <sup>c</sup>	91.2% R 1.1% K	79% E 16.4% D	94.7% K 3.8% R	87.3% H 6.8% Y	99.1% R 0.5% K	92% H 4% W	99.6% Y
Functions <sup>d</sup>	TS	S	GAB	TS, GAB	TS	S, TS	N

<sup>a</sup>Residue numbers refer to the vaccinia/variola virus topoisomerases. <sup>b</sup>A lysine residue is located in this position of the sequence in the TopIB enzymes, but is directed away from the active site. <sup>c</sup>The values given are based on analysis PSI-BLAST hits from the NCBI RefSeq database using the Cre catalytic domain as query. The percentages for the catalytic lysine are estimates, as described in Methods. <sup>d</sup>Roles of active site residues in catalysis, based on published biochemical and structural data and on the work described here. TS=transition state stabilization; S=structural; GAB=general acid/base catalysis; N=nucleophile.



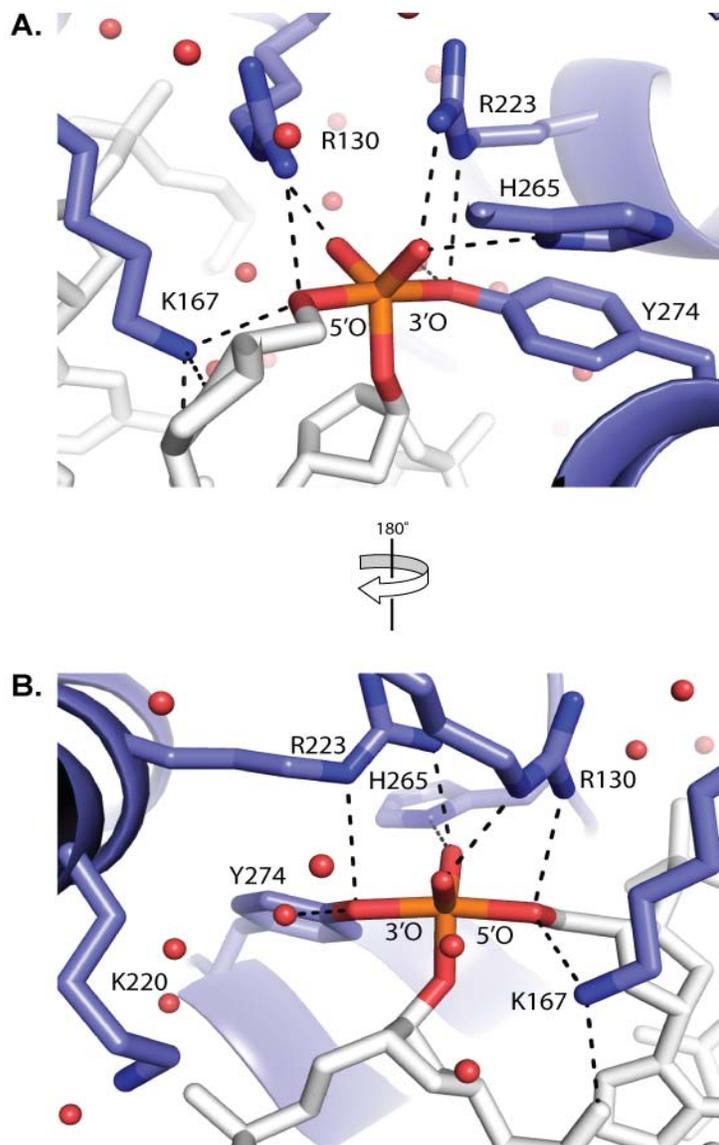
**Figure 1.4** *Proposed mechanism of phosphoryl transfer in YRs and TopIB*

Active site interactions depicting mechanistic states in phosphoryl transfer in YRs and TopIB enzymes as derived from structural snapshots and supported with biochemistry. (A) Precleavage. (B) transition state. (C) Covalent intermediate. A constellation of conserved active site residues surround the scissile phosphate. The reaction proceeds through an  $S_N2$  like mechanism where attack by the tyrosine nucleophile, assisted by a general base generates a pentavalent transition state as observed in TopIB structure (pdb 3IGC). The two strictly conserved arginines stabilize negative charge on the scissile phosphate in the transition state and the conserved Trp/His III donates a hydrogen bond to a nonbridging oxygen O1 of the phosphate. The conserved lysine acts as a general acid to donate a proton to O5' of the leaving strand, generating a 3'phosphotyrosine covalent intermediate (C). Ligation is the reverse of cleavage. The 5'OH of the original strand or in the case of YRs, from a partner site is activated by the lysine, now acting as a base, for attack on the 3'-phosphotyrosine generating the transition state where the a base, now acting as an acid donates a proton to the tyrosine leaving group generating intact DNA and noncovalently attached enzyme as in (A).



**Figure 1.5** *Pre-cleavage and covalent intermediate active sites of Cre and TopIB.*

Structural snapshots depicting the arrangement of active site residues of Cre (top, A and B) and V. topoisomerase (bottom, C and D). With the exception of E176, H289 the active sites are nearly superimposable. In place of E176, TopIB has two well ordered water molecules providing similar interactions as E176 with Arginine and Lysine. K220 in topoisomerase is the equivalent of H289, but is oriented away from the active site making an interaction if a phosphate. In the precleavage (C) and transition state mimic (not shown) structures of TopIB, two well ordered water molecules occupy the space as histidine. Similar well ordered water is not observed in the TopIB covalent intermediate structure (2H7F), possibly due to the 2.7 Å resolution.



**Figure 1.6** *Active site of Variola topoisomerase transition state intermediate*

The active site of Variola (smallpox) topoisomerase active site as viewed in the transition state where the scissile phosphate has been replaced with vanadium (pdb 3IGC). Notice the well ordered water in B, located 2.7 Å from phenolic oxygen of Y274 in the same approximate location as Nε of H289 in Cre in the covalent complex (1CRX – see figure 1.5).

The identification of a general base responsible for activation of the catalytic tyrosine has remained elusive in both families of enzymes. The conserved histidine in YRs, H289 in Cre, is a lysine in the TopIB enzymes. In structures of YRs, the histidine is positioned next to the catalytic tyrosine leading to the suggestion that the histidine could act as a general base to activate the tyrosine for attack on the scissile phosphate (Guo *et al.*, 1997). However, mutagenesis of this histidine is not as deleterious to recombination as mutations at the catalytic lysine (Ghosh, Lau, Gupta & Van Duyne, 2005a; Parsons, Prasad, Harshey & Jayaram, 1988; Whiteson *et al.*, 2007) (Appendix A). Surprisingly, the lysine of TopIB forms an interaction with the phosphate backbone of DNA away from the active site. The equivalent space is instead occupied by two well ordered water molecules that have been proposed to assist in the activation of tyrosine (Davies *et al.*, 2006). It remains an open question why YRs have a conserved histidine when TopIB enzymes have no equivalent residue.

Another notable difference in active sites between YRs and TopIBs is the conservation of an acidic residue, E176 in Cre with no structural equivalent in TopIB (Nunes-Düby *et al.*, 1998). In Cre, E176 accepts hydrogen bonds from the backbone amide nitrogens of K201 and R173 in addition to N $\eta$ 1 of R173 (Ghosh *et al.*, 2007; Ghosh *et al.*, 2005b; Guo *et al.*, 1997). As with the histidine, water fills this space in TopIB structures. Although aspartate in the equivalent as Cre E176 has been studied in Flp and  $\lambda$ -int (Chen *et al.*, 1992a; Nunes-Düby *et al.*, 1998), no specific function has been suggested that would explain why this residue is conserved in YRs. E176 has essentially been ignored and until now, has not been considered among the conserved active site residues in YRs (Azaro & Landy, 2002; Chen & Rice, 2003b).

The specific roles of all conserved active site residues are explored further in Chapter 2 of this thesis.

## 1.4.2 Architecture

Crystal structures of YRs have had a tremendous impact in the understanding the assembly of the activated tetramer and the process of strand exchange that defines the recombination pathway (reviewed in (Van Duyne, 2008; Van Duyne, 2002)). Structures are available for isolated domains, unliganded protomers and complete tetrameric protein-DNA complexes.

The tetrameric complex consisting of two sets of recombinase dimers bound to the core site are brought together in a process called synapsis. It is within this tetramer that cleavage, strand exchange and ligation take place, forming a HJ intermediate (Fig. 1.2A). Crystal structures of Cre have captured the pre-cleavage synaptic complex, the post-cleavage covalent intermediate and the HJ intermediate providing visual snapshots of several crucial steps in the recombination pathway (Gopaul, Guo & Van Duyne, 1998; Guo *et al.*, 1997; Guo, Gopaul & Van Duyne, 1999) (Fig. 1.7). The structures of a Flp HJ complex (Chen *et al.*, 2000) and  $\lambda$ -int HJ and synaptic intermediates (Biswas *et al.*, 2005) have also been determined, providing a means of directly comparing a number of YRs to understand conserved and variable features of the recombinase-DNA assemblies (Fig. 1.8).

Strikingly, the structures of Cre in the three states are almost superimposable, exhibiting a pseudo four-fold symmetry with the largest differences being in the connectivity of the DNA in the cross-over region. Each subunit makes extensive protein:protein interactions, burying nearly 3400 Å<sup>2</sup> of surface area per subunit. The N-terminal domains on one side of the DNA plane form a set of interactions separate from those of the C-terminal catalytic domains (Fig. 1.7 and 1.8). The very C-terminus of Cre forms a helix (hN) that exchanges in a cyclical manner with neighboring subunits in the tetramer. The similarity in the three states suggest the entire recombination pathway occurs with little change in the overall architecture.

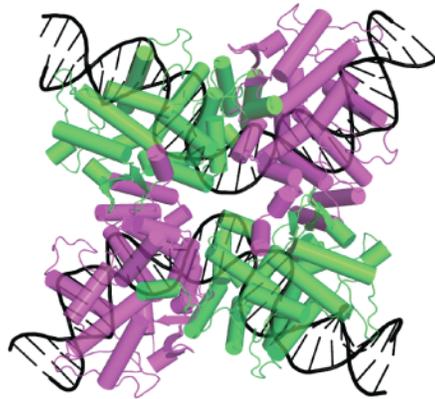
The inter-subunit interfaces observed in the Cre tetramers differ in the corresponding Flp and  $\lambda$ -int complexes. The Cre complex is the most rigid of the three, with  $\lambda$ -int forming very few interactions between the CB and catalytic domains (Biswas *et al.*, 2005; Chen & Rice, 2003c)(Fig. 1.8). This observation is consistent with the fact that  $\lambda$ -int exhibits weak binding specificity to core sites and requires binding of P site DNA to the ARM-domain for efficient synapsis. Interestingly, the ARM domains are cyclically exchanged in the integrase tetramer (Fig. 1.8).

The pseudo four-fold symmetry observed in the tetrameric complexes reveals two distinct interfaces. At the start of the reaction, synapsed sites form a pair of interfaces between two recombinases bound to the same core site (the dimer interface). A second interface is formed between recombinases bound to different core sites (a synaptic interface) (Van Duyne, 2002). HJ isomerization swaps these interfaces for the second half of cleavage and strand exchange. Differences in the protein:protein interfaces formed by the catalytic domains at dimer and synaptic interfaces are at least partially responsible for activating only two of the four recombinases for cleavage in the tetramer. The mechanism of activation in Cre is not completely understood, however regions containing the active site lysine and tyrosine in Cre are involved in protein:protein interfaces at the synaptic and dimer interfaces.

The catalytic lysine is located in a loop between two well conserved  $\beta$ -strands in YRs and TopIB enzymes. However, the size of this loop and placement of the lysine in the loop is variable. In Cre, this element appears to be quite mobile, as the lysine is poorly ordered or misoriented in all but the activated subunits. In the active subunit, this region forms inter-subunit interactions at the synaptic interface, while the region is displaced from the active sites of inactive subunits mediated by the dimer interface. Therefore, it seems possible that this region could act as a form of allosteric control to regulate the activity of the recombinase by delivery of lysine to the active site. Since this interaction is mediated through a synaptic partner, this would imply that synapsis may be required for cleavage.

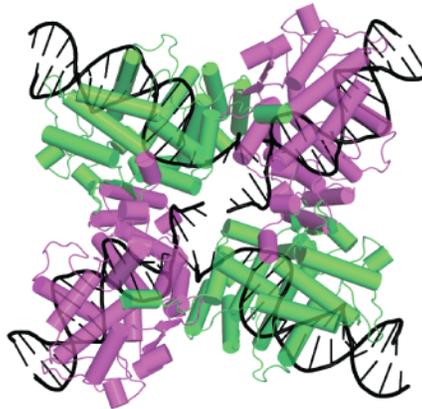
A. Synaptic complex

(pdb: 2H0I)



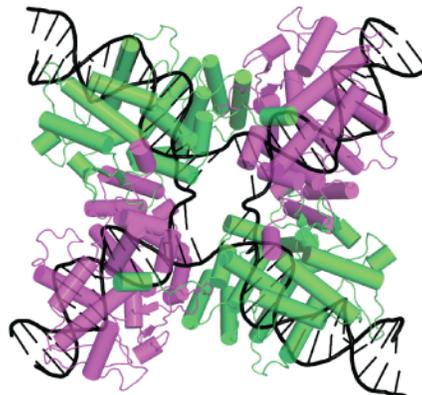
B. Covalent intermediate

(pdb: 1CRX)



C. Holiday Junction intermediate

(pdb: 3CRX)

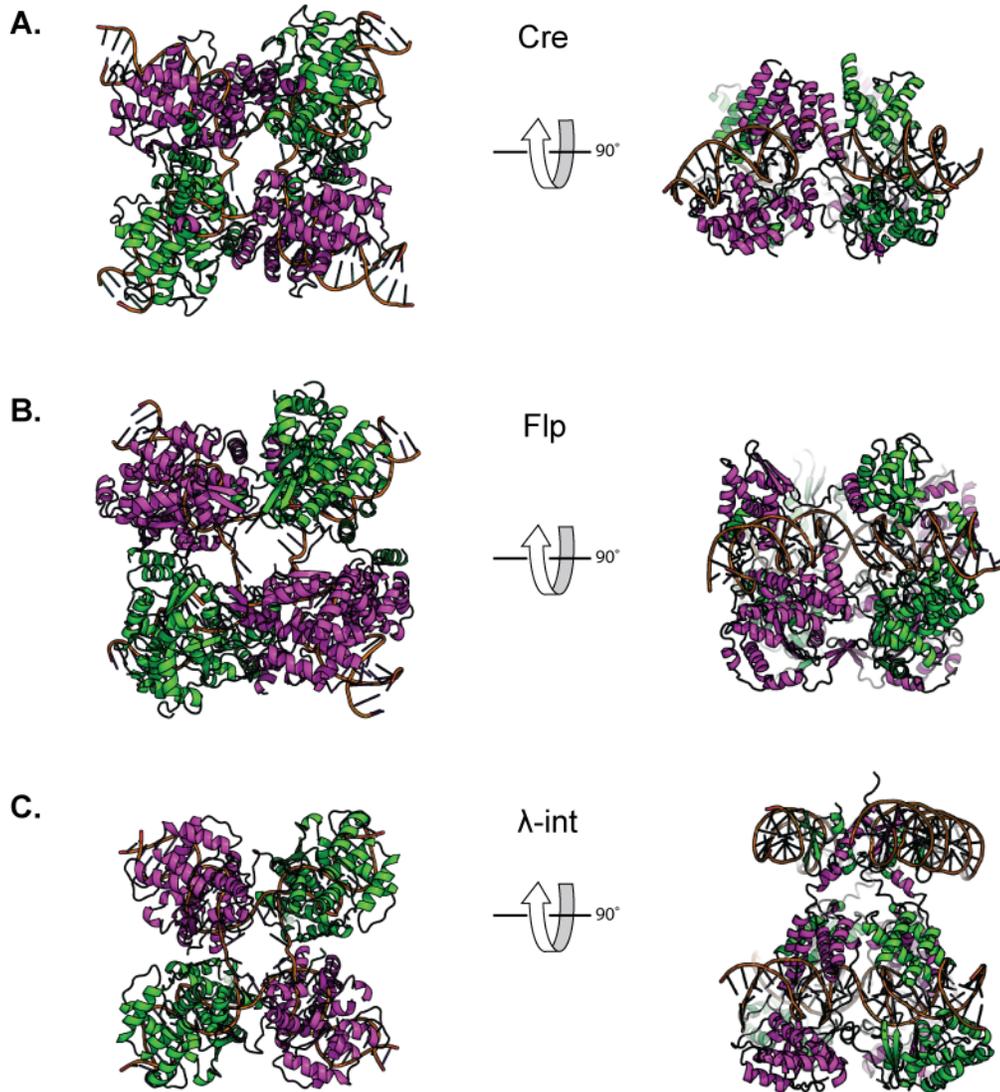


**Figure 1.7** Crystal structures of higher-order Cre-DNA complexes.

Cre subunits are colored green and purple. The DNA is yellow. The tyrosine nucleophile is colored red and represented as spheres. The architecture of the tetramer does not change significantly in the three states with exception of the central DNA region where cleavage and strand exchange occur.

(A) Pre-cleavage synaptic complex. (B) Covalent intermediate structure (C) A HJ intermediate structure.

The C-terminal region of Cre adopts a very different conformation between the two interfaces. In the dimer interface, hN of an active Cre subunit is donated to the catalytic domain of an inactive subunit. However, hN of inactive subunits is donated across the synaptic interface to the catalytic pocket of an active subunit. Although hN binding is similar in both instances, the position of hM, Y324, and the hM-hN linker, is different, suggesting a role for the C-terminus of Cre in regulating activity by modulating the position of Y324 (Gopaul *et al.*, 1998). This region is not conserved within the YR family. Flp for instance, exchanges the equivalent of hM with the catalytic tyrosine to a neighboring subunit and the equivalent of hN returns to the original subunit. Thus, the active site of Flp is composed of residues from two subunits (Chen & Rice, 2003b). Since the tyrosine is donated to a neighboring subunit, cleavage in Flp occurs in trans, as opposed to other YRs such as Cre,  $\lambda$ -int and XerCD that construct an active site entirely from the same subunit where cleavage occurs in cis. The C-terminal residues of  $\lambda$ -int are exchanged in a cyclical manner similar to Cre, however the element exchanged forms a  $\beta$ -strand and binds a different region of the catalytic domain (Biswas *et al.*, 2005). The only structural information from the XerCD system is of unliganded XerD, however biochemical studies investigating the C-terminus indicate that this region is important for activity and is likely to be structurally similar to Cre (Hallet, Arciszewska & Sherratt, 1999; Kazmierczak *et al.*, 2002; Spiers & Sherratt, 1999; Tekle *et al.*, 2002).



**Figure 1.8** *HJ intermediate structures of YRs*

Comparison of YRs structures captured in the same mechanistic state of HJ intermediate. Cre (3CRX) makes extensive protein:protein contacts at the interfaces of subunits, while Flp (1FLO) and  $\lambda$ -int (1Z1G) have far fewer interactions. The cyclic exchange of N-domains interacting with ARM site DNA are visible for  $\lambda$ -int (C, right).

## 1.5 Thesis objectives

The overall goal of this thesis is to understand the mechanism of Cre-*loxP* recombination. Despite the extensive biochemical and structural studies that have been reported in Cre, TopIB and other YR systems, a number of crucial questions remain.

***YRs contain seven active site residues. How conserved are these residues and how tolerant is Cre to change at each of these conserved positions?***

Although each of the conserved YR active site residues has been studied previously, no more than seven of the nineteen possible substitutions have been tested. In some cases, a clear understanding of the function has been derived from this limited set of mutants, however the function for others remains unclear. To understand why these seven residues are conserved in YRs, all nineteen single site variants were constructed in Cre and tested for the ability to support recombination. To further aid the analysis of the active site and requirements for catalysis, I determined the structure of Cre-DNA-vanadate complex, trapped as a transition state mimic, where the scissile phosphate has been replaced by pentavalent vanadium. This structure represents the first such intermediate for a YR system and complements a similar set of structures of TopIB (Davies *et al.*, 2006; Perry *et al.*, 2010) to aid in understanding the evolutionary relationship between YRs and TopIBs.

***Why do YRs contain two highly conserved active site residues that are not found in TopIB enzymes?***

YRs and TopIB share a similar mechanism of phosphoryl transfer and active site with two exceptions. H289 and E176 in Cre are well conserved residues in YRs with undefined function, but well positioned to serve important catalytic roles. The equivalent space in the TopIB active site

is replaced with water. I performed detailed biochemical analysis of mutants at H289 and E176 to better understand the catalytic function of these residues and suggest why they are not found in TopIB.

***Given similarities to TopIB enzymes that function as monomers, how does Cre, perhaps the most simple YR, regulate activity to prevent uncontrolled cleavage of DNA?***

TopIB enzymes efficiently cleave and ligate DNA as monomers, in some cases with limited DNA specificity. Cre and other YRs systems regulate their activities to prevent undesirable cleavage of DNA with potentially devastating consequences. Complex YR systems such as  $\lambda$ -int and XerC/D use accessory factors to regulate activity. However, the mechanisms controlling simple recombinases such as Cre are poorly understood. Work described in this thesis defines the minimally active Cre complex, and investigates the regulatory function of the C-terminal helix (hN).

## Chapter 2. Requirements for catalysis in the Cre active site (adapted from Gibb, B. *et al.* *Nucleic Acids Research*. 2010.)

### 2.1 Introduction

Members of the tyrosine recombinase (YR) family of site-specific recombinases catalyze DNA rearrangements using phosphoryl transfer chemistry that is identical to that used by the type IB topoisomerases (TopIBs). This similarity also extends to the level of structure, as the TopIB enzymes have similar catalytic domain folds to YRs and share a subset of the YR active site residues (Cheng, Kussie, Pavletich & Shuman, 1998).

The YR active site is composed of seven conserved residues (Table 2.1). In addition to the tyrosine nucleophile, one lysine and two arginine residues are thought to be essential for efficient recombination. Two histidine residues are also found in most YR active sites; one of these (His/Trp III) is a tryptophan in the Cre and Flp recombinases. The seventh active site residue is either Glu or Asp in the YR enzymes. A comparison of ~100 tyrosine recombinases in 1998 revealed aspects of the protein sequences that were strongly conserved, that were similar, and that differed among family members (Nunes-Düby *et al.*, 1998). To obtain an updated estimate of the degree of conservation of the active site residues that includes data from microbial genomes sequenced over the past decade, we analyzed over 12,000 PsiBlast (Altschul *et al.*, 1997) hits from the NCBI RefSeq protein database (Pruitt *et al.*, 2007) using the Cre recombinase catalytic domain as a query sequence. As shown in Table 2.1, conservation ranges from the more variable box I histidine (87%) to the highly conserved catalytic tyrosine (99.6%) and box II arginine (99.1%). Two residues (His/Trp III and Glu/Asp I) are present as one of two alternative amino acid types in the YR enzymes. Four of the seven active site residues listed in Table 2.1 are strictly conserved among the TopIBs.

Table 2.1: Tyrosine recombinase and topoisomerase IB active site residues

	<b>Arg I</b>	<b>Glu/Asp I</b>	<b>Lys</b>	<b>His II</b>	<b>Arg II</b>	<b>His/Trp III</b>	<b>Tyr</b>
P1 Cre	R173	E176	K201	H289	R292	W315	Y324
$\lambda$ Int	R212	D215	K235	H308	R311	H333	Y342
<i>S. cerevisiae</i> Flp	R191	D194	K223	H305	R308	W330	Y343
E.coli XerC/D	R148	D151	K172	H240/244	R243/247	H243/247	Y275/279
Poxvirus TopIB <sup>a</sup>	R130	None	K167	None <sup>b</sup>	R223	H265	Y274
Conservation <sup>c</sup>	91.2% R 1.1% K	79% E 16.4% D	94.7% K 3.8% R	87.3% H 6.8% Y	99.1% R 0.5% K	92% H 4% W	99.6% Y
Functions <sup>d</sup>	TS	S	GAB	TS, GAB	TS	S, TS	N

<sup>a</sup>Residue numbers refer to the vaccinia/variola virus topoisomerases. <sup>b</sup>A lysine residue is located in this position of the sequence in the TopIB enzymes, but is directed away from the active site. <sup>c</sup>The values given are based on analysis PSI-BLAST hits from the NCBI RefSeq database using the Cre catalytic domain as query. The percentages for the catalytic lysine are estimates, as described in Methods. <sup>d</sup>Roles of active site residues in catalysis, based on published biochemical and structural data and on the work described here. TS=transition state stabilization; S=structural; GAB=general acid/base catalysis; N=nucleophile.

A number of active site mutants have been previously described for YR and TopIB enzymes (summarized in Appendix I). Even when the data for all YR enzymes are combined, however, the substitutions that have been studied for a given active site residue number only 6-7 at most out of the 19 alternatives and in some cases relatively conservative substitutions have not been reported.

As is the case for the YRs (Van Duyne, 2008), there are now several high resolution crystal structures available for TopIBs bound to DNA substrates, providing a corresponding set of snapshots of the enzyme active sites at different steps in the phosphoryl transfer reaction (Dong & Berger, 2008). One such model is a vanadate transition state mimic for both trypanosomal and variola virus TopIB (Davies *et al.*, 2006; Perry *et al.*, 2010), which has provided additional insight into the nature of TopIB catalysis. A corresponding transition state mimic structure for a YR system has not yet been reported.

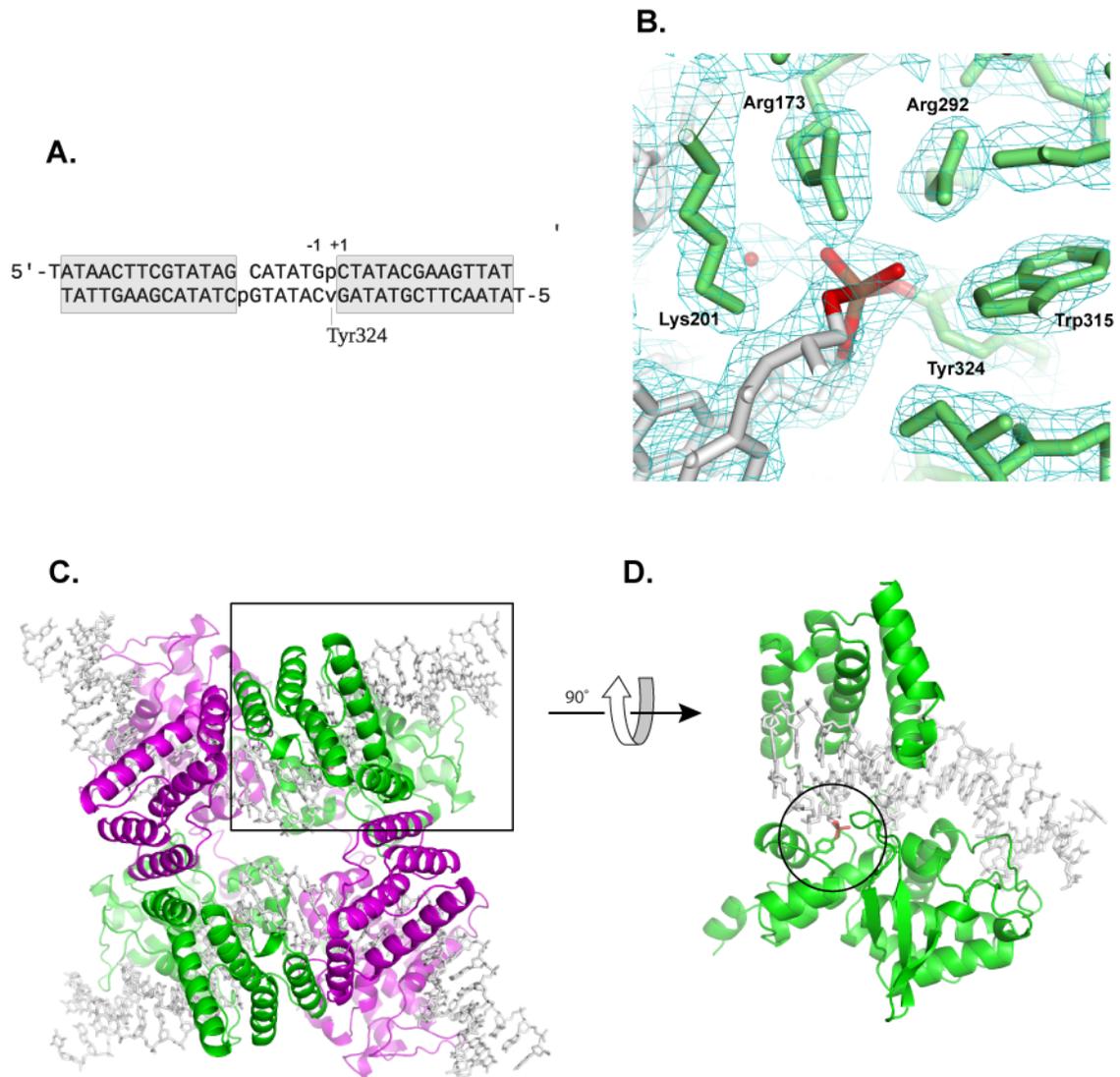
To better understand the requirements for YR catalysis and the relationship between the YRs and the TopIBs, I have analyzed the *in vivo* and *in vitro* recombination activities of all substitutions of the seven active site residues in Cre recombinase. I have also determined the structure of a Cre-*loxP* reaction intermediate trapped as a transition state mimic with vanadate that facilitates interpretation of mutant activities and allows for a comparison with similar structures from the related topoisomerases. The findings indicate that active site residues shared by YRs and TopIBs are most sensitive to substitution. Only two residues, the tyrosine nucleophile and a conserved lysine residue that activates the 5'-hydroxyl group leaving group, are strictly required to achieve greater than 5% of wild-type activity. The two conserved arginine residues each tolerate one substitution that results in modest recombination activity and the remaining three active site positions can be substituted with several alternative amino acids while retaining a significant amount of activity. The results in this chapter are discussed in the context of YR and TopIB structural models and data from related YR systems.

## 2.2 Structure of a Cre-*loxP* transition state mimic

To trap a vanadate transition state mimic for the Cre-*loxP* reaction, we used a symmetric version of the *loxP* site where the scissile phosphates are missing in both the top and bottom strand cleavage positions (Fig. 2.1A). I chose to symmetrize the more G-C rich half-site of *loxP* for these experiments (resulting in a site which I refer to as loxS1 (Ghosh *et al.*, 2007)), which provides a transition state model depicting the DNA functional groups that would be present in the initial cleavage step of the reaction (Hoess, Wierzbicki & Abremski, 1987), as opposed to the HJ-resolution step later in the pathway. When activated potassium orthovanadate was added to the Cre-loxS1 complex, one VO<sub>2</sub><sup>+</sup> equivalent was incorporated into the active site and covalent linkages were formed with O3', O5', and Tyr324 as observed in the related TopIB transition state mimic structures (Davies *et al.*, 2006; Perry *et al.*, 2010).

Crystals of the Cre-loxS1-vanadate complex were grown from hanging drops using conditions similar to those used previously for several Cre-DNA structures (Gopaul *et al.*, 1998; Guo *et al.*, 1997; Guo *et al.*, 1999) and diffracted to 2.3 Å resolution at synchrotron sources. The structure was determined by molecular replacement using the synaptic Cre-DNA complex (Ghosh *et al.*, 2007) lacking active site residues as a search model. A summary of diffraction data and refinement results is given in Table 2.1 and electron density within the active site following refinement is shown in Fig. 2.1B.

The overall structure is similar to that observed in the pre-cleavage and covalent intermediate Cre-DNA structures previously described (Ghosh *et al.*, 2007; Guo *et al.*, 1997), with a tetramer of interlocked Cre subunits bound to two bent DNA substrates (Fig. 2.1C-D). With the exception of the active sites, the protein subunits in the current structure superimpose with an r.m.s.d of 2 Å to those in the synaptic complex and covalent intermediate structures, with no significant differences observed. My discussion of the transition state mimic structure will therefore focus only on the active site of the enzyme.



**Figure 2.1** *Structure of Cre-DNA-vanadate transition state mimic.*

(a) Sequence of the DNA substrate used, with vanadium incorporation indicated by 'v'. The 14 bp recombinase binding elements that surround the central 6 bp crossover region are shaded. The '+1' and '-1' base-pair labels are used to facilitate comparison with related TopIB structures. (b) Weighted 2Fo-Fc electron density following refinement at 2.3 Å resolution. The map is contoured at 1.2  $\sigma$ . Glu176 and His289 are not visible in this view. (c) Overall structure of the synaptic complex, viewed from the N-terminal domains of the recombinase subunits. The activated subunits containing vanadate linkages are colored green. (d) View of a single Cre subunit bound to a *loxP* half-site within the tetrameric complex, rotated from the boxed subunit in (c). The tyrosine linkage to vanadate is circled.

Table 2.2: Crystallographic data for the Cre-DNA-vanadate transition state mimic

Diffraction Data	Cre-Vanadate-DNA Complex
Resolution (Å)	2.3 Å
Space Group	P322 <sub>1</sub>
Unit Cell	a = b = 136.046 Å c = 218.431 Å
Mosaicity	0.53
Wavelength (Å)	1
Completeness (%)	97.2(95.4)
R <sub>merge</sub>	0.103(0.462)
Total Reflections	775,986
Unique Reflections	102,079
I/σ	16.17 (4.15)
Redundancy	7.6(7.3)
<b>Refinement</b>	
R <sub>free</sub>	0.222(0.285)
R <sub>work</sub>	0.193(0.238)
Number of atoms	13,407
Protein	10,204
DNA	2859
Water	346
Average B factors (Å <sup>2</sup> )	
Protein	27.75
DNA	36.75
Water	25.6
Rmsd	
Bond lengths (Å)	0.01
Bond angles (°)	1.62

Values in parentheses refers to the highest resolution shell (2.38-2.30 Å).

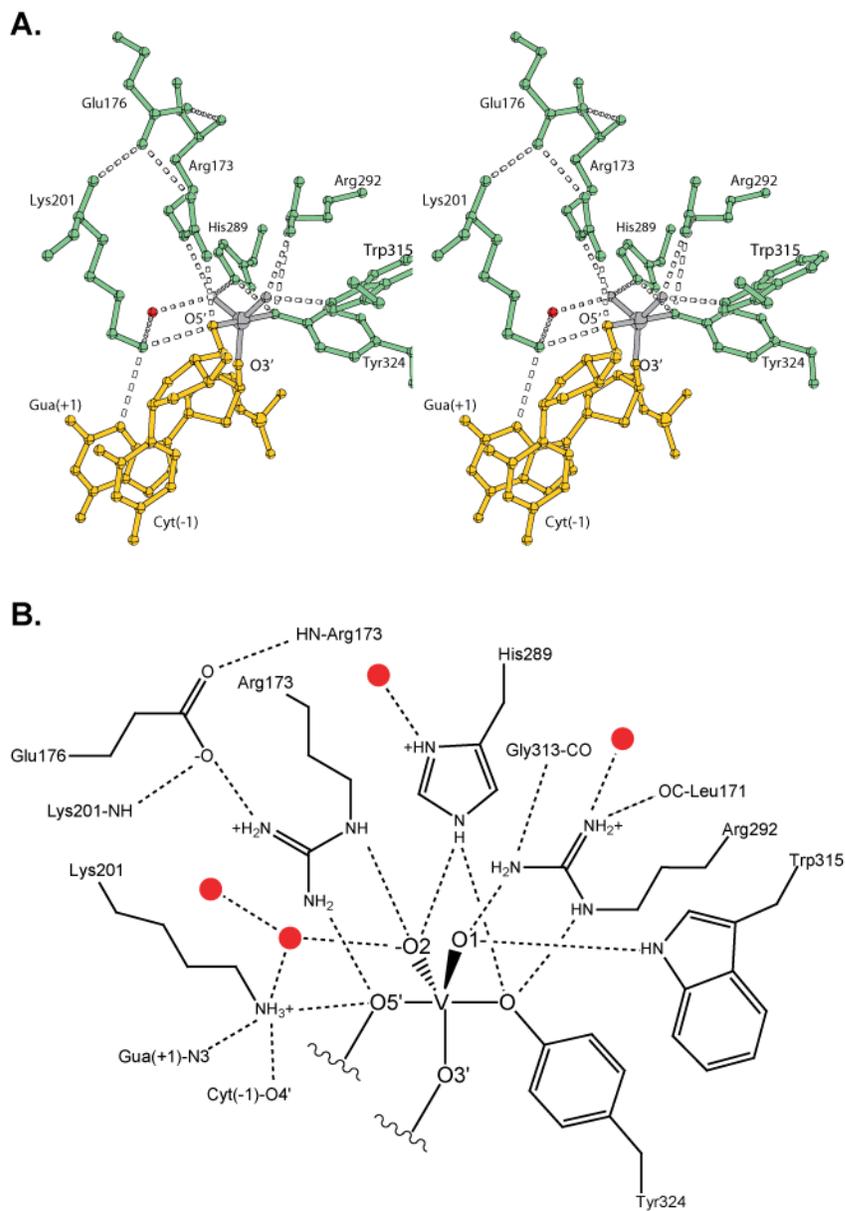
$R_{sym} = \sum |I_h - \langle I_h \rangle| / \sum I_h$ , where  $\langle I_h \rangle$  is the average intensity over the symmetry equivalents.

R<sub>work</sub> includes 95% of the reflection data used in refinement.

R<sub>free</sub> includes 5% of the reflection data excluded from refinement.

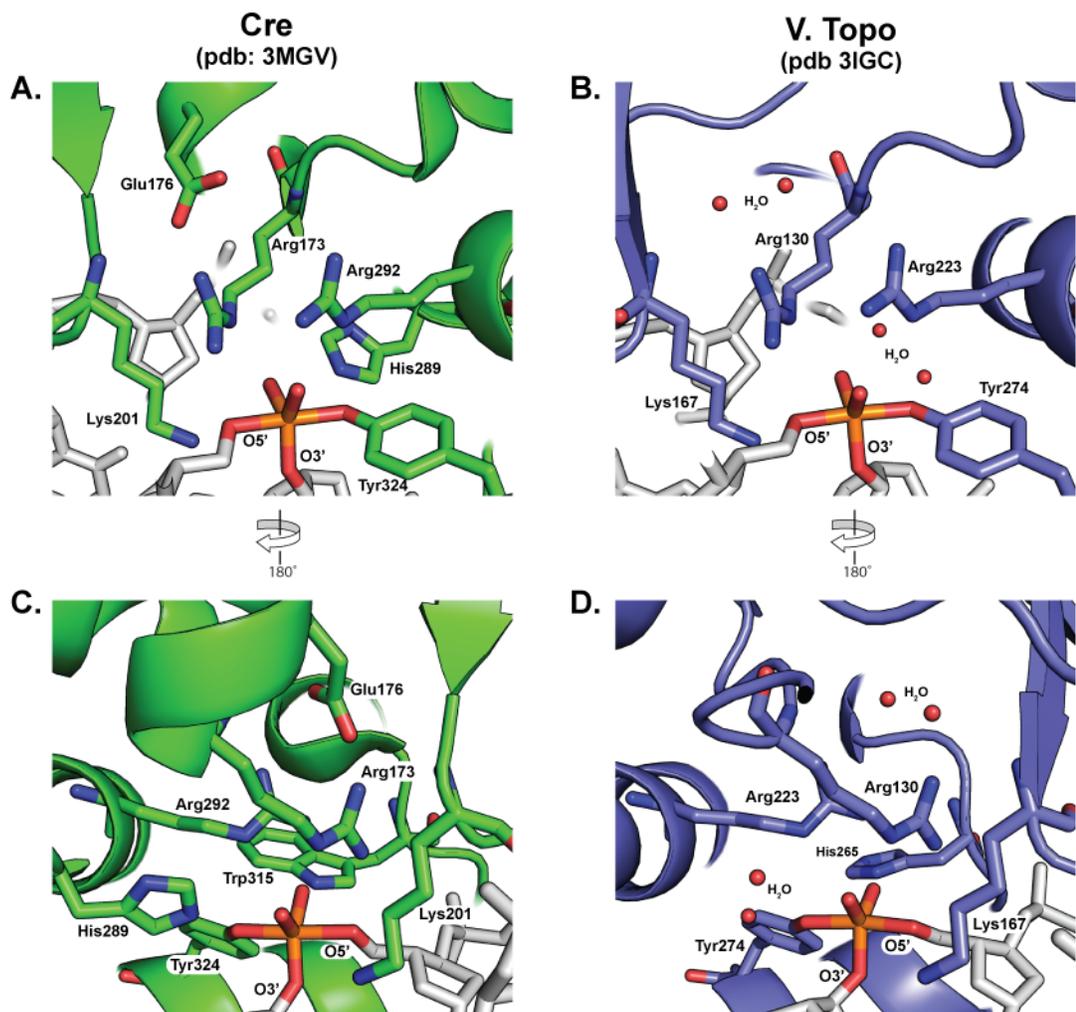
In principle, vanadate could be incorporated into each of the four active sites present in the Cre-loxS1 complex shown in Fig. 2.1C. In the Cre-DNA reaction intermediate structures that have been reported, two of the active sites in the Cre tetramer adopt 'active' configurations that are poised for catalysis of strand exchange (referred to here as the 'activated sites') and the other two are in 'inactive' configurations ('inactive sites') that are not competent to catalyze cleavage of the DNA (Van Duyne, 2001). In the current structure, vanadium is incorporated only into the activated sites, reflecting the inability of the remaining two subunits to stabilize a transition state configuration. With the exception of Tyr324 and Lys201, the protein and DNA residues involved in coordinating vanadium in the activated sites (Fig. 2.2) adopt similar positions in the inactive sites of the Cre-loxS1 complex. Tyr324 is shifted by 3.5 Å in the inactive site relative to the activated site position and Lys201 is withdrawn from the minor groove. These alternative configurations for Tyr324 and Lys201 are typical of those observed in the inactive sites of other Cre reaction intermediates and I assume that these effects are also responsible for preventing stable incorporation of penta-coordinate vanadium in the inactive sites of the Cre-DNA complex described here.

In the activated sites, vanadium is coordinated by five oxygen atoms in a trigonal bipyramidal arrangement with Tyr324-O $\eta$  and O5' in the axial positions and with O3' and the two non-bridging oxygen atoms in equatorial positions (Fig. 2.2). The two conserved arginine residues each make double hydrogen bonding interactions; Arg173 hydrogen bonds to O5' and the non-bridging oxygen corresponding to O2P and Arg292 hydrogen bonds to the Tyr324 phenolic oxygen and to O1P. Lys201 is within close hydrogen bonding distance of four groups: O5' of the -1 sugar, N3 of the +1 guanine base, O4' of the -1 sugar, and a tightly bound water molecule. Each of these features is also found in the trypanosomal and viral TopIB transition state structures (Davies *et al.*, 2006; Perry *et al.*, 2010), supporting the close mechanistic and evolutionary relationship between the two enzyme families. Several other aspects of the Cre active site differ from the TopIB enzymes and will be discussed below in the context of the specific residues. A comparison of the Cre and TopIB active sites trapped in the transition state is shown in Fig. 2.3.



**Figure 2.2** *The Cre-DNA-vanadate transition state mimic active site.*

(a) Stereo view of the 'activated' site where vanadate has been incorporated. (b) Schematic of the activated site with additional residues and hydrogen bonding interactions indicated. Well-ordered water molecules are drawn as red spheres in (a) and (b).



**Figure 2.3** *Cre and TopIB active sites in the transition state*

The active sites of Cre and TopIB vanadate complexes are compared. A and C. Cre pdb 3MGV (this work). B and D. Variola topoisomerase, pdb 3IGC (Perry, Hwang, Bushman & Van Duyne, 2010).

## 2.3 Comprehensive analysis of Cre active site substitutions

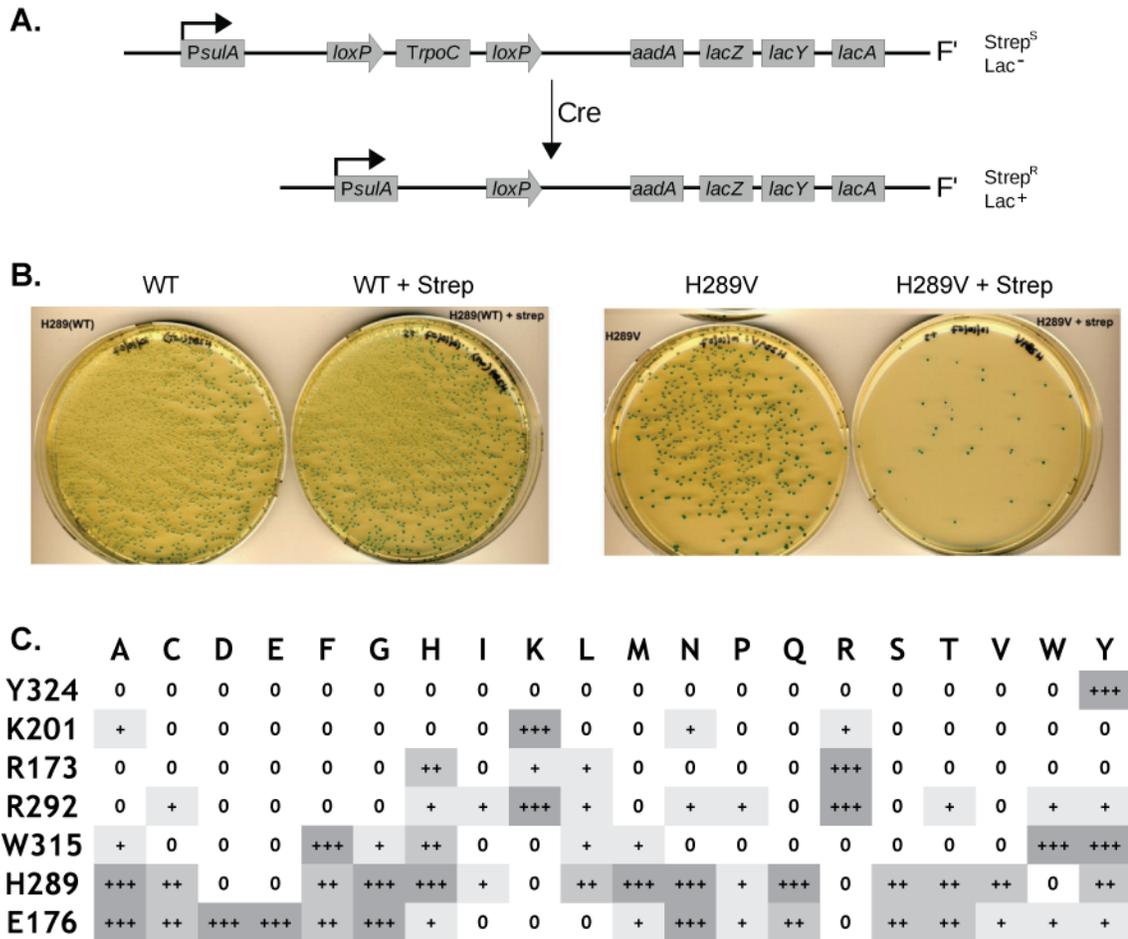
To establish which Cre active site substitutions would support recombination, I generated all single site variants for the residues shown in Table 2.1 and measured their ability to catalyze recombination *in vivo* and *in vitro*.

To test for recombination activity in bacterial cells, I designed the single copy reporter shown schematically in Fig. 2.4A. The *aadA* streptomycin resistance and *lacZYA* genes located on an F' episome are transcribed from a strong promoter when an intervening transcriptional terminator has been excised by Cre-*loxP* recombination, but full-length transcripts are only produced at low levels when the terminator is present. Cells that are Lac<sup>-</sup> and harbor the reporter F' will be sensitive to streptomycin and Lac<sup>-</sup> in the absence of active Cre but will become streptomycin resistant and Lac<sup>+</sup> when active Cre is present. For the experiments described here, I used streptomycin resistance and  $\beta$ -galactosidase activity to score for Cre-mediated site-specific recombination (Fig. 2.4B). Growth on minimal lactose plates can also be used and is more effective in identifying mutants with very low activity.

To assay for *in vivo* Cre activity, I transformed *E. coli* strain CSH142 containing the reporter F' with a plasmid containing a Cre coding sequence (described below) and compared the number of colonies that formed on plates containing 20  $\mu$ g/ml streptomycin to the number that formed on plates lacking streptomycin. Approximately 45 min is required for expression of the *aadA* gene (data not shown). A sixty minute outgrowth prior to plating therefore provided a ~15 min time window for recombination to occur. To facilitate *in vitro* mutagenesis and limit *in vivo* expression, I cloned the wild-type Cre coding sequence into pBluescriptIIKS+ where it is oriented opposite that of the *lac* and *bla* promoters. Due to the high plasmid copy number, this promoter-less construct still provides moderate levels of constitutively expressed Cre. When I compared Cre expression from this plasmid to basal Cre expression from uninduced pET21a (T7/*lac* promoter) and to a P1 lysogen by western blotting, I were not able to detect Cre in lysates from the P1 lysogen but could

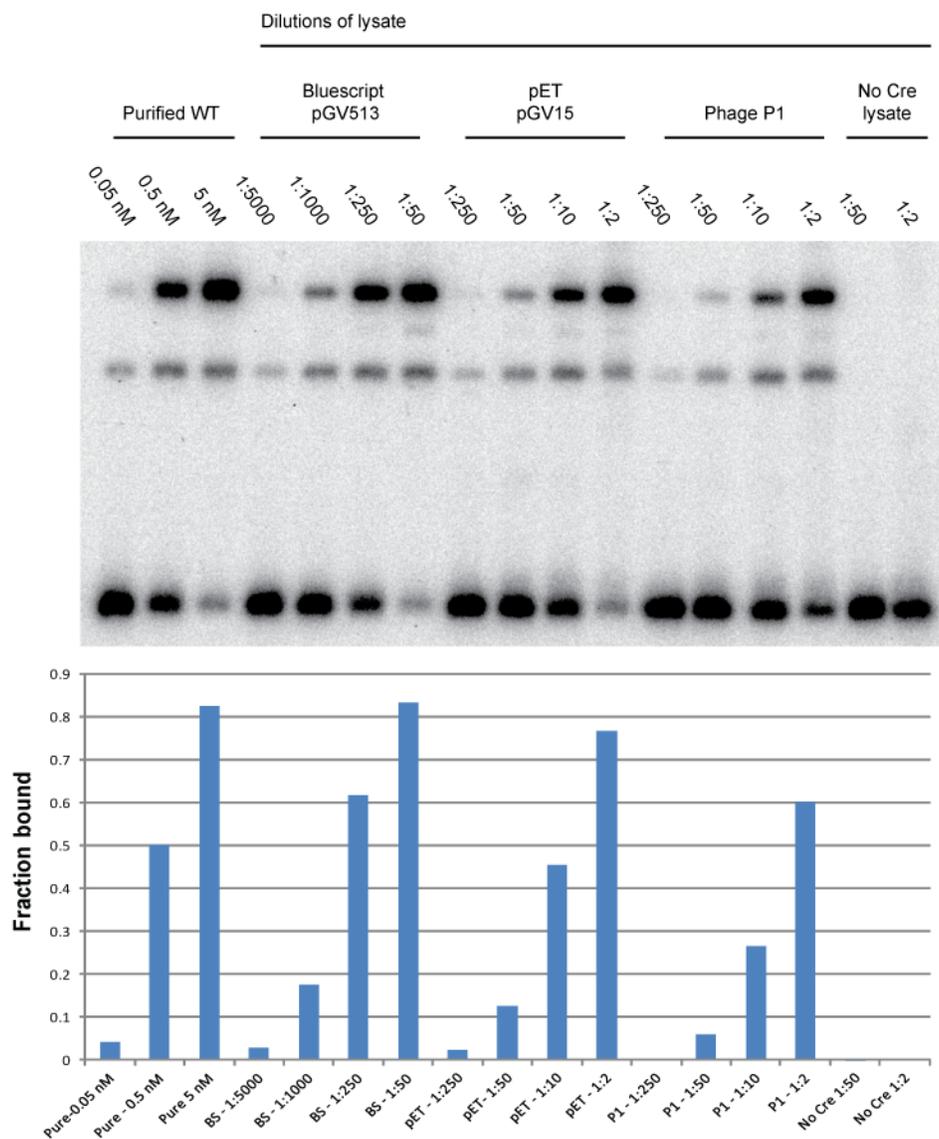
detect small amounts from pET21a and larger amounts from pBluescriptIIKS+ . Electrophoretic mobility shift assays (EMSA) confirmed that active Cre is present in all three cases and allowed us to estimate the relative concentrations (Fig. 2.5). Thus, the plasmid used for Cre expression provides protein levels that are higher than that found in a P1 lysogen, but much lower than would be produced by a promoter-driven expression system.

Using the pBluescriptII-Cre plasmid as template, I generated all nineteen amino acid substitutions at each of the seven active site residues shown in Table 2.2. Four of these (Arg173, Lys201, Arg292, and Tyr324) are strongly conserved among both the YRs and the TopIB enzymes. Trp315 is most often a histidine in the YRs and is always histidine for the TopIBs. The final two residues are Glu176 and His289, both of which are unique to the YR enzymes. Glutamic acid is favored in the Glu176 position, although a significant number of family members use aspartic acid. His289 is moderately conserved among the YRs but is absent from the TopIB proteins. The results of *in vivo* recombination for the 134 Cre variants are summarized in Fig. 2.4C. I found a surprising range of flexibility at several active site positions, including a number of mutants that I had not predicted would be active.



**Figure 2.4** *In vivo recombination activities of Cre active site mutants.*

(a) F' reporter used to score *in vivo* activity. If Cre excises a transcriptional terminator located downstream of the promoter, the streptomycin resistance and LacZYA gene products are expressed. (b) example data from *in vivo* recombination. After a 60 minute outgrowth, cells are plated on non-selective plates and selective plates containing 20ug/ml streptomycin (+strep). (c) *in vivo* recombination results for Cre active site substitutions. Activity definitions: 0: no activity observed; +: <5% wild-type, ++: 5-50% wild-type, +++: >50% wild-type.



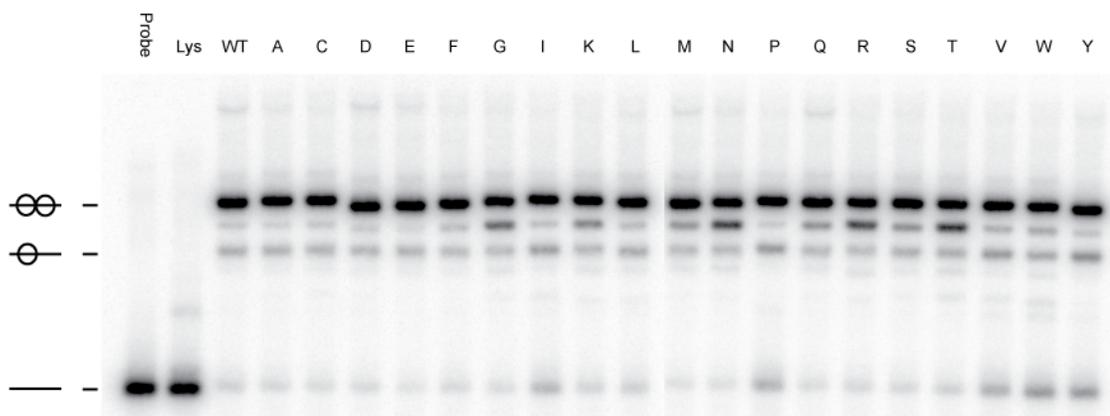
**Figure 2.5** *Cre concentration in vivo*

Dilution of Cre lysate and purified Cre were tested for binding *in vitro* to compare the expression levels of WT Cre on from the highcopy Bluescript vector used for the characterization of Cre mutants compared with the expression of uninduced Cre from pET21A and Cre expression in a P1 lysogen. Comparing the intensity of the shifted probe to that of purified Cre, allows for a determination of the concentration of Cre in the undiluted lysate. Based on the comparison to purified Cre, the approximate concentrations of Cre in the undiluted lysates are ~250 nM from the Bluescript vector, ~5 nM from pET21A, and >1 nM Cre from phage P1.

## 2.4 *In vitro* recombination of Cre mutants

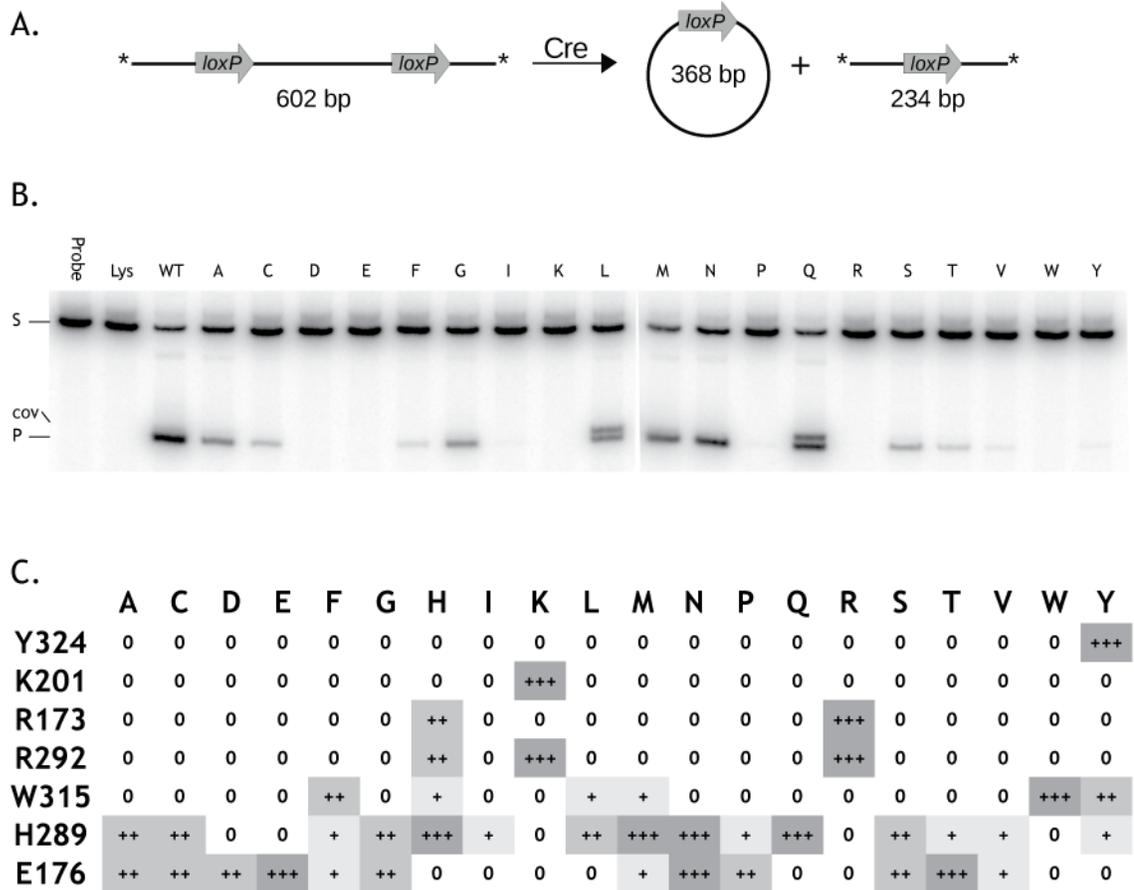
To determine if any of the active site substitutions disrupt the ability of Cre to bind to the *loxP* site, I used a gel-shift assay to test cleared lysates from the transformed reporter strain (Fig. 2.6). The lysates in these experiments were diluted in binding buffer so that wild-type Cre lysate had similar binding activity as that of a 4 nM Cre control. This allowed a qualitative comparison of DNA-binding activities in the lysates relative to wild-type Cre. As shown in Fig. 2.6, the lysates for all substitutions at His289 contain DNA-binding activities similar to wild-type Cre. Out of all 134 Cre variants tested, only the Y324T mutant displayed a significant *loxP*-binding defect, where the apparent  $K_D$  is more than 10-fold greater than wild-type Cre. For the remaining mutants, these results indicate that the Cre variants are being produced in a soluble form that bind to *loxP* site DNA, eliminating expression, folding, or gross DNA-binding defects as explanations for a lack of recombination activity. For many of the mutants, I noticed a small, but reproducible variation in mobility of the shifted EMSA bands that may be caused by subtle differences in DNA-bending.

To confirm and extend our findings from the *in vivo* recombination experiments, I tested each Cre mutant for *in vitro* recombination, again using lysates from the transfected reporter strain. In this assay, Cre excises a 334 bp segment flanked by *loxP* sites in a labeled, linear substrate, producing an unlabeled 368 bp circle and a labeled 234 bp linear product (Fig. 2.7A). The results of *in vitro* recombination for the H289 panel are shown in Fig. 2.7B and a summary of results for all of the active site substitutions is shown in Fig. 2.7C. In general, there is good agreement between *in vitro* and *in vivo* recombination, but for mutants with higher levels of activity the *in vitro* assay provides better discrimination. For example, the W315F and W315Y substitutions have near wild-type activity in the *in vivo* assay, but are shown have less than 50% wild-type activity in the *in vitro* excision assay. The *in vitro* recombination experiments also identified active site mutants that accumulate covalent and/or HJ intermediates in excess of wild-type Cre. These results are summarized in Table 2.3 and in Fig. 2.8.



**Figure 2.6** *In vitro loxP-binding for Cre H289 substitutions.*

A 64 bp *loxP*-containing probe for electrophoretic mobility shift assays using lysates that had been diluted 50-fold to give an apparent concentration of 5 nM for wild-type Cre based on gel shift assays of further serial dilutions and the known binding affinity to the *loxP* site. Assignment of the shifted bands was based on preliminary experiments where purified Cre was analyzed in the same gels for comparison. The additional band between the singly and doubly bound species is lysate-dependent and could be due to Cre degradation products (observed in Western blots) or to bacterial components in the lysates.



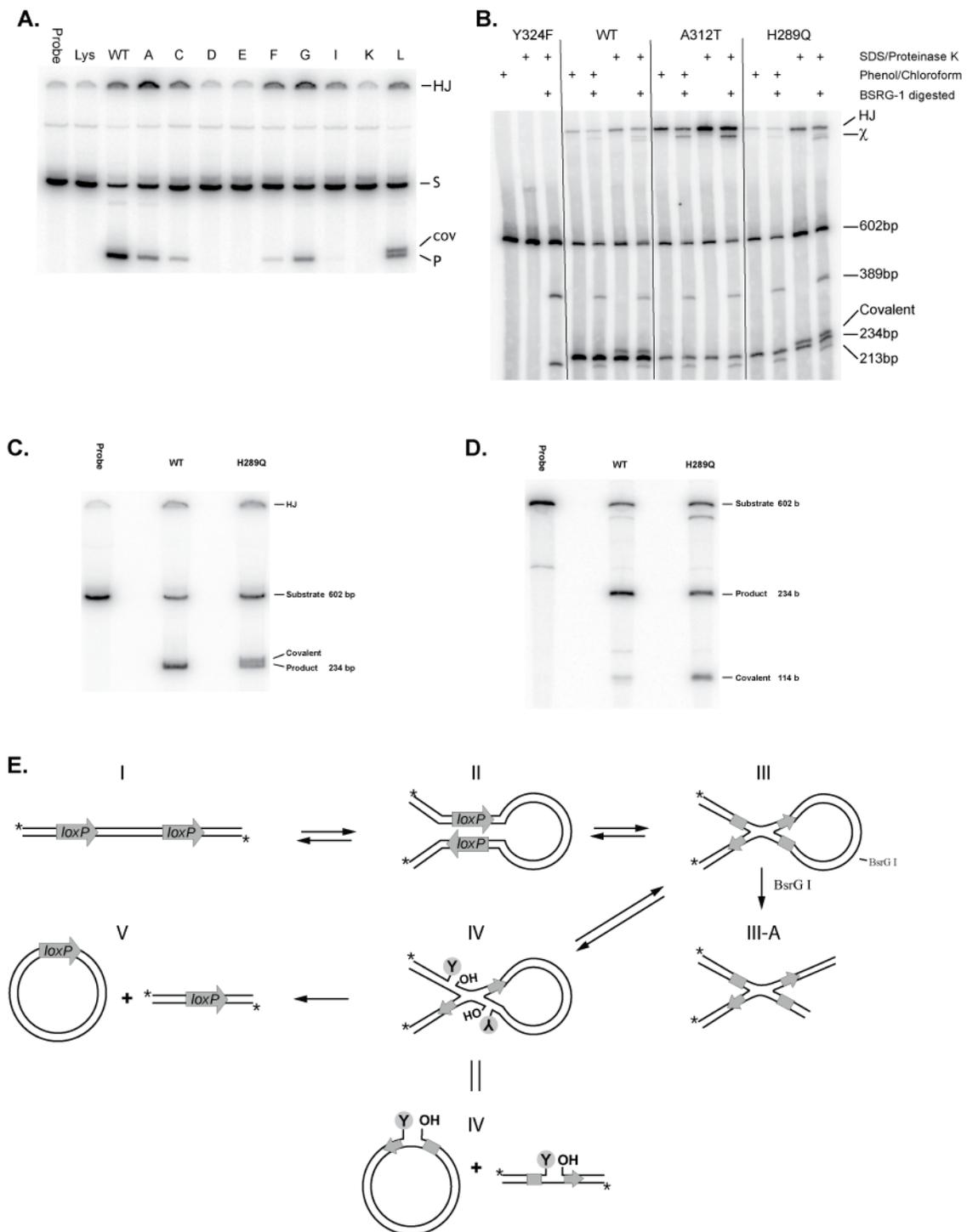
**Figure 2.7** *In vitro* recombination activities of Cre active site mutants.

(a) Schematic of the excision assay. (b) Native PAGE of *in vitro* recombination reactions following protein digestion for the H289 panel of substitutions. In addition to the 602 bp substrate (labeled 'S') and the 234 bp product (labeled 'P'), a band corresponding to a covalent Cre-DNA intermediate can be observed migrating slightly above the product band for some mutants. This is a nicked 234 bp covalent 3'-phosphotyrosine adduct. The corresponding covalent intermediate associated with the substrate is not well-resolved. Identification of covalent and HJ intermediate bands are described in more detail in Figure S2. (c) *in vitro* recombination results. Activity definitions: 0: no activity observed; +: <5% wild-type, ++: 5-50% wild-type, +++: >50% wild-type.

Table 2.3: Accumulation of Holliday junction and covalent intermediates during *in vitro* recombination

		A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y		
Y324	HJ																					+	
	COV																					+	
K201	HJ										+												
	COV										+												
R173	HJ								++											+			
	COV								++											+			
R292	HJ								++	+											+		
	COV								+++													+	
W315	HJ					+++																+	+++
	COV																					+	+
H289	HJ	++	+				+	++	+	+	+	++	+	+	+	+	+	+	+				
	COV								+				++	+	+	++							
E176	HJ	+++	+++	+++	+	+++	+++	+				++	++	++	+	+	+++	++			+		
	COV	+				+											+					+	

Key: + indicates levels similar to wild-type Cre; ++ indicates greater than 2x wild-type; +++ indicates greater than 4x wild-type. Levels reported are after a 60 min reaction.



**Figure 2.8** Identification of intermediates formed during *in vitro* recombination.

Legend on next page.

**Figure 2.8 continued. Legend**

(a) Expanded view of the 6% acrylamide gel in Fig. 2.5b. HJ intermediates do not migrate far into the gel matrix, but nicked 3'-phosphotyrosine linked products resulting from digestion of covalent Cre-DNA intermediates are resolved from product bands. (b) *in vitro* recombination reactions with Cre WT, Y324F, A312T and H289Q were incubated for 60 minutes, quenched, and then separated on a 4% acrylamide gel. Cre A312T accumulates HJ intermediates (1). The reactions were quenched with either SDS/Proteinase K or Phenol/Chloroform. Covalent intermediates partition into the organic phase during extraction with phenol/chloroform. The DNA substrate contains a central BsrGI site; digestion results in the formation of 389 bp and 234 bp bands. Digestion of the HJ  $\alpha$ -intermediate (III in panel e.) yields a  $\chi$ -structure with increased mobility (III-A in panel e.). Covalent intermediates formed during HJ resolution are identified by comparing samples quenched with SDS/Proteinase K vs. Phenol/Chloroform. (c and d) *In vitro* recombination reactions with Cre WT and H289Q were quenched after 60 min with SDS/Proteinase K and separated by native PAGE (c) or by denaturing PAGE in 7 M Urea (d). The denatured, digested covalent intermediate is separated into a labeled 114 -mer containing a covalently attached tyrosine and a labeled 234-mer that is uncleaved. (e) Schematic of the reaction showing intermediates that are formed.

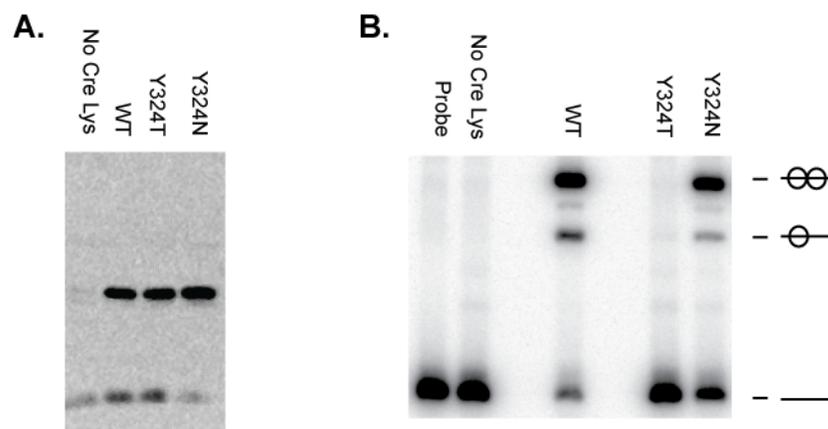
## 2.5 Results and discussion of the active site

### 2.5.1 *Tyr324 substitutions*

As anticipated, Tyr324 is strictly required for recombination. None of the nineteen alternative residues showed any detectable recombination activity *in vivo* or *in vitro*. As noted above, Cre Y324T was the only mutant identified that is substantially defective in DNA-binding (Fig. 2.9B). Western blot analysis indicates that this mutant is expressed in a full-length, soluble form and at levels comparable to wild-type Cre (Fig. 2.9A). It is possible that this mutant is misfolded or adopts an alternative structure that is not competent to bind to the *loxP* site.

The requirement for tyrosine in this position of the active site is evident from Fig. 2.2. Whereas some substitutions could in principle fulfill a role as nucleophile (e.g., serine, cysteine, lysine, histidine), the length and shape of the tyrosine side chain is unique. Indeed, a large superfamily of site-specific recombinases use serine rather than tyrosine as a nucleophile to catalyze strand exchange by a related phosphoryl transfer mechanism (Grindley *et al.*, 2006). However, even if a suitable general base were present to activate a serine residue for nucleophilic attack in the Cre active site, it is clear that the side chain of a Y324S substitution is too short reach the scissile phosphate.

The requirement for tyrosine in the YR and TopIB active sites may also extend to the identity of the phenolic hydroxyl group. Semi-synthetic vaccinia virus TopIB mutants with the tyrosine oxygen replaced by nitrogen (tyramine) or sulfur (thio-tyrosine) are both inactive, suggesting a unique requirement for tyrosine in the cleavage and ligation chemistry catalyzed by this enzyme superfamily (Gao *et al.*, 2005).



**Figure 2.9** *Y324T defective in binding, but soluble.*

Cre Y324T was the only mutant to display a significant deficiency in binding with soluble lysates.

A. 20 ul of cleared Cre lysate (grown to 1 OD) run on SDS-PAGE for western analysis and detected with Cre specific antibody. 'No Cre Lys', contains Bluescript vector lacking Cre insert.

B. Cre binding assay with 1:50 dilution of cleared lysate. Probe contains labeled substrate in binding buffer, 'No Cre Lys', contains Bluescript vector lacking Cre insert.

## 2.5.2 Arg173 substitutions

The results of substitutions at Arg173 were surprising. For the Cre R173K mutant, previous studies showed no *in vitro* cleavage or recombination activity using linear substrates (Chapter 5). It was expected that no substitutions at Arg173 would result in recombination activity above the threshold of our *in vivo* assay. However, the R173K mutant is clearly active for recombination *in vivo*. Cre R173K was previously shown to be severely defective in synapsis of *loxP* sites (Ghosh *et al.*, 2007) and the corresponding R191K mutant in Flp shows altered patterns of DNA-bending (Friesen & Sadowski, 1992). It is conceivable that substrate supercoiling or other differences in the *in vivo* reaction (Adams, Bliska & Cozzarelli, 1992) are able to compensate to some extent for these deficiencies.

The second surprise was the relatively high activity observed for R173H, a mutant that to our knowledge has not been described for the tyrosine recombinases or the TopIB enzymes. Unlike R173K, which showed no *in vitro* recombination activity, the R173H mutant showed robust activity both *in vivo* and *in vitro*. These results suggest that a lack of interaction with the O5' leaving group is unlikely to be responsible for the devastating loss of activity observed in other substitutions at this position (e.g., R173A). The hydrogen bond to O5' made by R173 is via the  $\eta$ -nitrogen, at the very end of the arginine side chain (Fig. 2.2A). At best, the shorter histidine side chain in this position would be capable of a long-range interaction with O5' via its  $\epsilon$ -nitrogen. The more likely active site configuration for Cre R173H is one where His173 hydrogen bonds to the O2P non-bridging oxygen via  $N\epsilon$ , just as Arg173 does in all Cre-DNA crystal structures that have been reported. Thus, the activity of R173H is most likely due to its ability to stabilize a buildup of negative charge on O2P rather than its ability to activate O5'.

The R173H mutant also accumulates more HJ and covalent intermediates of the reaction than are observed for wild-type Cre. Accumulation of covalent intermediate implies that  $K_{eq} = k_{clv}/k_{lig}$  (where  $k_{clv}$  and  $k_{lig}$  are the first-order rate constants for cleavage and ligation,

respectively) has increased for the covalent intermediates in the recombination pathway. Although Arg173 has been implicated in activation of the O5' leaving group, the same interaction may also play a role in guiding O5' into position for efficient ligation. Loss of this contribution could explain the mild ligation defect observed in the R173H mutant. Interestingly, the Flp R191K mutant also accumulates covalent intermediates *in vitro*, but in this case the ligation defect is more severe and no recombinant products are formed (Friesen & Sadowski, 1992).

Arg173 is strongly conserved among the tyrosine recombinase and TopIB enzymes and conservative lysine substitutions in this position show greatly diminished *in vitro* catalytic activity in the systems that have been tested (Appendix A). The Cre-DNA transition state mimic structure shows this catalytic arginine making hydrogen bonding interactions with both a non-bridging phosphate (vanadate) oxygen and with the O5' leaving group (Fig. 2.2). Thus, Arg173 plays a dual role in the transition state of the cleavage and ligation reactions.

### **2.5.3 Lys201 substitutions**

Lys201 is located in the loop that connects conserved  $\beta$ -strands 2 & 3 in the YRs and in the TopIBs, where it is inserted into the minor groove near the cleavage site and hydrogen bonds to the +1 base (Guo *et al.*, 1997; Redinbo *et al.*, 1998). Although Lys201 appears to be strongly conserved among the YRs, its position in a variable-length loop creates challenges for identifying the corresponding residue in large alignments of YR sequences (see Experimental Methods).

None of the Lys201 substitutions in Cre resulted in high recombination activity. However, several substitutions did have a low level of recombination activity *in vivo*, including the K201R mutant. Since the pKs of lysine and arginine differ by roughly two log orders, the arginine substitution would be expected to result in a 100-fold decrease in cleavage rate based only on its ability to activate the O5'-leaving group. This prediction holds true for poxvirus TopIB (Krogh & Shuman, 2000) and appears to be approximately true for Flp (Chen & Rice, 2003a) and for Cre

(this work). However, the Ala and Asn substitutions at this position also have weak *in vivo* activity, despite their lack of ability to participate in general acid catalysis. None of the K201 substitutions showed measurable *in vitro* recombination activity.

The important catalytic role of this conserved lysine was first demonstrated for poxvirus TopIB, where 5'-bridging phosphorothiolate DNA substrates were used to show that this residue is the general acid responsible for activation of the O5' leaving group during the cleavage step of the reaction (Krogh & Shuman, 2000). The same conclusion was later reached for Cre, using a similar approach (also Chapter 5) (Ghosh *et al.*, 2005a). Accordingly, the Cre K201A mutant and the corresponding mutants in Flp and Xer are inactive (Cao & Hayes, 1999; Chen & Rice, 2003a; Lee & Sadowski, 2003).

The Cre transition state mimic structure strongly supports the biochemical data and related structural results from the TopIB systems regarding the active site lysine (Fig. 2.3) (Davies *et al.*, 2006; Perry *et al.*, 2010; Krogh & Shuman, 2000; Krogh & Shuman, 2002). Lys201 hydrogen bonds directly to the O5' leaving group as well as to the +1 base, O4' of the -1 sugar, and a well-ordered water molecule that is also found in the TopIB transition state mimic structures (Fig. 2.3). The lysine side chain adopts a fully extended form, with all four side chain dihedral angles near 180°. This may explain in part why histidine is ineffective in this position; even if the bulkier imidazole ring could be accommodated in the minor groove, the  $\epsilon$ -nitrogen would fall well short of forming effective hydrogen-bonding interactions with O5'. Although the weak activity of the arginine substitution can be rationalized, it is not clear why the Ala and Asn substitutions have a comparable level of activity *in vivo*.

#### **2.5.4 Substitutions at Arg292**

Sequence alignments indicate that Arg292 is as conserved as Y324, with only a small number of lysine substitutions represented (Table 2.2). Structures of Cre-DNA (Van Duyne, 2008),  $\lambda$  Int-DNA (Aihara *et al.*, 2003a; Biswas *et al.*, 2005), and Flp-DNA (Chen & Rice, 2003b)

complexes all show this side chain hydrogen-bonding to the O1P non-bridging oxygen via N $\eta$ 2. Since lysine in this position should also be capable of interacting with O1P (based on modeling), it is not surprising that the corresponding R223K mutant in vaccinia virus TopIB retains ~10% activity (Cheng, Wang, Sekiguchi & Shuman, 1997). I also found that the R292K mutant was moderately active. In addition, I identified nine different substitutions for Arg292 that are weakly active *in vivo*. Of these, only the histidine mutant shows detectable recombination *in vitro*. The R292K substitution, on the other hand, is active both *in vivo* and *in vitro*. This mutant also accumulates the most covalent intermediate of any single active site substitution that I have studied (Table 2.3), indicating a substantial defect in substrate ligation. Similar observations have been reported for the corresponding mutants in vaccinia virus TopoIB (Cheng *et al.*, 1997), and XerCD (Arciszewska, Baker, Hallet & Sherratt, 2000).

The Cre-DNA-vanadate structure reveals a potentially important interaction involving Arg292 that was not evident in earlier recombinase-DNA crystal structures. In addition to forming a hydrogen bond with O1P, the arginine side chain hydrogen-bonds to the Tyr324 phenolic oxygen via N $\epsilon$ , forming a double hydrogen-bonding contact resembling that observed for Arg173 (but to different functional groups). A similar interaction was observed in the TopIB-vanadate structures (Fig. 2.3) (Davies *et al.*, 2006; Perry *et al.*, 2010), indicating that this is a conserved feature of the transition state in the YR and TopIB enzymes. As noted for the TopIB enzyme (Yakovleva, Chen, Hecht & Shuman, 2008), the interaction of Arg292 with the Tyr324 oxygen suggests that Arg292 could play a functional role in stabilizing the buildup of negative charge of the phenolate form of the active site tyrosine during the transition state of the reaction. The accumulation of covalent intermediate by the R292K mutant would be consistent with a role for Arg292 in activation of the tyrosine leaving group during the ligation reaction, since lysine in this position would not be expected to form optimal interactions with either the Tyr324 oxygen or with O1P for the same reasons discussed above for the R173K mutant. during the ligation reaction, since lysine in this position would not be expected to form optimal interactions with either the Tyr324 oxygen or with O1P for the same reasons discussed above for the R173K mutant.

### 2.5.5 Substitutions at Trp315

Trp315 is an outlier among the YR enzymes in that greater than 90% of family members have histidine in this position (Table 2.1). In the Cre-DNA-vanadate structure Trp315 hydrogen bonds to O1P phosphate oxygen via the indole nitrogen. A similar interaction was observed with the equivalent histidine in  $\lambda$ -integrase (Aihara *et al.*, 2003a) and for the TopIB enzymes (Perry, Hwang, Bushman & Van Duyne, 2006; Redinbo *et al.*, 1998). The results of substitutions at Cre Trp315 found that His is not as effective when substituted for tryptophan in Cre than Phe or Tyr. In addition, Ala, Gly, Leu and Met substitutions all have weak *in vivo* recombination activity. Of the weaker substitutions, His, Leu and Met have detectable recombination activity *in vitro* but the smaller Gly and Ala substitutions do not. The finding that W315M has recombination activity comparable to W315H further underscores the need for a large hydrophobic residue in this position. The *in vivo* activity of W315A and W315G would appear to contradict this model, but it is possible that these mutants adopt an alternative conformation in this region of the enzyme that is somehow able to partially compensate for the loss of W315.

The positions of Trp315 and Tyr324 in the transition state mimic structure (Fig. 2.2) provide an opportunity to carefully examine the contacts that are present at this crucial stage of the reaction. Trp315 is located on a turn between the  $\alpha$ L and  $\alpha$ M helices, where it provides a substantial amount of hydrophobic surface onto which the  $\alpha$ M helix docks via van der Waals interactions with Ile320 and Val321. Tyr324 is located at the carboxyl-terminus of  $\alpha$ M, where the tyrosine ring 'sits' on top of Trp315. Each of the carbon atoms in the Tyr324 phenyl ring contacts C $\zeta$  of Trp315 with a distance of 3.5-4 Å. Although the Phe and Tyr substitutions are predicted to be less effective than Trp in supporting the Tyr324 ring system, it is clear from modeling that a similar set of substantial contacts could be made in the transition state in both cases. Indeed, the structure of the Flp W330F/DNA complex shows that this is the case when the catalytic tyrosine has not yet attacked the scissile phosphate (Chen & Rice, 2003a). On the other hand, the smaller histidine in this position does not appear to be capable of forming similar contacts. I note

that both the Phe and Tyr substitutions at Trp315 cause substantial accumulations of HJ intermediates (Table 2.3), suggesting that repacking this hydrophobic region of the enzyme may also lead to conformational changes that influence aspects of the reaction pathway other than catalysis.

If Trp315 plays a crucial role in positioning the active site nucleophile, then it is instructive to consider how the majority of tyrosine recombinases are able to function with histidine in this position. In both Cre and Flp (and most other YRs that use Trp in this position), glycine is located two residues before His/Trp III. A side chain in this position would occupy the same space as the tryptophan side chain in the absence of large structural changes in the  $\alpha$ L- $\alpha$ M turn. This residue is normally Leu, Met, or Val in most of the YR enzymes, each of which would likely be able to fulfill the role of the active site tryptophan in Cre and Flp. In  $\lambda$  integrase, for example, a leucine side chain in this position packs against the catalytic tyrosine (Aihara *et al.*, 2003a).

In those cases where a smaller side chain is located two residues preceding His/Trp III, a large hydrophobic side chain is present four residues after Arg292 in the  $\alpha$ K helix. This residue is normally an alanine in most YRs, but a larger side chain would also project into the space occupied by the bulky Trp side chain in Cre and Flp. Thus, there appear to be a number of solutions for stabilizing the local structure and defining the position of the active site nucleophile in the YR enzymes (Ma *et al.*, 2007). In vaccinia virus TopIB, the corresponding His265 can be substituted by Asn or Gln with only a modest loss of activity (Petersen & Shuman, 1997b). In this case, the helix containing Tyr274 is anchored adjacent to the active site via direct contacts to the DNA backbone and to bases downstream of the cleavage site (Perry *et al.*, 2010), so the role of this catalytic residue for the TopIBs appears to primarily involve transition state stabilization and not positioning of the active site tyrosine. In the Flp recombinase, the W330A mutant shows an unexpected thio-effect on 5'-bridging phosphorothiolate substrates, suggesting a role in leaving group activation (Ma *et al.*, 2007). This observation cannot be readily explained based on the Cre transition state mimic structure.

### 2.5.6 *His289 is unique to tyrosine recombinases*

I found that all but five substitutions (Asp, Glu, Lys, Arg, and Trp) at His289 in Cre are active for recombination *in vivo*. Of these, Ile and Pro show only weak activity but several are nearly as efficient as wild-type Cre. All of the mutants that were active *in vivo* also showed detectable *in vitro* activity (Fig. 2.7). Two of the active mutants, H289A and H289N, have been previously studied (Gelato, Martin & Baldwin, 2005). In the transition state mimic structure, His289 adopts a curious hydrogen-bonding arrangement, with the N $\epsilon$  hydrogen directed midway between non-bridging O2P and the attacking/leaving oxygen of Tyr324. Thus, His289 is positioned in such a way that it could stabilize the buildup of negative charge on O2P and serve as a proton donor/acceptor for the Tyr324 phenol oxygen. A well ordered water molecule (B-factor  $\sim 18 \text{ \AA}^2$ ) hydrogen bonds to N $\delta$  of His289, an anchoring interaction that was also observed in the covalent intermediate structure.

The TopIB enzymes have no equivalent to His289 in their active sites. Although lysine precedes Arg223 by three residues in vaccinia and variola virus TopIB, differences in local structure allow Lys220 to be directed away from the scissile phosphate, where it instead interacts with the upstream (+1) phosphate on the cleaved strand (Fig. 1.5 and Fig 1.6)(Perry *et al.*, 2006). Consequently, the space occupied by His289 in the YRs is filled with water molecules in the TopIB structures. His289 is less strongly conserved among the tyrosine recombinases compared to the active site arginines and the tyrosine ( $\sim 87\%$ ), with  $\sim 7\%$  of sequences showing Tyr in this position and  $\sim 1\%$  with Gln. The original covalent Cre-DNA structure prompted the suggestion that His289 could be responsible for activation of Tyr324 during cleavage and ligation (Guo *et al.*, 1997; Sherratt & Wigley, 1998). Indeed, subsequent crystal structures of Cre-DNA, Flp-DNA and  $\lambda$  Int-DNA complexes supported this idea. Biochemical evidence for a role in general acid/base catalysis has come from the Flp system, where synthetic peptides with increased nucleophilicity of the catalytic tyrosine were shown to rescue cleavage activity of the His305Q mutant (Whiteson *et al.*, 2007).

The variety of amino acid substitutions at His289 that are active for recombination indicates that although moderately conserved, this active site residue is not strictly required for catalysis. This idea is consistent with the absence of an equivalent residue in the TopIB enzymes. Glutamine is probably the most conservative substitution for His289 in terms of size and ability to hydrogen bond, yet the Met substitution has similar activity *in vivo* and nearly the same activity *in vitro*. Even the Gly substitution is moderately active. As shown in Fig. 2.5B and Table 2.3, both the Gln and Leu substitutions at His289 lead to an accumulation of covalent intermediate. This effect was also observed for several His305 mutants in Flp, including H305L (Parsons *et al.*, 1988). Although one might conclude from an analysis of these substitutions alone that His289 is required for efficient ligation, the data indicate that covalent intermediate accumulation is not a general property of His289 mutants. The Ala, Gly, Met, and Asn substitutions are all moderately active, yet none of them appears to have shifted the cleavage-ligation balance in favor of cleavage. The H289M mutant does, however, accumulate a significant amount of HJ intermediate, with lesser amounts produced by the Ala and Gly substitutions. Interestingly, the Cre H289A mutant has also been shown to have an altered preference for the strand of *loxP* that is initially exchanged during recombination (Gelato *et al.*, 2005). The functional role of His289 in Cre is explored in detail in Chapter 4.

Of the inactive substitutions at His289, modeling indicates that Trp is too bulky to be accommodated in the active site without major changes in the positions of catalytic residues. Asp and Glu would in principle be capable of performing general acid/base catalysis, but the negative charges associated with the carboxylate side chains are clearly detrimental to recombination, presumably by destabilizing the transition state. The lack of detectable activity for the Lys and Arg substitutions was somewhat surprising. One possible explanation is that these side chains compete with Tyr324 for the same position adjacent to the scissile phosphate and sterically block the nucleophile. A large conformational change that occurs in helix-M and Tyr324 relative to the inactive active sites in the Cre tetramer suggests that this explanation is plausible. Alternatively, the side chains may simply be too long to be accommodated in the active site. It does not appear

possible for Lys289 or Arg289 to adopt conformations that resemble Lys220 in the viral TopIB structures due to steric interference with the inter-domain linker that is quite different in Cre compared to the viral TopIB protein.

### **2.5.7 Glu176 organizes the active site**

Glu176 does not interact directly with the scissile phosphate, but plays a role in organizing the Cre active site. Although Glu is favored in this position of the YR family (79%), Asp is also highly represented (16.4%), with over 95% using one or the other acidic side chain (Table 2.1). Fifteen substitutions for Glu176 in Cre were active for recombination *in vivo*, although the His, Met, Pro, Val, Trp, and Tyr substitutions were very weak. *In vitro*, the His, Trp, and Tyr mutants showed no detectable activity. Surprisingly, the strongest substitutions were Asn and Thr, each of which showed high *in vitro* recombination activity. The conservative E176D mutant was somewhat less active and the E176Q mutant showed no measurable activity at all in the *in vitro* assay. The behavior of E176Q was unexpected; presumably this residue stabilizes an alternative, but unproductive arrangement of active site residues. The modest activities resulting from Ala and Gly substitutions indicate that loss of the interactions involving Arg173 and Lys201 diminishes, but does not destroy catalysis in this system. The functional role of Glu176 and the interactions made with Arg173 and Lys201 are explored in detail in Chapter 5.

Several substitutions for the corresponding Asp194 in Flp recombinase have also been studied (Appendix A). Of these, it is interesting to note that the D194N mutant is deficient for cleavage of DNA half-sites (Chen *et al.*, 1992a). As noted above, the corresponding conversion from acid to amide in Cre is also defective. Although none of the Cre Glu176 mutants accumulate significant amounts of covalent intermediate, many of them do accumulate Holliday intermediates.

In the transition state mimic structure, Glu176 accepts three hydrogen bonds from catalytic residues. The side chain carboxylate forms a hydrogen-bonding bridge between the backbone amide and N $\eta$ 1 of Arg173, appearing to lock this key residue into a conformation that is optimal for catalysis. The Glu176 carboxylate also hydrogen bonds to the backbone amide of Lys201, helping to position the  $\beta$ 2- $\beta$ 3 loop in the minor groove adjacent to the active site.

Both  $\lambda$  integrase and Flp recombinase have aspartic acid in this position. In the  $\lambda$  Int-DNA complex, Asp215 anchors the  $\beta$ 2- $\beta$ 3 loop by hydrogen bonding with the backbone amide of Ser234 (which precedes the conserved active site Lys235) and makes a long-range interaction with Arg212- N $\eta$ 1 (Aihara *et al.*, 2003a; Biswas *et al.*, 2005). In the Flp-DNA complex, Asp194 hydrogen bonds to the backbone amide of Arg191 and interacts with the amide of Lys223 indirectly via bridging water molecules (Chen *et al.*, 2000). Thus, Asp and Glu in this position of the YR enzymes carry out similar structural roles, but the detailed interactions differ among the individual systems. The TopIB enzymes do not have a residue corresponding to Glu176 in Cre. Instead, tightly bound water molecules in the active site make a similar set of interactions with Arg130 and the Lys167 backbone (Fig. 2.3) (Davies *et al.*, 2006; Perry *et al.*, 2010).

## 2.6 Discussion

The Cre active site residues can be grouped into three categories based on their tolerance to substitution. In the first category, Tyr324 and Lys201 are essential for catalysis; I found no active substitutions of Tyr324 and only very weak ones for Lys201. Thus, a nucleophile and a general acid catalyst to activate the 5'-hydroxyl leaving group play the most unique and essential roles in the reaction. In the second group, Arg173 and Arg292 each play dual roles in catalysis and are symmetrically organized in the transition state with respect to cleavage and ligation. Both residues stabilize the buildup of negative charge by hydrogen-bonding to a non-bridging phosphate oxygen. These residues also hydrogen bond to the nucleophile and leaving groups of the reaction: Arg173 to O5' and Arg292 to the Tyr324 hydroxyl. Thus, the roles of the two

residues become interchanged during cleavage vs. ligation. The third category includes Trp315, His289, and Glu176, each of which is tolerant to substitutions that result in moderate to high levels of recombination activity, with additional substitutions that have lower levels of activity. No active site substitution resulted in an increase in *in vivo* or *in vitro* recombination activity relative to wild-type Cre.

For most active site residues, there is a high correlation between the degree of substitution allowed and the diversity of residues identified at the corresponding position in the protein database. The exception is Arg173, which is present as Arg in only 91% of the YR protein sequences. Interestingly, the lysine substitution is found in only ~1% of the sequences and the remainder are distributed among Ala, Ser, and other residues. One explanation could be that there are alternative active site compositions and/or geometries that do not require the box I arginine for activity. Alternatively, the YRs lacking Arg I may represent a group that is catalytically inactive, but has nonetheless retained Arg II and the active site tyrosine.

The active site residues most sensitive to substitution are those shared by the TopIB enzymes, which catalyze the same basic phosphoryl transfer reactions very efficiently without His289 and Glu176 equivalents. Why, then, is His II moderately conserved in the YR family and why are Asp and Glu nearly always found in the position corresponding to Glu176? Part of the answer may be related to the differences between the reactions catalyzed by the YR and TopIB enzymes. The TopIBs cleave their targets to form covalent 3'-phosphotyrosine intermediates and allow rotation of DNA downstream of the cleavage site about the site of the nick. TopIB evolution has presumably tuned the relative rates of cleavage and ligation to strike a balance between allowing escape of the 5'-nucleoside (and subsequent relaxation of supercoils) and minimizing exposure of the nicked chromosome to potentially damaging processes. In contrast, the YR enzymes need to coordinate the cleavage, exchange, and ligation of DNA strands between two

active sites in the recombinase tetramer, while rendering the other two sites inactive. In the absence of such allosteric regulation, YR enzymes could generate double-strand breaks across the 6-8 bp overlap region.

The details of the inter-subunit interactions responsible for coordinating cleavage in the YR systems differ considerably among the Cre,  $\lambda$  integrase, and Flp systems. However, one general feature in common to each of these systems involves positioning of the active site tyrosine, where availability of the nucleophile for catalysis is strongly influenced by interactions with the neighboring subunit (Discussed in Chapter 6) (Aihara *et al.*, 2003a; Arciszewska *et al.*, 2000; Chen & Rice, 2003b; Van Duyne, 2001). For Flp recombinase, the active site tyrosine is donated in trans from an adjacent subunit (Chen, Lee & Jayaram, 1992b). An additional source of regulation may involve positioning of the active site lysine, which is also influenced by inter-subunit contacts. One consequence of evolving a regulated recombinase active site from an ancestral topoisomerase may therefore have been a loss of some catalytic power, as the dynamics of active site residues were modified in order to allow allosteric control of their positions. I suggest that the YR active site residues corresponding to His289 and Glu176 in Cre were added in order to compensate for this loss in activity. Indeed, loss of either of these residues in Cre to generate a more 'Topo-like' active site results in a decrease in recombination activity and for those mutants in Cre and other systems where cleavage has been assayed specifically, a decrease in cleavage rates have been reported (Appendix B).

Ironically, a comparison of the static transition state models between the TopIB and YR enzymes might lead to the conclusion that the YRs should be more efficient in catalyzing cleavage and ligation. In the Cre transition state mimic structure, His289 provides both additional transition state stabilization and a potential activating function for Tyr324. Glu176 plays a structural role in positioning Arg173 and Lys201. In the corresponding variola virus and trypanosomal TopIB transition state mimics, these functional roles appear to be carried out by ordered water molecules (Fig. 2.3). However, estimates of the pseudo-first order cleavage rate

constants for half-site and full-site DNA substrates for Cre range from 0.001-0.003 s<sup>-1</sup> (Gelato *et al.*, 2005; Ghosh *et al.*, 2005a; Ma *et al.*, 2009a), compared to 0.3-0.7 s<sup>-1</sup> for vaccinia virus TopIB (Krogh & Shuman, 2000; Stivers *et al.*, 2000). Flp recombinase, on the other hand, cleaves half-site and full-site DNA at rates that are closer to that of the TopIB enzymes (Ma *et al.*, 2009b). Thus, the additional active site residues found in the YR enzymes do not appear to provide a substantial increase in their ability to cleave and ligate DNA substrates relative to their TopIB cousins.

The active site substitutions that I have described, when combined with transition state mimic structures of Cre and TopIB, provide a useful foundation for comparing the two systems and for understanding the requirements for catalysis during Cre-*loxP* recombination and presumably in the other YR family members as well. In the YR systems, however, it is an oversimplification to think of the active site as being devoted entirely to catalysis of cleavage and ligation. The Cre R173K and H289A mutants are examples where this is clearly the case. Cre R173K is defective in synapsis of *loxP* sites, which may contribute to the lack of *in vitro* activity for this mutant. In the structure of the pre-cleavage synaptic Cre-*loxP* complex (using the Cre K201A mutant), Arg173 hydrogen bonds to the scissile phosphate in the inactive sites, stabilizing a phosphate backbone conformation adjacent to the asymmetric *loxP* bend (Ghosh *et al.*, 2007). The H289A mutant has the unexpected property of promoting a top-strand-first order of *loxP* strand exchanges during recombination, the opposite of that preferred by wild-type Cre (Gelato *et al.*, 2005). His289 also plays an important role in the inactive site of the synaptic complex, nucleating a hydrogen-bonding network involving the Arg173 backbone, two water molecules, and the active site phosphate (Ghosh *et al.*, 2007). Related allosteric effects have been reported in the Xer system, where an R148K mutant in XerC is able to stimulate the catalytic activity of XerD on HJ substrates and vice versa (Arciszewska *et al.*, 2000). Thus, active site residues can also strongly influence the reaction via interactions involving the half-sites not undergoing catalysis.

Efforts thus far to correlate biochemical activities with structural models in the Cre, Flp, and  $\lambda$  Int systems indicate that while the YRs have evolved strategies for inter-subunit communication that share some overall themes, they do differ considerably in their molecular details. This is especially apparent when comparing Flp recombinase to the bacterial YRs, where the inter-subunit contacts in the recombinase tetramers are quite different. Consequently, the roles of active site residues in mediating the allosteric regulation of cleavage and ligation in these enzymes are likely to differ as well. An example of this can be seen in a comparison of the Cre H289Q mutant with the corresponding H305Q mutant in Flp. The Cre mutant is nearly as active as wild-type Cre (this work and Chapter 4), yet the Flp H305Q mutant is severely defective (Parsons *et al.*, 1988). Dissection of active site contributions to chemical catalysis vs. mediating the conformational changes and dynamics of the recombinase tetramer through the recombination pathway may therefore prove to be challenging.

## 2.7 Methods

### ***Reagents and strains***

Reagents were obtained from Sigma, unless otherwise noted. Strains CSH100 (F' lacproA+,B+(lacIq lacPL8)/ara  $\Delta$ (gpt-lac)<sup>5</sup>) and CSH142 (F-/ara  $\Delta$ (gpt-lac)<sup>5</sup>) were obtained from the E. coli Genetic Stock Center (CGSC8105 and CGSC8083) at Yale University. Plasmids pFW11 and pTSA29 were gifts from Dr. Fred Whipple and Dr. Gregory Phillips, respectively.

### ***Crystallization and diffraction data***

Cre was over-expressed and purified as previously described (Ghosh & Van Duyne, 2002). Oligonucleotides were synthesized by the Keck Biotechnology Facility at Yale University and purified by ion-exchange chromatography. Sodium orthovanadate (Sigma) was activated as

described (Gordon, 1991) and added to a premixed solution containing 40  $\mu\text{M}$  Cre, 75  $\mu\text{M}$  half-site DNA, 20 mM  $\text{CaCl}_2$ , 20 mM sodium acetate, pH 5.6, 1 mM dithiothreitol, and 25% 2-methyl-2,4-pentanediol (MPD) to a final concentration of 2 mM. Crystals were grown by hanging drop vapor diffusion against a well reservoir of 20 mM sodium acetate, pH 5.0, 20 mM  $\text{CaCl}_2$ , and 30-45% MPD. Crystals were flash frozen in liquid nitrogen prior to data collection. Diffraction data were measured at the Advanced Light Source beamline 8.2.1 and processed using HKL2000 (Otwinowski & Minor, 1997).

### ***Structure solution and refinement***

The structure was determined by molecular replacement with AMORE (18), using a search model based on the pre-cleavage synaptic complex (pdb entry 2HOI) where crossover region DNA and all active site residues had been removed. Initial rounds of refinement were performed with CNS (Brünger *et al.*, 1998) following model building and adjustment with COOT (Emsley & Cowtan, 2004). Final refinements were performed with REFMAC (Murshudov, Vagin & Dodson, 1997). Active site residues, the vanadate linkage, and omitted DNA residues were fit into clear, unbiased electron density. Vanadate geometry was restrained during refinement as described for the viral TopIB-DNA complex (Perry *et al.*, 2010). Coordinates for the refined structure have been deposited with the Protein Data Bank with accession code 3MGV.

### ***Generation of active site mutants***

An XbaI-HindIII fragment containing the wild-type Cre coding sequence was cloned into the same sites of pBluescriptIIKS+ (Stratagene) and used as a template for PCR-based mutagenesis using a modified version of the Phusion protocol (New England Biolabs; NEB). Most substitutions were generated using primers with degenerate codons and those remaining following initial sequencing were obtained using specific primers. In both cases, only the primer containing the altered active site codon was phosphorylated during oligonucleotide synthesis (Integrated DNA

Technologies; IDT) and a 1 hour DpnI digestion following ligation of the PCR products was used to remove the template. All mutant plasmids were sequenced twice to verify the correct substitution.

### ***In vivo recombination reporter***

A single copy reporter was generated by homologous recombination of an intermediate plasmid based on pFW11 into an F' containing an intact lac operon (Whipple, 1998). A modified *sulA* promoter (Dmitrova *et al.*, 1998) was cloned into the EcoRI and BamHI sites of pFW11 using oligonucleotides P1 and P2 (Table 2.4) and a fragment containing the *aadA* coding sequence with a 5'-KpnI site was amplified using primers P3 and P4 and ligated into the BamHI and Sall sites of *sulA*-FW11 to give pGV1026. A terminator cassette that could be excised by Cre was generated by inserting the *E. coli* *rpoC* terminator (Brosius, 1984) amplified using primers P5 and P6 into the BsrGI and HincII sites of pGV697, a pBluescript derivative containing directly repeating *loxP* sites flanking an 867 bp segment (GV, unpublished), to generate pGV1025. The resulting construct contains *loxP* sites flanking a 920 bp segment containing *TrpC*. The excisable element was then inserted into the BamHI-KpnI sites of pGV1026 to give pGV1035. pGV1035 was then transformed into CSH100 and double crossovers were selected following transfer of the recombinant episome into CSH142/pTSA29 by conjugation as described (Whipple, 1998). pTSA29 provides ampicillin resistance and a temperature sensitive pSC101 origin, allowing facile removal following incubation at 42°C (Phillips, 1999). The relevant segment of the resulting F' (pGV1048) is shown schematically in Fig. 2.4A.

*Table 2.4: Oligonucleotide sequences used in construction of in vivo recombination reporter*

P1 : 5' -AATCAATAGGGTTGATCTTTGTTGTCACTGGATGTACCGTACATCCATACAGTAACTCAC-AGGGCTG-3'  
P2 : 5' -GATCCAGCCCTGTGAGTTACTGTATGGATGTACGGTACATCCAGTGACAACAAAGATCAAC-CCTATTG-3'  
P3 : 5' -GTCAGTGGATCCTAGGTACCGAAAAAGGAAGAGTATGAGGG-3'  
P4 : 5' -GTCAGTGTCGACTTATTTGCCGACTACCTTGG-3'  
P5 : 5' -GTCAGTTGTACACGTGCTGCGGGTGAAGC-3'  
P6 : 5' -CTGACAAATGCTCTTTCCCTAAACTC-3'

Restriction sites are underlined

### ***In vivo recombination***

Cre mutant plasmids were transformed into chemically competent CSH142/pGV1048 cells and after a one hour incubation at 37°C were plated on LB containing ampicillin (100 µg/ml), kanamycin (35 µg/ml), and X-gal (20 µg/ml). An equal volume was plated on LB containing the same components plus 20 µg/ml streptomycin. Activity was recorded as the ratio of the number of colonies after 16 hours at 37° on the streptomycin plate divided by the number on the plate lacking streptomycin. For wild-type Cre, this ratio was reproducibly close to 1, with a colony count of ~1000. For the empty pBluescript and Cre Y324F negative controls, the background was zero; no streptomycin resistant colonies were observed in multiple experiments and colonies on the non-selection plate were white. For weakly active mutants where a small number of colonies were observed on the streptomycin plate, *in vivo* activity was corroborated by the appearance of blue colonies in the absence of streptomycin selection. *in vivo* assays were repeated 3-4 times and the results were tabulated as a percentage of wild-type activity with typical standard deviations of 9% for the most weakly active mutants.

### ***In vitro Recombination Assay***

The substrate for *in vitro* recombination was a linear 602 bp DNA fragment digested from pBG1701, dephosphorylated, and <sup>32</sup>P end-labeled with polynucleotide kinase. pBG1701 was constructed by ligation of a DNA segment containing two *loxP* sites flanking a 334 bp spacer, shown schematically in Fig. 2.7A. Cre-containing lysates were prepared from CSH142/pGV1048 transformants by sonication of cells from 2 ml culture grown at 37° to an optical density of ~1.0 in 100 µl lysis buffer (50 mM Tris pH 7.4, 150 mM NaCl, 2mM dithiothreitol, 1 mM EDTA, 0.5% Triton-X100 and protease inhibitors). Cleared lysates were diluted in lysis buffer and used directly for *in vitro* binding and recombination assays. Recombination reactions were carried out in 25 µl assay buffer (50 mM Tris pH 7.4, 150 mM NaCl, 2mM dithiothreitol, 1 mM EDTA, 50

µg/ml sheared salmon DNA) containing 200 pM labeled substrate and a 50-fold dilution of lysate. After incubation at 37°C for 1 hour, reactions were quenched with 0.5% SDS, heated at 70°C for 1 minute, and digested with proteinase K for one hour at 37°C. Reactions were analyzed on a 6% acrylamide gel in 1X TBE for 2 hours at 10 V/cm. Gels were dried and quantified on a Phosphorimager. Errors for the *in vitro* recombination assays were estimated at 5-10% of wild-type activity based on multiple repetitions of a subset of weak and moderately active mutants.

### ***In vitro loxP binding***

The probe for electrophoretic mobility shift assays was a 64-mer duplex containing the *loxP* site flanked by 15 bp G·C-rich segments. 10 pmol of PAGE-purified top strand was <sup>32</sup>P end-labeled with polynucleotide kinase and annealed to a purified, unlabeled bottom strand. Cleared lysate was added (50:1 final dilution) to 200 pM labeled substrate in assay buffer and the reactions were incubated for 30 min at 20°. Binding reactions were electrophoresed on a 6% polyacrylamide gel (29:1 acrylamide:bisacrylamide) in TBE that was pre-run for 10 V/cm for 45 min at 20 °C prior to loading. The gel was run at 13 V/cm for two hours at 20°C, dried, and analyzed with a phosphorimager.

### ***Tyrosine recombinase sequence alignments***

the Cre catalytic domain (residues 134-343) was used as a query sequence for three iterations of PSI-BLAST (Altschul *et al.*, 1997) searching of the NCBI RefSeq database (Pruitt *et al.*, 2007) and 12,049 alignments (maximum E-value = 0.01) that included all seven active site residues were analyzed by quantifying the amino acid frequency at each of the seven positions. The catalytic lysine is not well-aligned due to its position in a loop of variable size and sequence. For this residue, we counted lysine (or arginine) as present at this site if it was located within five residues of Cre Lys201 in the alignment. We obtained similar results when the λ integrase catalytic domain was used as a search query.

### ***Identification of covalent intermediate***

Purified Cre (wild-type and H289Q) was added to recombination substrate in assay buffer to 50 nM concentration and the reaction carried out as described in the main text. Following proteinase K digestion, one half of the reaction was ethanol precipitated, resuspended and denatured in formamide, and loaded on a 4% (19:1) 7M Urea-TBE gel (19:1 acrylamide:bisacrylamide) that was pre-run for 45 min at 7 watts. The samples were electrophoresed for 2 hours at 7 W, the gels fixed in 10% MeOH/10% Acetic Acid, dried, and analyzed by phosphoimager analysis. The other half of the reaction was electrophoresed on a native 6% TBE polyacrylamide gel.

### ***Identification of HJ intermediate***

Recombination reactions were performed with purified wild-type Cre, Cre H289Q and Cre A312T. The ability of Cre A312T to accumulate intramolecular HJ intermediates and the analysis of the resulting 'alpha' DNA structure has been described (Hoess *et al.*, 1987). Following a one hour incubation at 37°, half the volume of each reaction was digested with proteinase K and the other half was extracted with phenol-chloroform. Following ethanol precipitation, samples were resuspended in BsrGI reaction buffer (NEB) and again divided in half, where one half was digested with BsrGI and one was not. Since the 602 bp recombination substrate contains an BsrGI site between *loxP* sites, the HJ intermediate alpha structure is converted to the open  $\chi$  form upon digestion. Samples were electrophoresed on a 4% TBE acrylamide gel at 13 V/cm for 100 min.

### ***Detection of Cre by Western Blot***

20 ul of 1OD of cleared lysate prepared as described above, was loaded on SDS-Page and separated at 20 V/cm for 1 hour. Gels were transferred to nitrocellulose membrane (Bio-Rad) using a semi-dry transfer apparatus (Bio-Rad) at 17V for 30 minutes. Following transfer, the

membrane was soaked for 16 hours at 4 °C in TTBS buffer containing 3% BSA and 0.1% Triton. Membrane was exposed to a 1:5000 dilution of primary antibody that is specific for Cre (Novagen) for 1 hour at 20 °C and then washed three times with TTBS. The membrane was then incubated with a secondary antibody and developed with a chemiluminescent blotting substrate (Pierce).

## Chapter 3. A Dual Role in Catalysis for H289 in Cre

### 3.1 Introduction

Tyrosine recombinases (YR) contain a well conserved histidine that is not conserved in Type IB topoisomerases. This histidine, H289 in Cre is positioned in the active site proximal to the catalytic tyrosine nucleophile and is hypothesized to function as a proton acceptor/donor in the activation of tyrosine during cleavage and ligation reactions. TopIB enzymes have no analogous residue in their active sites. A lysine, K220 in viral topoisomerases, is related by sequence, however this lysine is positioned away from the active site making a non-specific interaction with the phosphate backbone of the cleaved strand. Mutation of K220 to alanine results in a small deficiency in relaxation and cleavage, which indicates that the interaction made by the lysine is important but not essential to topoisomerase activity (Cheng *et al.*, 1997). The K220D mutant however, is defective in relaxation and cleavage, but is capable of binding (Klemperer & Traktman, 1993). The space corresponding to the histidine is filled by two well ordered water molecules in TopIB-DNA complex structures (Davies *et al.*, 2006; Perry *et al.*, 2010).

The work described in Chapter 2 described activities for all nineteen substitutions at Cre H289. Fifteen of these retained activity *in vivo*, and all nineteen were capable of binding *loxP*, indicating that the mutants inactive for recombination were from a catalytic defect and not from a folding or solubility artifact. The fifteen active H289 mutants were confirmed for activity *in vitro* with soluble lysates. Surprisingly, the H289L and H289Q mutants accumulated a significant amount of covalent intermediate during *in vitro* recombination, prompting the suggestion that these mutations may have a deficiency in ligation. The leucine mutant has been demonstrated to accumulate covalent intermediate in other YR homologs (Cornet, Hallet & Sherratt, 1997; Lee & Jayaram, 1993; Vanhooff *et al.*, 2009). The glutamine mutant has also been studied in Flp, where H305Q could be rescued for cleavage and ligation with peptides containing modified tyrosine analogs supporting a role for histidine in acid/base catalysis (Whiteson *et al.*, 2007). The alanine

and asparagine substitutions of H289 in Cre were found to have an opposite strand bias from WT. These mutants cleaved the top strand of *loxP* preferentially and also accumulated Holiday junctions (Gelato *et al.*, 2005; Martin *et al.*, 2002).

This chapter examines the role of H289 and attempts to determine the basis for conservation in YRs but not in TopIB enzymes. Studies of all nineteen mutants of H289 reveal that the histidine is important but not necessary for recombination, binding, synapsis, and cleavage of *loxP*. Detailed investigation of the biochemical activity of the conservative mutant H289Q confirms that the histidine does not have an essential role in acid-base catalysis. However, mutants at H289 are defective in water mediated hydrolysis of the 3'-phosphotyrosine linkage. Ligation of *loxP* DNA appears to be substantially defective in some histidine mutants, including H289Q and H289L.

## **3.2 Studies of all 19 H289 mutants**

Fifteen of twenty amino acids at position 289 are active for recombination *in vivo*, yielding little direct insight into its conservation. To confirm these previous results and further characterize H289, all nineteen single site variants at position 289 were subcloned into a T7 expression vector and purified as described in Methods. The mutants were tested for their ability to bind, synapse, cleave and mediate recombination. A summary of these results is given in Table 3.1, and results from individual assays will be described below.

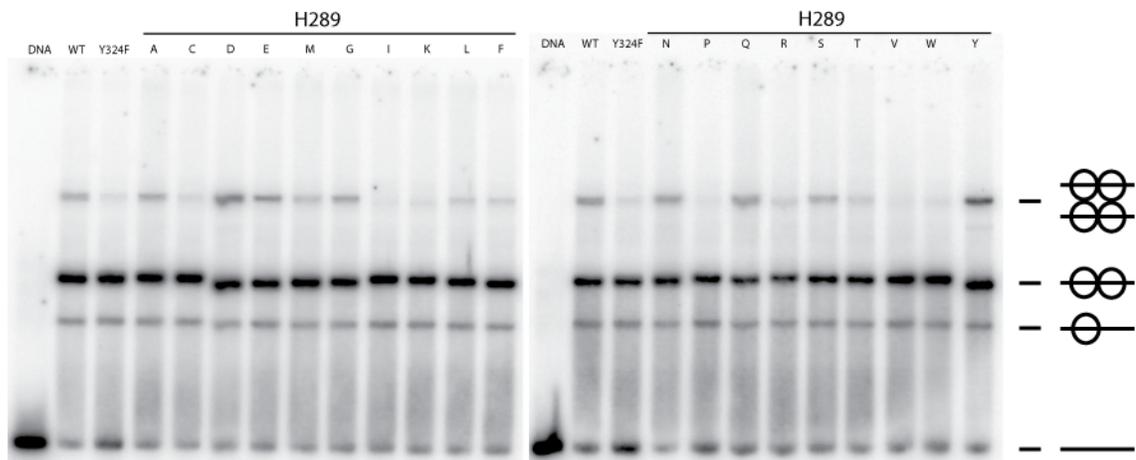
Table 3.1: H289 mutant activity summary

H289 Mutant	F' In vivo assay <sup>a</sup>	Lysate in vitro recombination <sup>b</sup>	Pure in vitro recombination <sup>b</sup>	Binding	Synapsis <sup>c</sup>	Cleavage <sup>d</sup> (1-hour)	Cleavage <sup>d</sup> (24-hour)	Covalent <sup>d</sup> (1-hour)	Covalent <sup>d</sup> (24-hour)
WT	+++	44.67	63.04	WT	3	84.40	92.51	8.87	4.91
Y324F	0	0.00	0.00	WT	1	0.55	0.31	0.00	0.00
A	+++	9.42	8.18	WT	3	12.26	60.43	3.72	4.91
C	++	2.42	0.38	WT	1	4.45	18.11	2.26	3.03
D	0	0.00	0.00	WT	3	0.00	0.00	0.00	0.00
E	0	0.00	0.00	WT	3	0.00	0.00	0.00	0.00
F	++	0.74	0.18	WT	2	2.55	8.53	1.16	3.10
G	+++	7.40	6.10	WT	3	8.63	37.20	3.56	4.50
I	+	0.13	0.07	WT	0	0.93	3.47	0.74	2.12
K	0	0.00	0.00	WT	1	0.72	0.16	0.00	0.00
L	++	12.89	1.97	WT	2	28.91	46.43	15.19	26.52
M	+++	21.15	25.92	WT	2	18.18	47.59	10.60	12.51
N	+++	30.37	27.11	WT	3	13.40	44.87	3.82	5.62
P	+	0.31	0.43	WT	0	3.93	35.35	1.45	2.61
Q	+++	30.90	38.99	WT	3	59.83	80.30	32.26	39.51
R	0	0.09	0.00	WT	2	0.76	1.72	0.00	0.00
S	++	4.70	1.31	WT	3	7.80	48.10	2.92	4.48
T	++	1.43	0.41	WT	2	4.16	23.37	2.29	3.96
V	++	0.26	0.09	WT	0	1.46	14.83	1.10	2.08
W	0	0.00	0.00	WT	1	0.00	0.00	0.27	0.48

a) Scoring of in vivo recombination the same as described in Chapter 3. b) In vitro recombination is reported as %excision. c) Synapsis is a qualitative assessment of the ability to synapse where 3 is WT, 2 is noticeably reduced, 1 is barely visible, and 0 no synapsis. d) Cleavage (suicide cleavage) and Covalent (equilibrium cleavage) reported as %total covalently bound Cre at 1 hour and 24 hours.

### 3.2.1 H289 mutant binding and synopsis

The binding and *in vitro* recombination activities of the purified H289 mutants agree well with those obtained from soluble lysates (Chapter 2). Minor differences may be attributed to the concentration of Cre protein in the lysate or cellular artifacts in the lysates. The binding assay was performed using a 64 bp *loxP* substrate at roughly 1 nM (instead of 0.2nM), which permits the formation of a small amount of synopsis and allows for a crude assessment of the ability of the H289 mutants to form synaptic complexes. The Cre concentration was 10nM, nearly 50 fold above  $K_D$  for binding, where WT Cre is completely bound to *loxP* as a dimer (Ringrose *et al.*, 1998). As Fig. 3.1 shows, all mutants at H289 are competent for binding to *loxP*, but the ability to synapse varies considerably. Several H289 mutants such as, Asp, Glu, Tyr and Met show a slight but reproducible change in mobility of the dimer species bound to *loxP*. Other mutants also appear to have altered mobilities, however their change is less apparent. These altered mobilities suggest that H289 mutants effect bending of *loxP* (Lee, Chu & Sadowski, 2003). Eight of the H289 mutants, Cys, Ile, Lys, Pro, Arg, Thr, Val, and Trp appear to have a severe deficiency in synopsis. The qualitative assessment of synopsis, shown in Table 3.1, is useful, however, full synopsis assays titrating *loxP* concentration would be necessary to be able to make detailed comparisons.



**Figure 3.1** *Binding and synapsis of purified H289 mutants*

Binding of 20 nM Cre to 1 nM *loxP*. All mutants bind *loxP* without significant deficiency, but some mutants display altered mobility relative to WT Cre. The mutants at H289 varied in the ability to synapse *loxP* sites.

### 3.2.2 H289 mutants in suicide cleavage

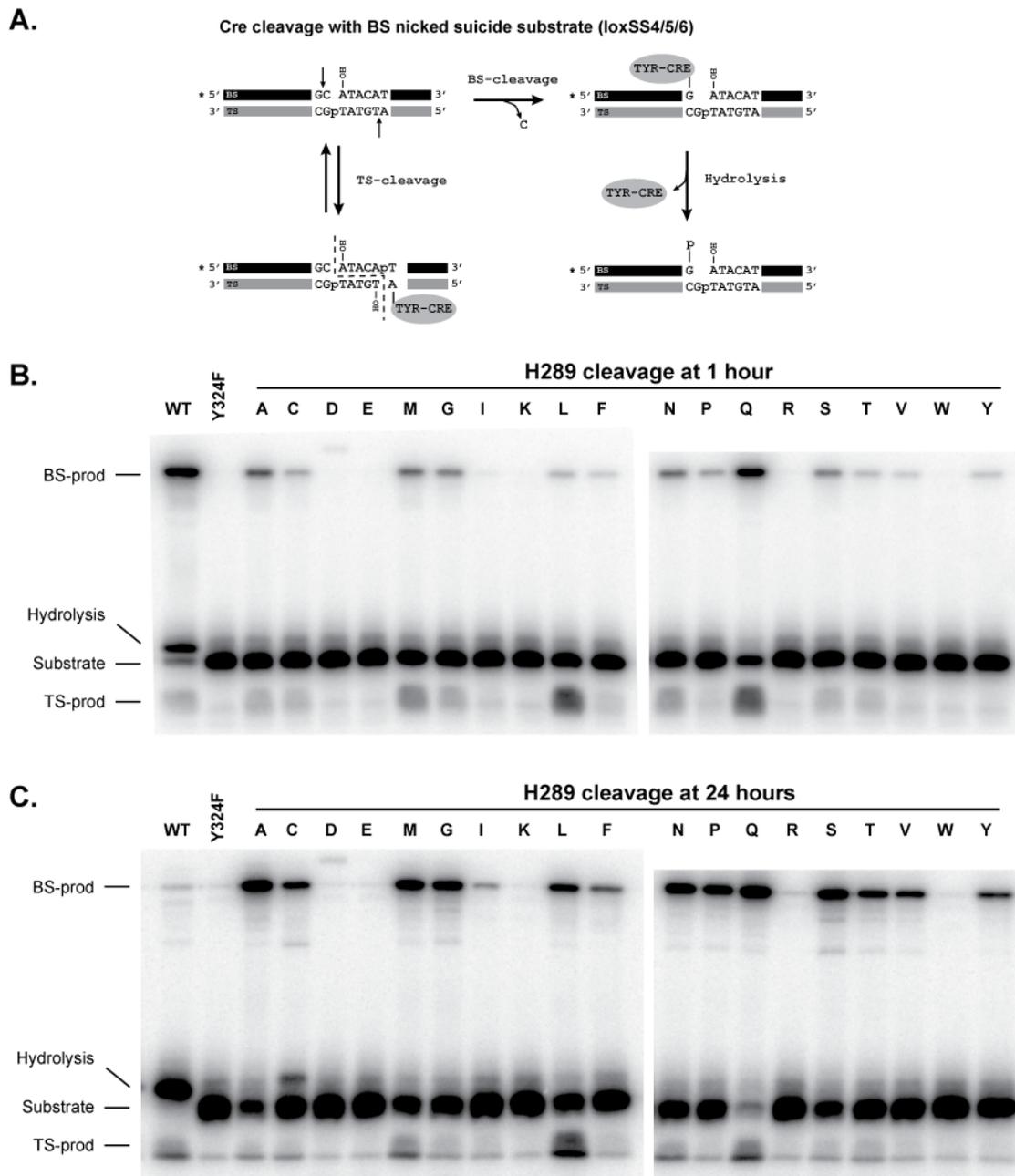
If H289 in Cre is involved in acid-base chemistry during phosphoryl transfer, then mutants at this position should have deficiencies in the ability to cleave *loxP*. Cre cleavage activity is measured using several different substrates shown in Appendix B, and described below. The cleavage ability of the H289 mutants was assayed using a suicide substrate that contains a nick at the +1 base on the bottom strand (BS; *loxSS4/5/6*; Fig. 3.2A for reaction diagram). The bottom strand of *loxP* is defined as the strand containing the cleavage site between C and G while the top strand (TS) is defined as the strand containing the cleavage site between A and T (Fig. 1.2B). Cre cleavage of *loxSS4/5/6* at the BS results in the dissociation of the +1 Cyt nucleoside, shortening the crossover region and effectively preventing religation. As a result, the 3'-phosphotyrosine intermediate is trapped. The TS of this substrate remains intact, so Cre is capable of cleavage, ligation and strand exchange using this substrate strand.

Cleavage of the H289 mutants was assayed with *loxP* concentrations at 50 nM, which favors synapsis formation and cleavage of the BS (Ghosh *et al.*, 2005a). Reactions were incubated at 37 °C for 1 hour and 24 hours to allow weakly active mutants to reach completion (Figs. 3.2B and C). Quenched reactions were separated by SDS-PAGE. Cre covalently bound to the labeled BS appears as a single band well above unreacted substrate. Cleavage on the intact strand results in a double strand break separated by 6 bp in the crossover region, which will dissociate forming two half site duplexes. The first duplex contains the labeled strand with greater mobility than the unreacted substrate. The second duplex is an unlabeled half site duplex containing covalently bound Cre to the TS site and is not visible in the gel (Fig 3.2A).

The H289 mutants active in recombination generated detectable levels of cleavage product at 1 hour and 24 hour intervals. Most mutants displayed greater cleavage product at 24 hours, indicating that the reaction at 1 hour was incomplete. Of the nineteen mutants, H289Q was the most active in cleavage, with near WT activity. The H289L, H289M, and H289Q mutants

generated TS cleavage product in greater amounts than WT Cre (Fig. 3.2). Cleavage of the intact TS is in equilibrium between cleavage and ligation, so covalent product accumulation on the TS means that either the rate of cleavage has increased with respect to the rate of ligation or the rate of ligation has decreased with respect to the rate of cleavage. Previous reports have observed that the leucine and glutamine mutants in related YRs also accumulate covalent intermediate, which is consistent with the histidine being conserved to function in ligation (Cornet *et al.*, 1997; Pan *et al.*, 1993b; Parsons *et al.*, 1988; Vanhooff *et al.*, 2009; Whiteson *et al.*, 2007). In addition, these mutants appear to preferentially cleave the TS half site. The observed TS bias could be an artifact of the nick, however the effect is only observed for a subset of mutants at H289, and the TS cleavage bias has been observed previously in Cre for H289A and H289N (Gelato *et al.*, 2005). The strand cleavage preference is examined in greater detail in section 3.3.

A second observation from the cleavage assays was that the covalent cleavage product for wildtype Cre disappeared at 24 hours with the concomitant appearance of a unique band with reduced mobility relative to unreacted substrate (Fig. 3.2). Of the H289 mutants, only H289C produced a similar band after 24 hours of incubation. Separate experiments discussed in section 3.3 helped to identify this band as hydrolysis product of the 3'-phosphotyrosine intermediate. WT Cre efficiently hydrolyzes covalent intermediate, generating nearly 50% of total counts at 1 hour and essentially 100% after 24 hours. The hydrolysis reaction is predicted to proceed in a mechanistic manner similar to that of ligation (Petersen & Shuman, 1997a). Since the H289 mutants are defective in hydrolysis, this provides further evidence that H289 may have a role in ligation. Hydrolysis is examined further in section 3.3.



**Figure 3.2** Suicide cleavage of purified H289 mutants at 1 hour and 24 hours.

(A) Schematic showing reaction products generated by Cre with the BS-nicked suicide substrate (loxSS4/5/6)

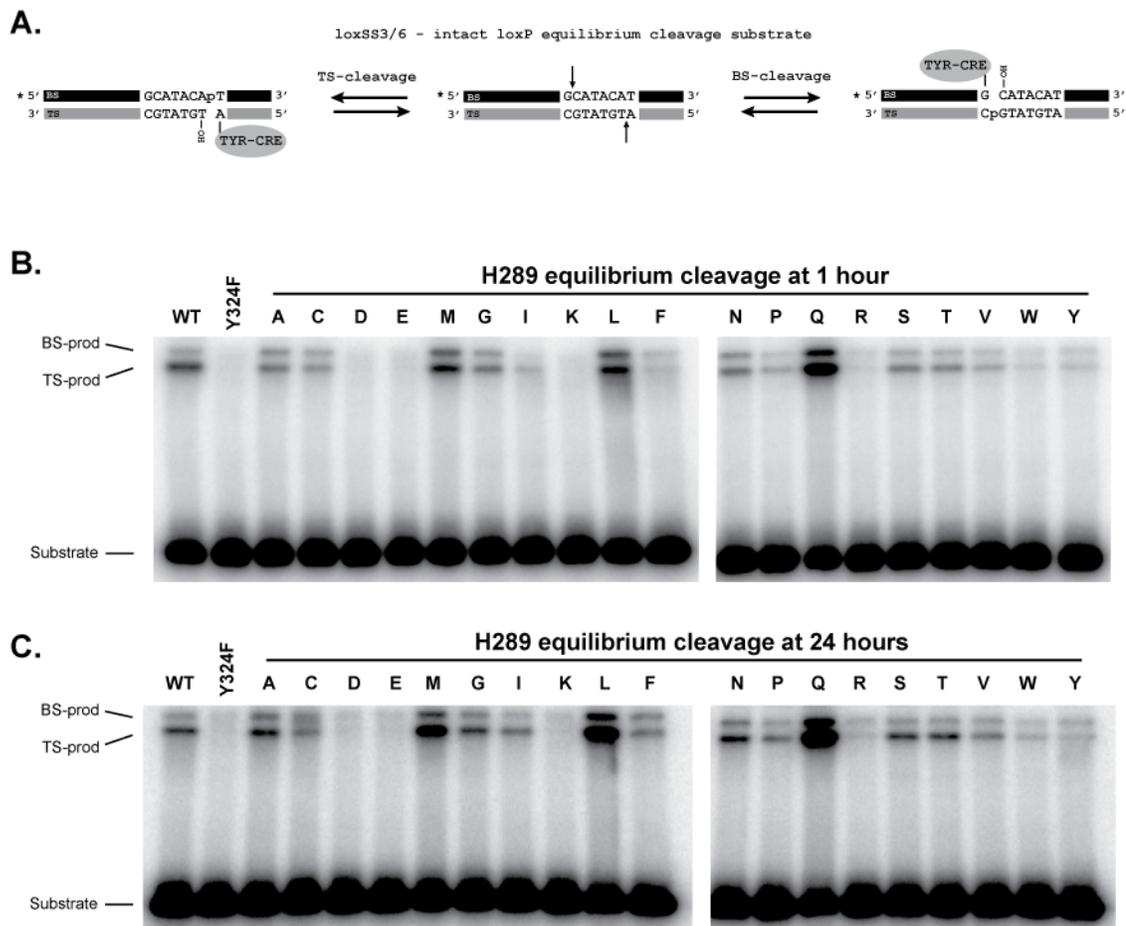
(B) BS-nicked suicide cleavage substrate loxSS4/5/6 was used to measure cleavage of purified H289 mutants at 1 hour  
 (C) BS-nicked suicide cleavage substrate loxSS4/5/6 was used to measure cleavage of purified H289 mutants at 24 hours.

Some H289 mutants accumulate cleavage product preferentially on the intact strand corresponding to TS cleavage. WT Cre hydrolyzes the BS-phosphotyrosine intermediate, which runs as unique band above unreacted substrate. H289C is the only mutant capable of detectable hydrolysis after 24 hours.

### 3.2.3 H289 mutants in equilibrium cleavage

Accumulation of covalent intermediate on a natural *loxP* site offers a way to indirectly measure  $K_{eq}$  and assess the balance of cleavage and ligation. A duplex 64 bp *loxP* substrate (the same used to measure binding (loxSS3/6)), was used to assay the accumulation of covalent intermediate in a method similar to that employed for suicide cleavage at 1 hour and 24 hours for all H289 mutants (Fig. 3.3). Since the substrate is an intact *loxP* site, all intermediates in the recombination pathway are accessible. TS and BS cleavage products are able to be ligated back to substrate or exchanged to form HJ intermediate and then resolved. However, the starting substrate is identical to the product of a complete recombination cycle. The 64 bp substrate is asymmetric, containing flanking DNA of 10 bp and 20 bp on either side of the *loxP* site. The asymmetry changes the location of the cleavage sites for TS and BS relative to the DNA such that covalent TS product is able to be resolved from covalent BS product on an SDS-PAGE Gel.

All mutants at H289 that were capable of cleavage had detectable levels of covalent intermediate with WT Cre generating approximately 10%. Most H289 mutants generated covalent product at levels similar to WT at equilibrium. However, the Met, Leu and Gln mutants accumulated 3-4 times more covalent intermediate than wild type. This result supports a defect in ligation and partially explains the accumulation of intact TS cleavage in the suicide cleavage assay. WT Cre and H289 mutants active on the *loxP* substrate all appeared to accumulate more TS covalent complex than BS covalent complex. Since the substrate allows used in this assay allows the entire recombination reaction to be sampled, it is impossible to assign the observed strand bias in equilibrium cleavage to a higher  $K_{eq}$  for TS or BS.



**Figure 3.3** *Equilibrium cleavage of intact loxP at 1 hour and 24 hours.*

(A) Schematic showing cleavage products generated by Cre with the intact *loxP* substrate, loxSS3/6.

(B) Covalent product accumulation with intact *loxP* by WT Cre and H289 mutants at 1 hour.

(C) Covalent product accumulation with intact *loxP* by WT Cre and H289 mutants at 24 hours.

Reactions at equilibrium (24 hours) are reflective of the relative rates of cleavage and ligation. H289M, H289L and H289Q accumulate significant covalent product relative to WT Cre, while other mutants do not.

### **3.2.4 H289 mutant summary**

An amino acid that is branched at C $\beta$  is more defective than one branched at C $\gamma$  or not branched at all. Leu and Met are more active in recombination and cleavage than Ile and Val; Ser is more active than Thr and Val. Aromatic substitutions are deficient or inactive, presumably due to steric bulk. Ala and Gly are modestly active *in vitro*, possibly due to bound water as observed in the TopIB-DNA complex structures (Davies *et al.*, 2006; Perry *et al.*, 2010). The deficiency in cleavage observed for H289A and H289G suggest that the histidine may have a role in catalysis during phosphoryl transfer. The H289Q mutant is the most active of all H289 substitution in cleavage and recombination, suggesting the histidine is dispensable in acid/base catalysis but the size and hydrogen bonding potential of histidine are important for Cre activity.

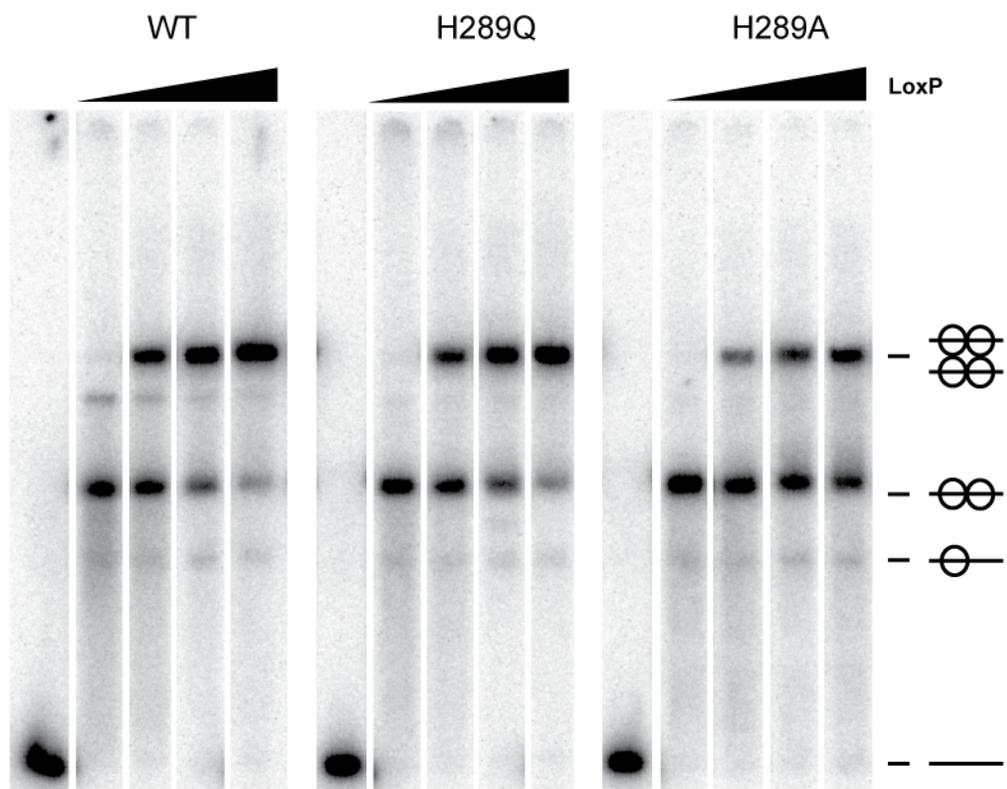
## **3.3 Detailed investigation of H289Q and H289A**

### **3.3.1 Synopsis of H289A and H289Q**

Initial binding studies with all nineteen mutants at H289 indicated that both H289A and H289Q were competent for binding and synopsis. However, the preliminary measurements were performed well below the  $K_D$  for synopsis (Ghosh *et al.*, 2007). To better measure the synopsis ability of H289Q and H289A, synopsis was assayed at higher concentrations of *loxP* and in conditions where Cre is saturating (1  $\mu$ M). As Fig. 3.4 shows, H289Q forms synaptic complexes nearly as well as WT Cre, but H289A appears to be as much as 4-fold deficient in synopsis. This result suggests that the histidine in Cre provides favorable van der Waals surface in addition to the hydrogen bonding interactions with the scissile phosphate, which Gln is able to mimic, but Ala does not. H289Q and H289A were among the more competent substitutions at 289 for synopsis, indicating that many of the other H289 mutants are more deficient than the H289A mutant.

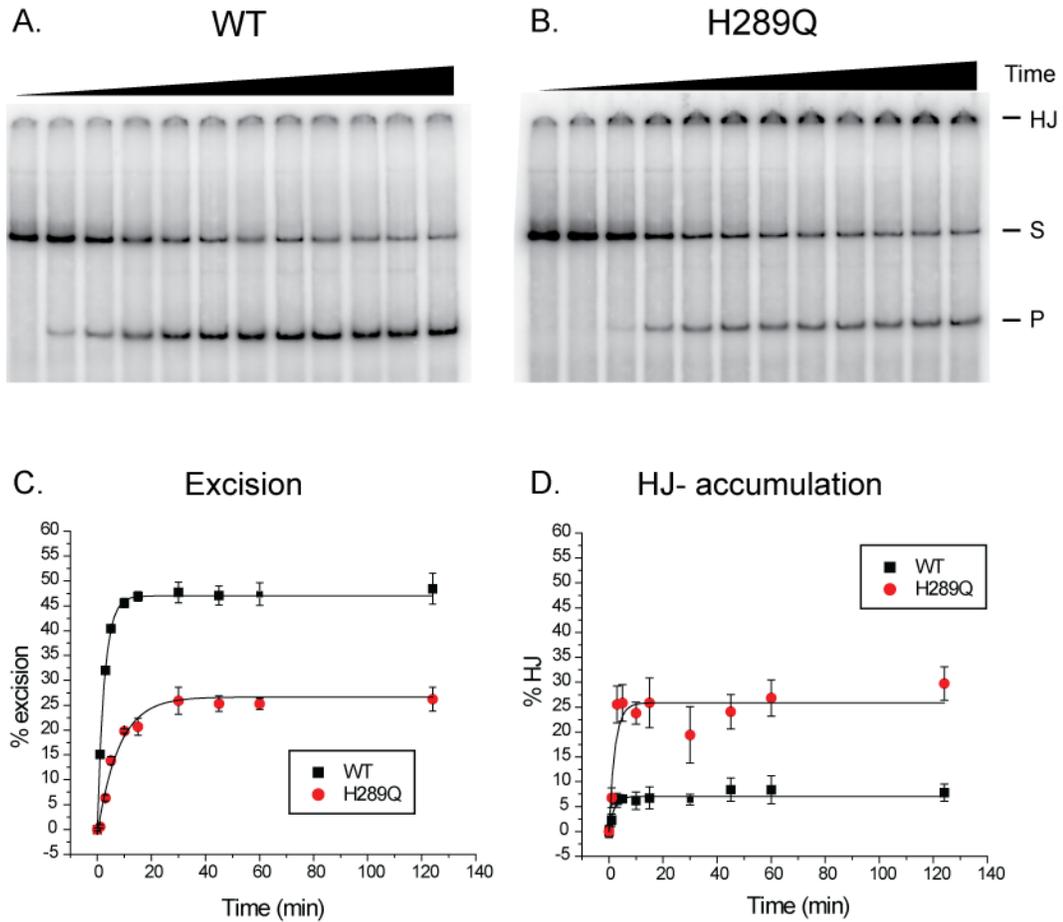
### 3.3.2 Rate of *in vitro* recombination for H289Q

Time course reactions containing 0.2 nM  $^{32}\text{P}$ -labeled recombination substrate in NCB assay buffer were initiated by addition of 50 nM Cre and time points were quenched in phenol/chloroform. Figs. 3.5 A and B show representative gels of *in vitro* recombination for WT Cre and H289Q. Graphs of product vs. time as well as HJ accumulation vs. time are shown in Figs. 3.5 C and D. The data fit well to a single exponential where the rates of excision for WT and H289Q were  $0.0065 \text{ sec}^{-1}$  and  $0.002 \text{ sec}^{-1}$  (see table 3.2). The minor decrease in recombination rate for H289Q is not consistent with the histidine functioning in acid/base catalysis in Cre as has been suggested for H305 in Flp (Whiteson *et al.*, 2007). However, H289Q did not reach the same maximum extent of excision as WT, indicating that the equilibrium for the recombination reaction was different. A significant buildup of HJ intermediate was observed for H289Q and H289A (not shown) suggesting that H289 may have a role in catalytic steps unique to recombination such as efficient isomerization or activation of Cre in the Holiday intermediate. The accumulation of HJ intermediate for Cre H289A and H289N had been noted previously (Gelato *et al.*, 2005; Martin *et al.*, 2002), providing additional evidence of the role of H289 in HJ resolution.



**Figure 3.4** *Synthesis of WT, H289Q and H289A*

The first lane of each series contains substrate alone. Cre at 1  $\mu$ M to saturate binding to loxP. Synapsis measured at 0.2, 10, 40, and 160 nM total loxP substrate. H289Q is not deficient, but H289A is moderately deficient in synapsis.



**Figure 3.5** *In vitro* recombination time course

Cre mediated excision of 234 bp linear product from 602 bp substrate as a function of time for WT (A) and H289Q (B).

C. Graph of excision for WT and H289Q fit to single exponential (black line).

D. Graph of HJ intermediate formation as a function of time. H289Q accumulates HJ intermediate in greater amounts than WT.

*Table 3.2: In vitro recombination rates for WT Cre and H289Q*

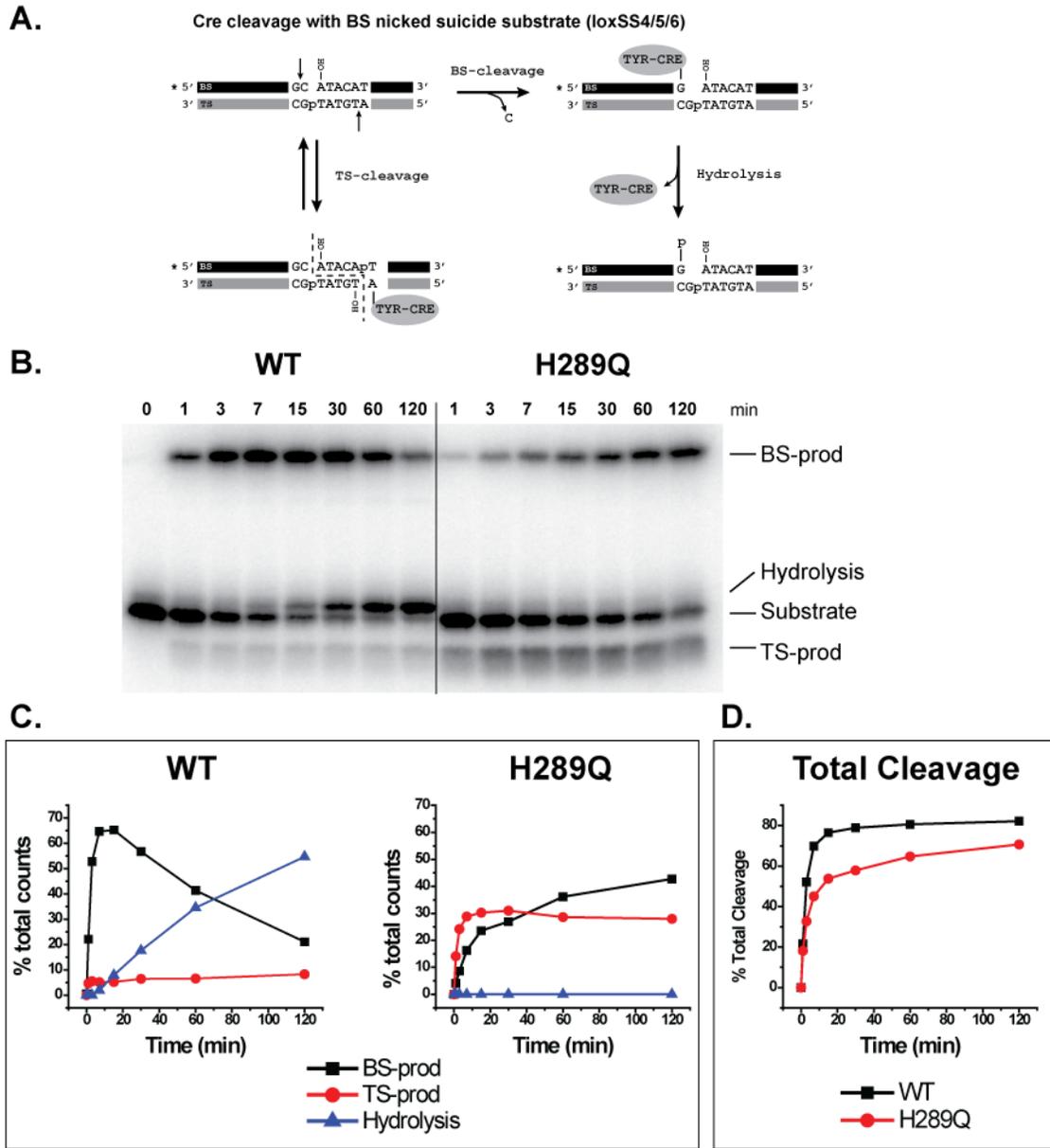
	WT	error	H289Q	error
Excision	0.39	0.003	0.12	0.021
HJ formation	0.53	0.102	0.45	0.114

Rates reported as (sec<sup>-1</sup>)

### 3.3.3 pH rate profile of Cleavage

Results from the panel of H289 mutants in recombination and cleavage suggest that the activation of Y324 does not require assistance from H289. If H289 in Cre functions in a rate limiting step in acid/base catalysis at cleavage, then the H289Q mutant should have a significantly different cleavage rate profile than WT Cre as a function of pH. The pH profile of cleavage for WT Cre and H289Q was measured using a suicide *loxP* substrate containing a nick on the bottom strand (*loxSS4/5/6*). Cleavage at the nick on the BS of *loxSS4/5/6* is irreversible, but cleavage on the intact TS is not. At physiological pH, WT Cre efficiently ligates any cleavage products on the top strand and rapidly accumulates trapped BS-covalent complex. The trapped BS-covalent complex is a substrate for water mediated hydrolysis of the phosphotyrosine linkage, depleting the covalent product (Fig. 3.6).

The initial velocities of cleavage for Cre WT and H289Q were measured in triplicate with the BS-nicked substrate under conditions favoring synapsis (50nM substrate) from pH 5.0 to 9.5 (Fig. 3.7). No hydrolysis product was observed in the time interval of the reaction. A plot of log initial velocity of total cleavage versus pH is shown in Fig. 3.7. The shape of the pH profile for WT is similar to that of H289Q, which suggests that histidine is not functioning in a rate-limiting step of cleavage. This is inconsistent with H289 having an essential role as a general base in Cre. cleavage. The pH dependence of Cre may reflect some other pH sensitive step in cleavage other than phosphoryl transfer (Ghosh *et al.*, 2007). Sedimentation equilibrium (Ghosh *et al.*, 2007) and velocity (data not shown) experiments show that Cre is capable of binding *loxP* over the entire pH range tested, but synapsis is disrupted at high pH.

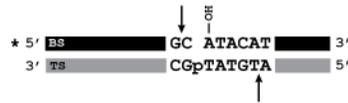


**Figure 3.6** Representative cleavage on BS-nicked substrate at pH 7.5

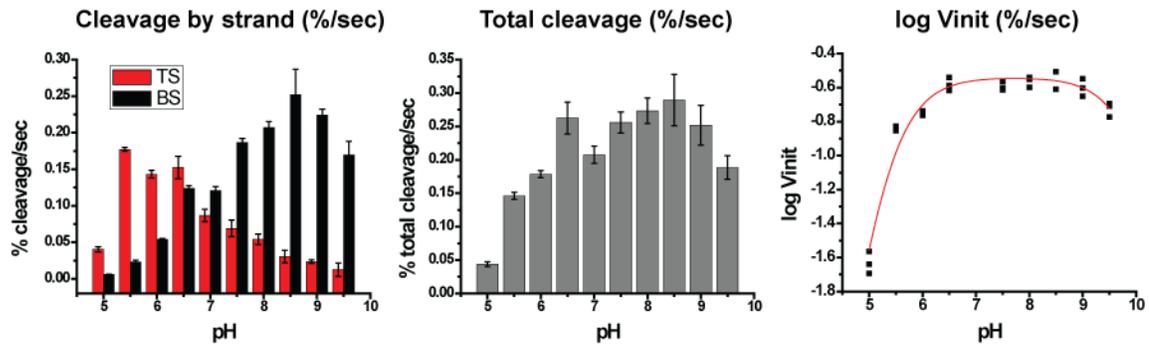
(A) Schematic of Cre cleavage products generated with BS-nicked *loxP* suicide substrate loxSS4/5/6. (B) Representative gel of WT Cre and Cre H289Q (500 nM) cleavage with loxSS4/5/6 (50 nM) as a function of time at pH 7.5. (C) Plots of bottom strand (BS), top strand (TS), and hydrolysis product for WT and H289Q as a function of time. (D) Plot of total cleavage, which is the sum of TS, BS and Hydrolysis products for WT Cre and H289Q as a function of time.

Cre can cleave the BS-nicked substrate (loxSS4/5/6) irreversibly on the BS to trap covalent intermediate (BS-product), or cleave the intact TS, which is able to be ligated back to starting substrate. The TS cleavage results in a double strand break since the substrate already contains a nick on the BS. The labeled TS-product, does not contain covalently bound Cre, but is half of the original substrate. The trapped BS-covalent product is a substrate for water mediated hydrolysis forming a band with slower mobility than the unreacted substrate. At pH 7.5, hydrolysis of BS-product for WT is readily apparent for WT, but absent for H289Q. TS-product accumulates more rapidly and to a greater extent for H289Q than for WT Cre.

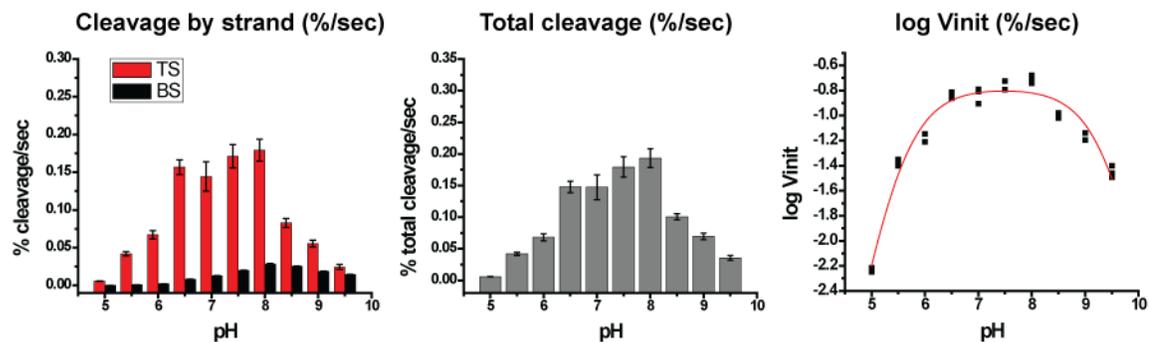
**A. BS nicked suicide substrate (loxSS4/5/6)**



**B. WT**



**C. H289Q**



**Figure 3.7 Initial Velocity of Cre Cleavage vs. pH**

(A) Diagram of BS-nicked substrate, loxSS4/5/6.

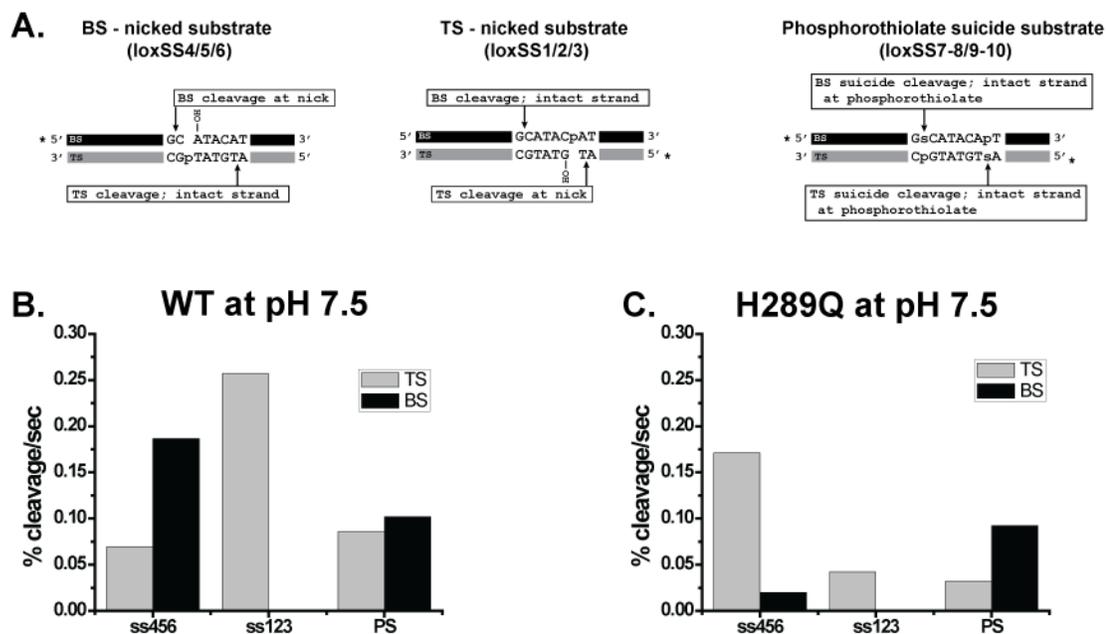
(B and C) Plots of initial velocity vs. pH for WT Cre (B) and Cre H289Q (C).

The profiles of total cleavage (center) for WT and H289Q look similar, however the relative proportions of TS and BS product (left) contributing to total cleavage differ substantially between the WT Cre and H289Q. WT is dominated by TS cleavage at low pH and shifts to primarily BS cleavage above pH 7. H289Q is dominated by TS cleavage over the entire pH range, with BS-product contributing only minimally to the initial velocity at the higher pH values tested. The log Vinit of total cleavage (right) was fit to equation  $\log V_{init} = \log V_{init(max)} - \log (1 + 10^{-pK_{a1}}/10^{-pH} + 10^{-pH}/10^{-pK_{a2}})$  (Stivers, Shuman & Mildvan, 1994).

For WT Cre, the TS cleavage rate drops as a function of pH while BS cleavage increases as a function of pH and the two cross at pH 6.5. H289Q is dominated by TS cleavage through pH 8.0 after which the rate rapidly decreases, and is less efficient at cleaving the BS with only a small increase in rate at higher pH values. Cleavage pH profiles were also determined for additional suicide substrates where the nick was moved to the TS (loxSS1/2/3) or the *loxP* substrate was intact without a nick using 5'-bridging phosphorothiolate substrates (loxSS7-8/9-10). The pH profile of total cleavage with these substrates was similar to that observed in Fig. 3.7 with loxSS4/5/6, but the relative proportions of TS and BS cleavage were different and are discussed below.

### **3.3.4 H289Q strand cleavage preference**

In the synaptic complex, *loxP* is bent asymmetrically at one of the two cleavage sites, where the site opposite the bend is active for cleavage (Guo *et al.*, 1997). The TS cleavage preference observed for H289Q with loxSS4/5/6 could be a symptom of alternative bending where H289Q is better able to bend at the half site with the nick to activate the other half site for cleavage. If H289Q promotes an alternative bend, moving the nick to the other half site should reverse the strand preference for H289Q. Alternatively, if H289Q preferentially cleaves the TS, then cleavage on a substrate with the nick on the TS (loxSS1/2/3) should improve cleavage since it is irreversible. WT Cre and H289Q were incubated for 60 minutes with substrate containing a BS-nick (loxSS4/5/6), TS-nick (loxSS1/2/3), or no nick (5' bridging phosphorothiolate DNA; loxSS7-8/9-10) (Fig. 3.8). WT Cre efficiently cleaves either substrate on the strand with the nick, however H289Q preferentially cleaves the TS with both nicked substrates, but is cleaves the TS better when the nick is on the BS. WT Cre and H289Q cleave more BS than TS with the phosphorothioate substrate, which is consistent with previous findings, which used the same substrate (Ghosh *et al.*, 2005a).



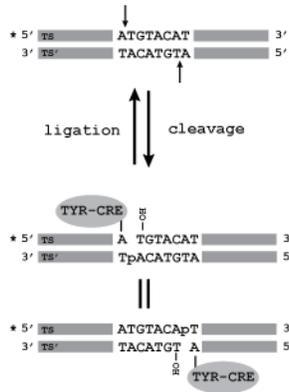
**Figure 3.8** *Representative data for strand cleavage preference*

(A) Cartoon representation of three cleavage substrates tested. LoxSS4/5/6 contains a BS-nick with intact TS, while loxSS1/2/3 is the opposite, containing a TS-nick and intact BS. The phosphorothiolate substrate is an intact duplex with sulfur instead of oxygen at the 5' position of the scissile phosphate. (B and C) Initial velocities for cleavage with each substrate on TS and BS at pH 7.5 for WT (B) and H289Q (C). Notice the strong reversal in cleavage preference on the loxSS4/5/6 (BS-nick) substrate for WT and H289Q. When the nick is moved to the TS (loxSS1/2/3), WT cleaves robustly, but H289Q is deficient. Both WT and H289Q accumulate very little intact BS cleavage product with loxSS1/2/3. With the phosphorothiolate substrate (PS), BS cleavage is preferred over TS for both WT Cre and H289Q.

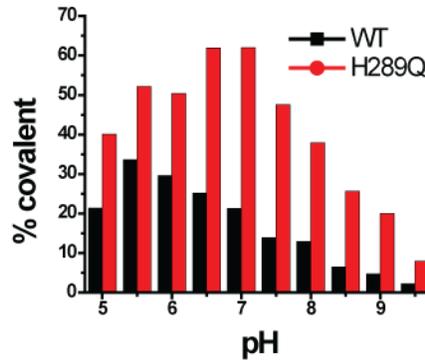
### 3.3.5 pH profile of equilibrium cleavage

To determine if the pH dependence of intact TS cleavage is due to a shift in the equilibrium cleavage rate constant ( $K_{cl}$ ), to favor the ligated state rather than a DNA bending effect, covalent intermediate was measured for WT and H289Q on a symmetrized and intact *loxP* site containing two AT half sites (*loxS2*) at equilibrium. Since both half sites are identical, any strand bias is eliminated. WT Cre and H289Q were incubated with *loxS2* substrate in different pH buffers for extended time intervals to allow the system to reach equilibrium. Fig. 3.9A shows a plot of covalent product versus pH at equilibrium for WT and H289Q with substrate *loxS2*. The shape of the pH profile is different for WT Cre and Cre H289Q. Covalent intermediate with WT Cre reaches a maximum of 35% at pH 5.5 and steadily declines with increasing pH. H289Q accumulates more covalent intermediate than WT Cre with a maximum of 62% at pH 6.5 or 7.0, and decreases sharply at higher pH. Since the rate of cleavage by WT Cre does not display a corresponding drop in cleavage activity with pH, the drop in equilibrium covalent product at higher pH values is likely due to an increase in the rate of ligation. The shape of the plot in Fig. 3.9A looks nearly identical to the shape of the initial velocities of TS cleavage determined with *loxSS4/5/6* for WT and H289Q in Fig. 3.7. It is unfair to compare initial velocity with covalent intermediate at equilibrium. However, cleavage of the intact strand in time course reactions with *loxSS4/5/6* and *loxSS1/2/3* did reach equilibrium (Fig. 3.10 for an example) and is plotted in Fig. 3.9B as a function of pH for H289Q. Though the extent of covalent product is different, the trend in covalent intermediate as a function of pH is the same, showing a decrease with pH after a maximum at pH 6.5 suggesting the drop of intact strand cleavage at high pH on nicked substrates is due to a pH dependent increase in ligation relative to the rate of cleavage. WT Cre does not accumulate significant intact strand product on nicked substrates above pH 6.

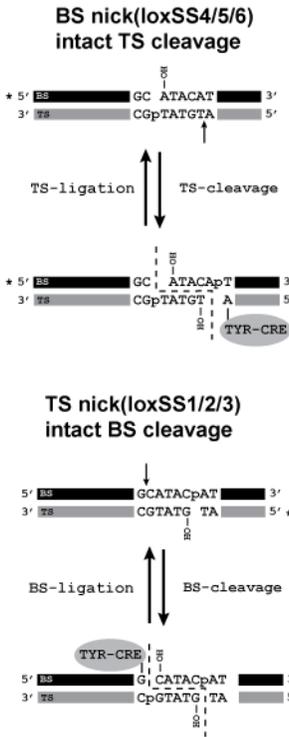
**A. Symmetric TS substrate (loxS1)**



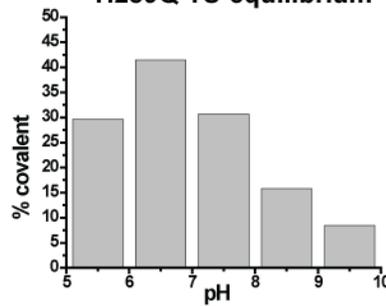
**Equilibrium cleavage vs. pH**



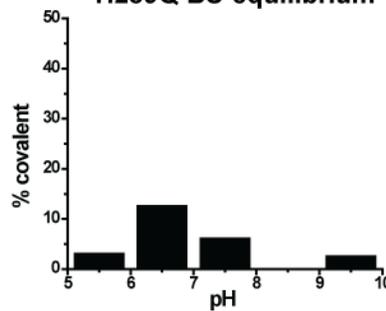
**B. H289Q covalent intermediate from cleavage on intact strand of nicked suicide substrates**



**H289Q TS-equilibrium**



**H289Q BS-equilibrium**



**Figure 3.9 Equilibrium cleavage as a function of pH**

A. Cleavage of symmetric substrate loxS2 containing TS halfsites was measured at equilibrium at pH 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0 and 9.5 for WT and H289Q. H289Q accumulates more covalent intermediate and the shape of the curve is shifted to the right where the maximum product is at 6.5 and 7.0 where the maximum for WT is at 5.5.

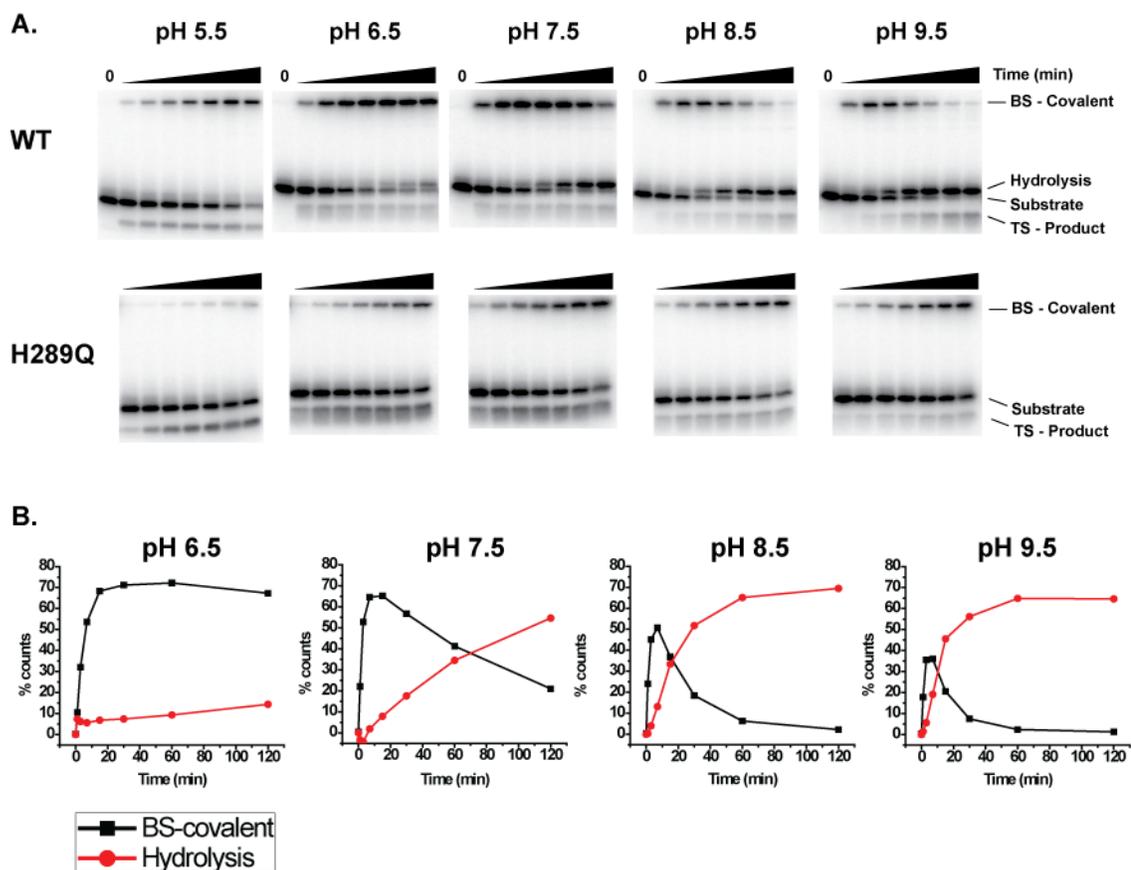
B. Cleavage of the intact strand of nicked substrates loxSS4/5/6 (intact TS) and loxSS1/2/3 (intact BS). Covalent product plotted for H289Q at or near equilibrium for intact strand cleavage.

### **3.3.6 pH Dependence of Hydrolysis**

Water mediated hydrolysis of the trapped covalent intermediate was observed in cleavage reactions with all three suicide substrates and increased with pH. This reaction is shown schematically in Fig. 3.6 for loxss4/5/6 where the trapped BS-covalent complex disappears as a function of time. Fig. 3.10 shows the effect of pH on the rate of hydrolysis catalyzed by WT Cre compared to H289Q.

The rate of hydrolysis in the Cre cleavage assays is dependent on the amount of cleavage product. However, a conservative measure of the hydrolysis rate can be attained by determining the slope of the linear portion of the hydrolysis product formation curve. As shown in Fig. 3.10, at pH 9.5 the rate of hydrolysis of the covalent intermediate is linear between the 3 and 15 minute intervals during the cleavage time course with loxSS4/5/6. The observed rate determined from the slope during this interval was 0.055% hydrolyzed/sec. This conservative estimate suggests that Cre mediated hydrolysis at pH 9.5 is more than 600 times faster than topoisomerase mediated hydrolysis at the same pH (Petersen & Shuman, 1997a).

H289Q does not catalyze the hydrolysis of covalent intermediate at any pH with any of the suicide substrates that were tested. None of the H289 substitutions were capable of hydrolyzing the covalent intermediate after 24 hours of incubation, where WT Cre had converted nearly 100% of substrate into hydrolysis product (Fig. 3.2). This suggests that H289 plays a critical role in promoting hydrolysis. The increased rate of hydrolysis by Cre compared to TopIB further implicates histidine in hydrolysis, especially since TopIB is faster in cleavage and ligation chemistry (Krogh & Shuman, 2000). H289 in Cre may therefore have a more important role as a proton donor to tyrosine in hydrolysis than in ligation.



**Figure 3.10** *pH profile of cleavage and hydrolysis timecourse*

A. Cleavage reactions measured at 0,1,3,7,15,30,60 and 120 minutes at pH 5.5, 6.5, 7.5, 8.5 and 9.5 for WT and H289Q on loxSS4/5/6 (BS-nicked) suicide substrate in conditions favoring synapsis (50nM total substrate). BS suicide cleavage accumulates over time, but is susceptible to hydrolysis at high pH. H289Q does not hydrolyze the BS cleavage product.

B. Graph of WT BS-covalent product and hydrolysis as a function of time. pH 5.5 not included since no hydrolysis is observed.

### 3.3.7 Activities of H289Q and H289A in ligation

Results of equilibrium cleavage suggested that some mutants at H289, including glutamine, are partially defective in ligation since they accumulate large amounts of covalent intermediate. Ligation activity has been measured directly, independent of cleavage in Cre, vaccinia topoisomerase (Woodfield, Cheng, Shuman & Burgin, 2000) and Flp (Whiteson *et al.*, 2007) by using a synthetically modified DNA substrate containing a tyrosine analog attached to the 3' end of a scissile phosphate. The tyrosine analog pNP (para-nitro-phenol) with a pKa of 7.1 was esterified to the 3' phosphate of either a TS or BS substrate to mimic an activated tyrosyl leaving group during ligation. The modified DNA substrate was radiolabeled and annealed to complementary and adjacent strands, forming a three stranded *loxP* substrate containing a nick on either the TS or BS at the scissile phosphate as depicted in Fig. 3.11. Cre catalyzes the ligation of the two adjacent strands by release of the pNP without the need for Y324 (this work) (Woodfield *et al.*, 2000). Fig. 3.11 shows ligation of Cre mutants using TS and BS pNP substrates at pH 7.5. At this pH, ligation does not require a proton donor to assist the pNP leaving group. The 60 minute reactions were quenched with SDS and digested with proteinase K to remove covalently attached Cre and then separated by denaturing PAGE. The unligated modified 36mer (BS) or 26mer (TS) substrate containing pNP is a single band with greater mobility than the ligated 64mer *loxP* product. A second band appears above the unreacted TS substrate for mutants that contain Y324 but not those containing the Y324F mutation indicating that this species is the proteinase K digested product of a phosphotyrosine covalent intermediate. The product of the ligation reaction is an intact *loxP* site, so active Cre mutants are able to cleave the ligated *loxP* product and generate covalent intermediate. However, K201A, which is inactive for cleavage, accumulates significant amounts of covalent intermediate with the ligation substrates (data not shown), which suggests that this species is either not covalent intermediate or is generated by some other mechanism.

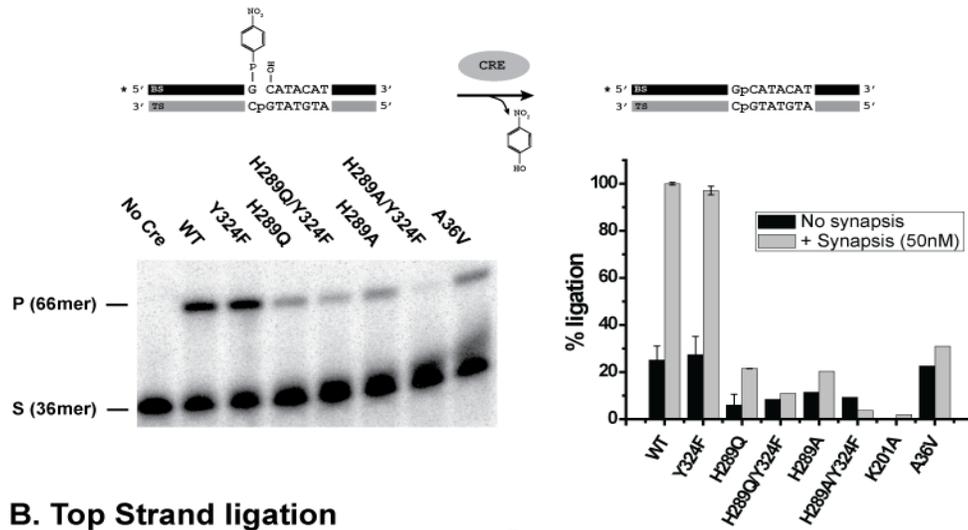
The observed covalent intermediate could be due to the direct attack of the pNP by tyrosine (Fig. 3.11B). If pNP assumes a position in the active site similar to O5' of an intact *loxP* site, then tyrosine is able to attack the pNP directly in a mechanism similar to cleavage. K201A is able to generate covalent intermediate because the pKa of pNP is much lower than 5'-OH, and as has been determined with 5'-bridging phosphorothiolate substrates (Ghosh *et al.*, 2005a; Krogh & Shuman, 2000), the leaving group does not require a proton donor. K201A does not generate any appreciable ligated *loxP* product, consistent with its role in activating the incoming 5'-OH. Any Cre mutants containing Y324 could exploit this secondary pathway to ligated products complicating analysis. For this reason, any mutants studied by this assay should also mutate Y324.

WT Cre and a selection of mutants were tested for BS and TS ligation of pNP modified substrates under conditions unfavorable (0.2 nM substrate) and favorable (50 nM substrate) for synapsis formation. Ligation product was normalized for each mutant to the maximum value measured with each substrate. The data shown in Fig. 3.11 is consistent with strand preference observations made for Cre cleavage, where WT Cre and Y324F are highly active on the TS half site in the absence of synapsis, and only become active at the BS half site in conditions favoring synapsis (Ghosh *et al.*, 2005a). Congruent with this observation, the A36V mutant, which is defective in synapsis is robust at ligation of the TS and deficient in BS ligation at both substrate concentrations. In general, Cre appears less active on the BS ligation substrate than on the TS substrate. However, time course reactions comparing both substrates demonstrate that ligation is nearly complete at sixty minutes, indicating that 100% conversion to product will not occur on the BS substrate (data not shown). Perhaps an artifact during synthesis of the BS-pNP generated a fraction of unreactive substrate.

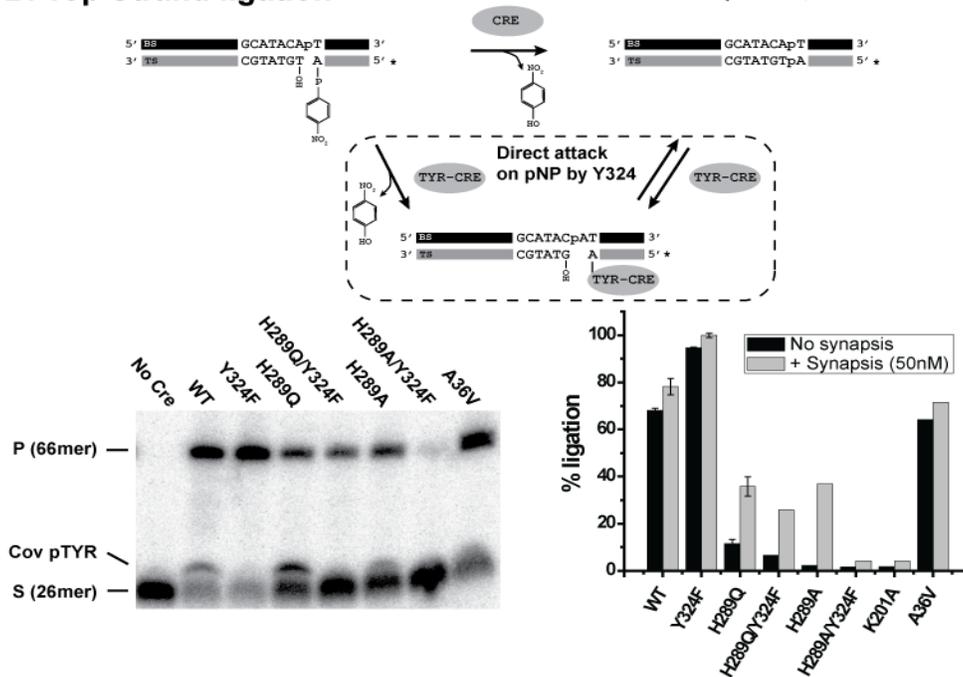
H289Q also appears to be more active on the TS ligation substrate than on the BS ligation substrate. If the H289Q generated covalent intermediate observed in the TS reaction is from ligation product that was then a substrate for H289Q cleavage, then the actual extent of ligation

for H289Q is nearly double that reported in Fig. 3.11. However, since the H289Q/Y324F double mutant does not generate this additional ligation product, either Y324 is aiding ligation or the covalent intermediate generated by direct attack on pNP by Y324. Although H289Q and H289Q/Y324F have similar extents of ligation, the H289A and H289A/Y324F mutants are significantly different. The H289A mutant generates ligation product on the order of H289Q, however the H289A/Y324F double mutant abolishes ligation activity. Perhaps the hydroxyl group of Y324 aids in stabilization of phosphate in the absence of H289. Alternatively, H289A may have greater ligation activity on the covalent intermediate than with the pNP substrate. The disruption of ligation observed with H289A and H289Q, particularly the double mutants with Y324F, support a role for H289 in ligation. However, since pNP does not require an acid to donate a proton at physiological pH, these experiments are unable to determine if H289 is acting as a proton donor in ligation.

## A. Bottom strand ligation



## B. Top Strand ligation



**Figure 3.11** Ligation of pNP modified TS and BS loxP DNA substrates by Cre

A. Bottom strand ligation of labeled 36mer DNA substrate (S) containing a 3'pNP, to an adjacent unlabeled 30mer to form a full 66mer loxP BS (P). B. Top strand ligation of labeled 25mer DNA substrate (S) containing 3' pNP, to an adjacent 40mer to form a full 66mer loxP TS (P). Mutants containing Y324 were found to accumulate Cre covalent intermediate via a 3'-phosphotyrosine linkage, which when digested with proteinase K runs just above the substrate on the gel (Cov pTYR). Covalent intermediate can form via two distinct mechanisms. The first involves direct attack of pNP phosphate by Y324 and the second involves Y324 mediated cleavage of the intact loxP product resulting from ligation. (A and B) Reactions were performed in conditions favorable (50nM substrate) and unfavorable for synapsis (0.2 nM substrate). The 60 minute reactions were quenched and separated by UREA Page. Reactions for WT, Y324F and H289Q are the averaged value of three experiments with associated standard deviation.

### 3.4 Discussion

Fifteen amino acid substitutions at 289 are tolerated by Cre as measured by recombination *in vivo* and *in vitro*, which prompted further study to understand the function and reason for conservation in YRs. The cleavage activity of the purified H289 mutants correlated well with the observed *in vitro* recombination activity. Studies of this histidine residue in YRs have suggested a role in ligation that is primarily based on the accumulation of covalent intermediate by the H305L in Flp (Pan *et al.*, 1993b; Parsons *et al.*, 1988; Serre & Jayaram, 1992), H240L in XerC and H244L in XerD (Arciszewska *et al.*, 2000; Cornet *et al.*, 1997), and H234L in Tnpl (Vanhooff *et al.*, 2009). The ligation deficiency of H305Q in Flp was observed directly using synthetic substrates (Whiteson *et al.*, 2007). All nineteen H289 mutants in Cre were studied for cleavage of an intact *loxP* site under equilibrium conditions. The Gln, Leu and Met mutants all accumulated covalent intermediate at levels above WT, but this does not appear to be a general property, as the other H289 mutants accumulated covalent intermediate at levels similar to or less than WT Cre. Accumulation of covalent intermediate under equilibrium conditions suggests that the rate of cleavage has increased relative to the rate of ligation. Since cleavage by H289Q, H289L and H289M is slower than WT Cre, a defect in ligation is mostly likely responsible for the buildup of covalent intermediate. These three mutants are the most competent in suicide cleavage, but several other mutants that do not accumulate covalent intermediate such as the alanine and asparagine are nearly as active.

Leu, Met and Gln are good structural mimics of histidine. Amino acid substitutions at 289 with branch points at C $\beta$  are disruptive to activity. This is evident by comparing the activities of Leu, Ile and Val as well as Ser and Thr. Modeling suggests that smaller amino acids such as alanine, asparagine and serine leave enough space to accommodate a water molecule in a similar position as N $\epsilon$  of histidine as observed in transition state structure of TopIB (Davies *et al.*, 2006; Perry *et al.*, 2010). H289 may participate in ligation, but solvent can partially or fully substitute in many H289 mutants, masking potentially significant defects. Leu, Met and Gln fill the

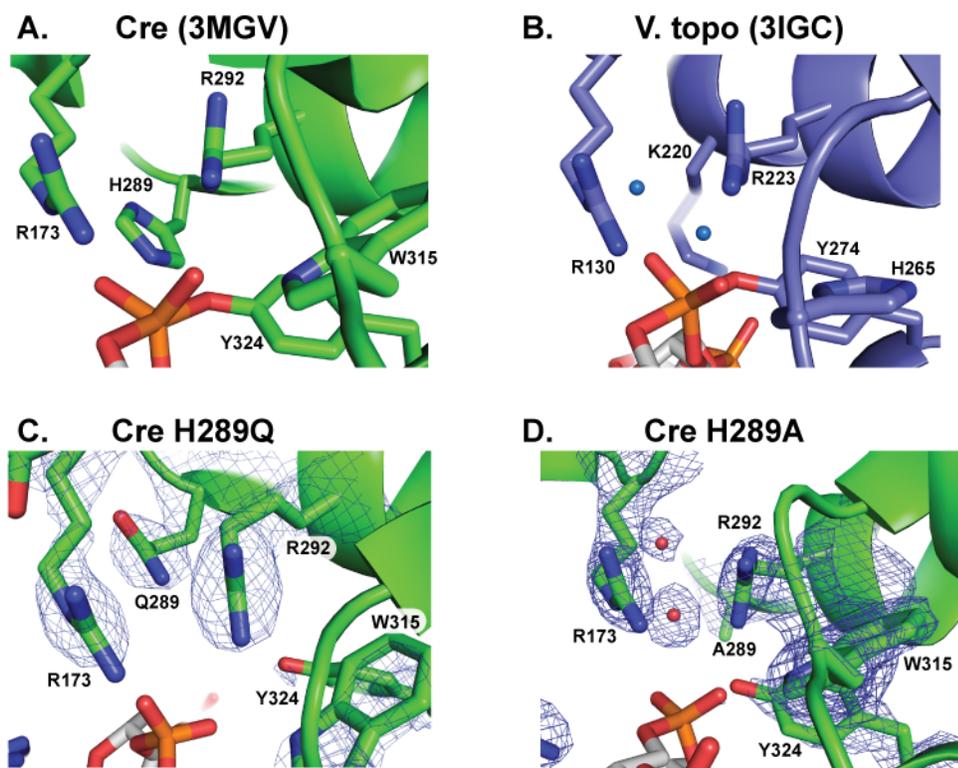
space occupied by histidine and may prevent solvent from efficiently assisting in ligation. Structures of the H289A and H289Q mutants done in collaboration with Kushol Gupta in the Van Duyn group are consistent with this hypothesis and are shown in Fig. 3.12. with statistics presented in Table. 3.3. Clear density for a well ordered water molecule in H289A is observed in the same space as the histidine in the active subunit. As expected, glutamine occupies the same general space as histidine, capable of providing stabilizing hydrogen bonds to the scissile phosphate, but unable to participate in acid/base catalysis. The density for the glutamine is well-defined, with no evidence of density corresponding to solvent in this region.

Despite a possible defect in ligation, H289Q is the most active of all the histidine mutants, which underscores the importance of stabilizing hydrogen bonds to the scissile phosphate. The pH profiles examining cleavage demonstrated that assistance of a general base is not required to activate the tyrosine for cleavage. This may be due in part to the differences in the pKa of tyrosine (~10) compared to the 5'-hydroxyl group (~15), and is consistent with observations from model reactions of nucleophilic substitution at phosphodiester linkages that show a modest dependence on the basicity of the nucleophile ( $\beta_{\text{nuc}} \sim 0.2-0.3$ ) but a large dependence on the basicity of the leaving group ( $\beta_{\text{lg}} \sim -0.8$ ) (Cassano, Anderson & Harris, 2004; Kirby & Younas, 1970a; Kirby & Younas, 1970b; Maegley, Admiraal & Herschlag, 1996). Therefore, the dependence on a general acid to assist the leaving group O5' is more important than a base to aid in the activation of the tyrosine nucleophile. During ligation, the histidine in YRs or water in TopIB would serve as an acid to assist in protonating the tyrosine leaving group. Experiments using modified tyrosine peptides were used to argue that H305 in Flp recombinase is acting as a general base in cleavage and an acid in ligation (Whiteson *et al.*, 2007). In these experiments, Flp H305Q was rescued for cleavage using a modified flourotyrosine peptide with a lower pKa than tyrosine, however Flp H305A was not rescued. Histidine is perhaps the only amino acid that could provide an important contribution to stabilization of the transition state and also be able to participate in acid/base

catalysis. Although water can substitute in acid/base catalysis, other amino acids at this position are either unable to stabilize the transition state or sterically exclude water from effectively participating in catalysis.

### **3.4.1 Top Strand Bias**

H289Q preferentially cleaves the TS half site of nicked suicide substrates. This preference is more pronounced for cleavage of the intact TS when the nick was located at the BS half site than when the nick is on the TS half site. This strand bias appears is not observed with the phosphorothiolate substrate, which lacks a nick. Cleavage of the phosphorothiolate substrate by H289Q was similar to that determined for WT Cre, where BS cleavage is slightly preferred to TS cleavage under conditions favoring synapsis (Ghosh *et al.*, 2005a). This suggests that the H289Q mutant may have a defect in bending *loxP* associated with the activation of cleavage. However, if the defect was strictly a bending artifact, cleavage of the BS should be more pronounced with a nick on the TS. A bending defect would also likely be apparent in binding and synapsis assays, but H289Q is unperturbed. This apparent opposite strand bias has been noted previously in Cre for the H289A and H289N mutants, which preferentially form and resolve TS Holiday junction intermediates (Gelato *et al.*, 2005; Martin *et al.*, 2002). The H289A-DNA structure (Fig. 3.12) is clearly a trapped TS-covalent complex, which is suggestive of an altered strand bias. The mechanism of this altered strand bias is unclear, however other YR systems such as  $\lambda$ -int and XerCD do not rely on site asymmetry alone to control the direction and order of strand cleavage during recombination (Azaro & Landy, 2002; Barre & Sherratt, 2002). Therefore the strand bias observed in Cre is likely only relevant to a small portion of YRs and does not account for conservation of histidine in the active site of the vast majority of YRs.



**Figure 3.12 Comparison of structure at H289**

A. Cre transition state (pdb 3MGV). H289 makes interaction with O2P and phenolic oxygen of Y324.

B. Smallpox topoisomerase transition state (pdb 3IGC). K220, the equivalent residue is oriented away from the active site. Two well ordered water molecules replace the nitrogens of H289 in Cre making similar bonding interactions.

C. Structure of Cre H289Q mutant in complex with *loxP*.  $2F_o - 2F_c$  electron density is contoured at  $1.5 \sigma$ . The glutamine occupies the same space as H289.

D. Structure of Cre H289A mutant in complex with *loxP*.  $2F_o - 2F_c$  electron density is contoured at  $1.5 \sigma$ . Clear density for two water molecules, similar to those observed in TopIB (B) replace the histidine.

Table 3.3 Crystal statistics for H289Q and H289A loxP DNA complexes

Diffraction Data	Cre H289Q – LoxP	Cre H289A – LoxP
Resolution (Å)	2.6 Å	2.3 Å
Space Group	C222 <sub>1</sub>	P322 <sub>1</sub>
Unit Cell	a = 106.29 Å b = 122.29 Å c = 178.64 Å	a = b = 136.2 Å c = 217.45 Å
Mosaicity	1	0.46
Wavelength (Å)	1	1
Completeness (%)	95.2 (91.9)	94.2(87.6)
R <sub>merge</sub>	0.067(0.584)	0.06(0.55)
I/σ	12.8 (2.3)	23.4 (3.1)
Redundancy	4.5 (4.5)	2.5 (2.0)
Refinement		
R <sub>free</sub>	0.241(0.289)	0.234(0.284)
Average B factors (Å <sup>2</sup> )	49.82	54.36
Rmsd		
Bond lengths (Å)	0.008	0.01
Bond angles (°)	1.303	1.53

### 3.4.2 Hydrolysis

Another interesting observation was made from the pH rate cleavage experiments where WT Cre hydrolyzed the 3'-phosphotyrosine intermediate efficiently as a function of pH, while H289Q did not generate any hydrolyzed product. The pH dependence of hydrolysis was observed in TopIB (Christiansen, Knudsen & Westergaard, 1994; Petersen & Shuman, 1997a) and the increased rate of hydrolysis was presumed to be due to the hydroxide ion concentration, which increases in alkaline conditions. Unlike the rates of cleavage and ligation, where TopIB is faster than Cre, the rate of hydrolysis observed in Cre is faster than TopIB at pH 9.5. This agrees with published comparisons of hydrolysis measured at pH 7.5 in Cre, Flp and Vaccinia topoisomerase (Ma *et al.*, 2009a; Ma *et al.*, 2007; Ma *et al.*, 2009b; Tian, Claeboe, Hecht & Shuman, 2003).

In this published work, the rate of hydrolysis was measured on natural sites as well as modified DNA substrates containing methylphosphonate in place of either the S<sub>p</sub> or R<sub>p</sub> oxygen of the scissile phosphate. The uncharged methylphosphonate abrogates the need for charge stabilization on the corresponding non-bridging oxygens in the transition state. The rate of hydrolysis catalyzed by vaccinia virus topoisomerase increased 30,000 fold with the methylphosphonate DNA substrates (Tian *et al.*, 2003). However, a similar enhancement in the rate of hydrolysis with methylphosphonate substrates was not observed in Cre and Flp (Ma *et al.*, 2009a; Ma *et al.*, 2007; Ma *et al.*, 2009b). An explanation for this observed difference may involve H289. With the methylphosphonate DNA, R223 in TopIB (R292 in Cre) is not required to stabilize charge on scissile phosphate and is then able to adopt an alternative conformation where N<sub>η</sub>1 and N<sub>η</sub>2 are positioned in the equivalent space as N<sub>ε</sub> of H289 in Cre. R223 is more efficient in activating Y274 which results in an enhancement in the rate of hydrolysis. This mechanism is

unavailable to Cre (and Flp) since H289 prevents R292 from adopting a similar activating conformation. If this hypothesis is true, then Cre H289A should see a large rescue in hydrolysis with methylphosphonate DNA substrates.

The stimulation of water mediated hydrolysis may be a protective mechanism in YRs. A stalled reaction, trapped as a covalent intermediate is problematic for the integrity of the genome. Hydrolysis may be used as a means of fixing stalled reactions stuck as covalent intermediates. In the cell, the 3'-phosphate hydrolysis product can be readily dephosphorylated and ligated to restore intact DNA with host enzymes such as eukaryotic polynucleotide kinase (Karimi-Busheri, Lee, Tomkinson & Weinfeld, 1998). The faster rate of hydrolysis in YRs may reflect the complexity of recombination where there is a greater chance having a stalled reaction. Relaxation catalyzed by the more efficient TopIB enzymes is a more simple reaction and hydrolysis to scavenge stuck covalent intermediates is unnecessary.

### **3.4.3 H289 in ligation**

The mechanism of water mediated hydrolysis is similar to that of ligation, where the incoming 5'-OH is activated for attack on the 3'-phosphotyrosine intermediate. The results from ligation assays using pNP modified substrates demonstrated that H289Q and H289A were deficient in ligation at physiological pH. Similar to the results from cleavage assays, H289Q and H289A generated more TS ligation product than BS ligation product. It was also interesting to note that the Y324F mutant has ligation activity similar to WT Cre, but the Y324F/H289Q and Y324F/H289A mutants were significantly defective relative to the H289Q and H289A single mutants (Fig. 3.11). It is possible that the hydroxyl group of tyrosine contributes a stabilizing interaction to the transition state of the pNP ligation reaction. In Flp, the H305Q/Y343F mutant was not as severely defective in ligation with the pNP substrate, but was defective in ligation with

a substrate containing cresol, which more closely mimics a covalently attached tyrosine with a pKa of 10 and is more dependent on a proton donor to assist the cresol leaving group (Whiteson *et al.*, 2007). These results support a role for H289 in ligation.

### 3.4.4 Summary

H289 may not have a single most important function to explain its conservation in YR active sites. Although fourteen mutants at 289 retain recombination activity, some of these mutants are significantly defective at recombination and cleavage. Examination of these mutants and detailed experiments primarily with the H289Q mutant suggests that H289 has a role in structural support of the active site, stabilization of the transition state through hydrogen bonding interactions, but is not important as a proton acceptor to activate Y324 in cleavage. H289 has a significant role in ligation and is essential for water mediated hydrolysis. In addition cleavage, ligation, and transition state stabilization, H289 may have important contributions to recombination that are not directly related to phosphoryl transfer. H289Q and H289A accumulate Holiday Junction intermediates in recombination, which reduces the amount of recombination product. On nicked suicide substrates H289Q displayed a TS cleavage preference, which has been previously observed for mutants at H289 in Cre (Gelato *et al.*, 2005; Martin *et al.*, 2002). This opposite strand preference may be unique to Cre, but could also be the cause of the HJ accumulation.

Water mediated hydrolysis by WT Cre but not H289Q indicates that a proton donor must be present to activate the tyrosine leaving group. Since H289A and H289G are also unable to generate hydrolysis product, suggesting that water less able to function in proton donation, arguing for a dual role in catalysis by H289 in proton donation and transition state stabilization. The mechanism of hydrolysis is similar to mechanism of ligation and the experiments using pNP modified substrates indicate that H289A and H289Q are partially deficient in ligation. An intriguing idea based on the rapid rate of hydrolysis in Cre as opposed to TopIBs is that YRs have evolved to use histidine to accelerate hydrolysis in order to rescue stalled reactions containing trapped

covalently bound recombinases. Such intermediates may occur at a greater frequency for YRs due to the increased complexity of recombination, in contrast to more simple mechanism of relaxation performed by TopIBs.

## **3.5 Methods**

### **3.5.1 Cloning and purification of Cre mutants**

Cre mutants were subcloned from pBluescriptIIKS+ derived vectors carrying active site mutants from Chapter 3. The XbaI-HindIII fragment containing Cre was ligated into pET21a.

Cre protein expression and purification of mutants used for biochemical studies were modified from previously published procedure (Ghosh & Van Duyne, 2002). Cre expression vectors were transformed into BL21(DE3). Cells were induced with 0.5 mM IPTG and grown for 3 hours at 37 °C. Cells were harvested by centrifugation at 4000 x g for 30 minutes and resuspended in 20 mL of lysis buffer (20 mM Tris HCl pH 7.8, 0.5 mM EDTA, 300 mM NaCl, 0.5% Triton-X100, 10 mM 2-mercaptoethanol, and protease inhibitors). Following cell lysis by homogenization (Avestin), the crude lysate was clarified by centrifugation at 14,000 x g for 20 minutes. The supernatant of the crude lysate was loaded on Hi-Trap SP-sepharose column (GE Healthcare Life Sciences). Following extensive washing in buffer (300mM NaCl, 10 mM Na/K phosphate pH 6.5, 10 mM 2-mercaptoethanol), the protein was eluted in with high salt buffer (800mM NaCl, 10 mM Na/K phosphate pH 6.5, 10 mM 2-mercaptoethanol) and dialyzed for 16 hours at 4 °C in acetate buffer (20 mM Na acetate pH 5.0, 10 mM 2-mercaptoethanol, 40 mM NaCl). The dialyzed protein was loaded directly onto a 10 mL type I hydroxyapatite column equilibrated in HAP A buffer (10 mM Na,K-phosphate pH 6.8, 10 mM 2-mercaptoethanol). Cre was eluted with a 40ml 25-40% gradient with buffer HAP B (0.8 M Na/K phosphate pH 6.8, 10 mM 2-mercaptoethanol). Peak fractions were pooled and dialyzed at 4 °C for 16 hours in acetate

buffer. Dialyzed pooled hydroxyapatite fractions were concentrated using Amicon-30 devices (Millipore/Amicon) to >10-15 mg/mL. The concentrated protein was buffer exchanged into storage buffer (20 mM Na acetate pH 5.0, 40 mM NaCl, 2 mM dithiothreitol (DTT)). The resulting protein was >90% pure as judged by Coomassie Blue-stained SDS-PAGE. Concentrated purified protein was stored at -20 °C in storage buffer containing 40% glycerol.

### **3.5.2 *LoxP* assay substrate purification**

A table of single-stranded oligonucleotides and the annealed substrates used in Cre assays is provided in Appendix B.

DNA oligonucleotides were either purchased from Integrated DNA Technologies (IDT) or the Keck DNA facility at Yale University, or synthesized with a Mermaid 4 DNA synthesizer (Bioautomation). Crude oligonucleotides that had the terminal 5'-DMT (dimethoxytrityl) group cleaved during synthesis were purified by HPLC anion exchange using a DNAPac column (Dionex) followed by SepPak C18 cartridge (Waters) to remove trace salts. Oligonucleotides were then dried down and stored in annealing buffer (10mM Tris pH 7.4, 0.5 mM EDTA, and 100mM NaCl).

Oligonucleotides containing a 5'-DMT group following synthesis were purified using Glen-Pak cartridges (Glen Research). Oligonucleotides are deprotected with 3 mL of concentrated ammonium hydroxide at 55 °C for 16 hours. An equal volume of 100 mg/mL NaCl was added and then loaded on a Glen-Pak cartridge that has been washed with 0.5 mL acetonitrile, followed by 1 mL 2 M TEAA (Triethylamine Acetate pH 7.0). Following loading, the cartridges were washed twice with 1 mL Salt Wash Solution (5% acetonitrile in 100 mg/mL NaCl) and twice with 1 mL 2% TFA (Trifluoroacetic Acid). Immediately following TFA, the cartridges were washed twice with 1 mL water and then eluted with 0.5 mL Elution Solution (50 % acetonitrile in water with 0.5% concentrated ammonium hydroxide). The eluted oligonucleotides were lyophilized. The dried DNA is dissolved in annealing buffer and stored at -20 °C.

## **5' bridging phosphorothiolate DNA**

5'-bridging phosphorothiolate substrates, loxSS7-8 and loxSS9-10 (Appendix B) were constructed as previously described (Ghosh *et al.*, 2005a).

## **Synthesis of pNP-modified DNA**

3'-pNP (para-nitrophenol) modified oligonucleotides were synthesized using previously published methods (Whiteson *et al.*, 2007; Woodfield *et al.*, 2000). loxSS1c and loxSS4c were synthesized using 3'-PO<sub>4</sub> CPG (Glen Research) and standard phosphoramidites on a Mermaid 4 DNA synthesizer (Bioautomation). Following the removal of the 5'-DMT protecting group, the resin was incubated in concentrated ammonium hydroxide at 55 °C for 16 hours. The DNA was precipitated with the addition of 0.1 volume of 3M Na acetate and 3 volumes of absolute ethanol. The pellet was resuspended in 250 mL of 2 mM MgCl<sub>2</sub>, 100 mM MES (pH 5.5). Aliquots of 3M para-nitrophenol (pNP) (Aldrich) in acetonitrile and 0.048 g of 1-[3-(dimethylamino)propyl]-3-ethylcarbodiimide hydrochloride (EDC) (Aldrich) were added sequentially in the dark and mixed vigorously to form an emulsion. After 12-16 hours at 20 °C, 250ul of water was added and the aqueous layer was extracted three times with 500ul each of ethyl acetate and detritylated with anion exchange (Waters, QMA Sep-Pak cartridge). The 3'-pNP derivatized oligonucleotide was purified by reverse phase chromatography with a C18 column (Grace Vydac 218TP C18 5u (250mm x 4.6 mm)) using a 5-25% acetonitrile gradient at 1ml/min (buffered with 100 mM TEAA at pH 7.0) over 30 minutes. Peak fractions were pooled, ethanol precipitated and stored in annealing buffer.

### ***DNA <sup>32</sup>P-5'end labeling procedure***

20 pmol of DNA was end labeled with <sup>32</sup>P γATP (3000 Ci/mmol) using 10 units of polynucleotide kinase (PNK) in PNK reaction buffer (70 mM Tris-HCl pH 7.6, 10 mM MgCl<sub>2</sub>, 5 mM dithiothreitol). The reaction was incubated at 37 °C for 1 hour and quenched by addition of phenol/chloroform. The aqueous layer containing labeled oligonucleotide was purified with a size exclusion spin column, Centriscin-20 (Princeton Separations) and stored at -20°C in annealing buffer or annealed with complementary single stranded oligonucleotides in excess by slow cooling in a thermocycler and stored at -20 °C.

### ***3.5.3 Cre Binding assay***

The probe for electrophoretic mobility shift assays was a 64 bp duplex containing the *loxP* site flanked by 10 bp and 20 bp G-C-rich segments (*loxSS3/6*) (Appendix B). 10 pmol of purified top strand (*loxSS6*) was <sup>32</sup>P end-labeled with PNK and annealed with a purified, unlabeled bottom strand. Fixed or varied concentrations of Cre were added to 0.2 nM labeled substrate in binding buffer (10 mM Tris pH 7.4, 150 mM NaCl, 2mM dithiothreitol, 1 mM EDTA, 50 µg/ml sheared salmon DNA, 50 µg/mL BSA) and incubated for 30 minutes at 20 °C. Binding reactions were electrophoresed on a 6% polyacrylamide gel (29:1 acrylamide:bisacrylamide) in TBE that was pre-run at 10 V/cm for 45 min at 20 °C prior to loading. The gel was run at 13 V/cm for two hours at 20°C, dried, and analyzed with a phosphoimager.

### ***3.5.4 Cre Synapsis Assay***

Fixed Cre at saturating concentrations, generally 0.5 µM or 1 µM was added to 0.2 nM of labeled *loxSS3/6* substrate and varied amounts of unlabeled *loxSS3/6* substrate in 1X NCB buffer (20 mM Hepes, pH 7.5, 5 mM MgCl<sub>2</sub>, 150 mM NaCl, 2 mM dithiothreitol, and 50 µg/ml sheared salmon DNA) and incubated for 30 minutes at 20 °C. Reactions were electrophoresed on a 6%

polyacrylamide gel (29:1 acrylamide:bisacrylamide) in buffer containing 50 mM Hepes acid and 25 mM Tris base at pH 7.4 that was pre-run at 10 V/cm for 45 min at 20 °C prior to loading. The gel was run at 13 V/cm for 2 hours at 20 °C, dried and analyzed with Phosphoimager.

### **3.5.5 Cleavage Assays**

Cleavage assays measure the accumulation of Cre-DNA covalent intermediate. The use of three distinct DNA substrates measure Cre cleavage in different ways. Nicked and 5'-bridging phosphorothiolate (PS) cleavage assays use suicide substrates to trap covalent intermediate by preventing or slowing ligation. The equilibrium assay uses an intact natural *loxP* substrate to measure the amount of covalent product without inhibition of ligation.

#### ***Cre suicide cleavage assay (nicked substrates)***

The nicked cleavage assay used a *loxP* substrate that is composed of three oligonucleotides that anneal to form a *loxP* site with a nick 3' of the scissile phosphate. When Cre cleaves the substrate on the nicked strand, the nucleoside 3' of the scissile phosphate, which is required to attack the 3'-phosphotyrosine intermediate in ligation, dissociates away, effectively trapping the covalent intermediate. The nick can be located on either the top strand (*loxSS1/2/3*) or bottom strand (*loxSS4/5/6*) of *loxP* (Appendix B).

Unless otherwise stated, cleavage reactions were performed in conditions where Cre is saturating (500nM). The labeled substrate was added at a fixed and constant low level, ~0.2nM except where noted. When cleavage was measured in conditions favorable for synapsis, the reaction was supplemented with 50 nM unlabeled but otherwise identical substrate.

Reactions were initiated by the addition of Cre to substrate in NCB buffer (unless otherwise stated), and incubated at 37 °C. Reactions were quenched by the addition of 5X SDS loading buffer (10% SDS, 312 mM Tris pH 6.8, 100 mM dithiothreitol, 0.002% bromphenol blue and

xylene cyanol) and separated by SDS-PAGE (12.5% 37.5:1 polyacrylamide gel) for 1 hour at 20 V/cm. Gels were dried (Thermo-Savant gel drier) and analyzed with a phosphoimager (Storm 800 - GE Healthcare).

### ***Phosphorothiolate Cleavage***

Top strand (TS) 5'-bridging phosphorothiolate substrate loxSS7-8 was <sup>32</sup>P end-labeled and annealed to either unlabeled BS 5'-bridging phosphorothiolate, loxSS9-10, or unmodified intact BS (loxSS3). Cre cleavage at the site containing the phosphorothiolate results in release of a 5'-sulfhydryl group. The sulfhydryl group has a pKa five log orders lower than the hydroxyl and is a poor nucleophile to attack the 3'-phosphotyrosine intermediate, effectively trapping the covalent intermediate (Krogh & Shuman, 2000; Krogh & Shuman, 2002).

Unless otherwise stated, cleavage reactions were performed in conditions where Cre is saturating (500nM). The labeled substrate is added at ~0.2nM except where noted. When cleavage was measured in conditions favorable for synapsis, the reaction was spiked with 50 nM unlabeled substrate.

Reactions are initiated by the addition of Cre to substrate in 1x NCB buffer (unless otherwise stated), incubated at 37 °C and quenched by the addition of 5X SDS loading buffer (10% SDS, 312 mM Tris pH 6.8, 100 mM dithiothreitol, 0.002% Bromphenol blue and xylene cyanol) and separated by SDS-PAGE (12.5% 37.5:1 polyacrylamide gel) for 1 hour at 20 V/cm. Gels were dried (Thermo-Savant gel drier) and analyzed with a phosphoimager (Storm 800 - GE Healthcare).

## ***Equilibrium***

Cre cleavage on a full *loxP* site is able to be efficiently ligated, but a small fraction of covalent intermediate is trapped for wildtype Cre and larger amounts for mutants that are defective in ligation. This assay is performed in two modes. The first is a simple end-point cleavage experiment, usually a 60 minute incubation with Cre. More formally however, the assay can be used to assess covalent intermediate at equilibrium ( $K_{eq}$ ). For wildtype, this occurs in less than 60 minutes, but mutants defective in cleavage may take longer.

Single stranded *loxP*, *loxSS6* (top strand) was  $^{32}P$  end-labeled and annealed with complementary BS (*loxSS3*). Cleavage reactions were performed in conditions where Cre is saturating, generally 500nM. The labeled substrate was added at a fixed and constant low level, generally ~0.2nM. The cleavage reaction was then spiked with unlabeled substrate for a total *loxP* concentration of 50 nM to promote synapsis.

Reactions were initiated by the addition of Cre to substrate in 1x NCB buffer (unless otherwise stated), incubated at 37 °C, quenched by the addition of 5X SDS loading buffer (10% SDS, 312 mM Tris pH 6.8, 100 mM dithiothreitol, 0.002% Bromphenol blue and xylene cyanol), and separated by SDS-PAGE (12.5% 37.5:1 polyacrylamide gel) for 1 hour at 20 V/cm. Gels were dried (Thermo-Savant gel drier) and analyzed with a phosphoimager (Storm 800 - GE Healthcare).

## ***pH dependent cleavage assays***

Suicide cleavage reactions were performed at 37 °C with the following buffers (50 mM): Na acetate, pH 4.5-5.5; NaMES, pH 5.5-6.5; Na HEPES, pH 6.5-7.5; Bis-Tris-Propane, pH 7.5-9.5. Reaction buffers also contained 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 2 mM dithiothreitol, 50 ug/ml sheared salmon DNA, and 50 ug/ml BSA. Reactions containing 0.2 nM  $^{32}P$  end-labeled substrate and 50 nM unlabeled substrate in assay buffer were initiated by addition of Cre to 500 nM.

Aliquots (20 ul) were taken at appropriate time intervals and quenched by the addition of 5X SDS loading buffer (10% SDS, 312 mM Tris pH 6.8, 100 mM dithiothreitol, 0.002% Bromphenol blue and xylene cyanol) and separated by SDS-PAGE (12.5% 37.5:1 polyacrylamide gel) for 1 hour at 20 V/cm.

Equilibrium cleavage was performed with the same buffers and conditions described for suicide cleavage except an intact, symmetrized loxS2 substrate was used (Appendix B). Reactions were incubated at 37 °C and aliquots were removed at appropriate time intervals to determine when the reactions reached equilibrium.

### **3.5.6 Ligation assay with pNP substrate**

Ligation was measured by initiating a reaction containing 0.2 nM <sup>32</sup>P end-labeled pNP (loxSS1c/2c/3-TS or loxSS4c/5c/6-BS) substrate in NCB buffer and incubated at 37 °C. Reactions performed under conditions favoring synapsis were supplemented with 50 nM unlabeled but otherwise identical pNP substrate. Reactions were quenched after 60 minutes by the addition of 0.5% SDS and heated to 70 °C for 1 minute. The quenched reactions were digested with 5 ug of proteinase K for 1 hour at 37 °C and analyzed on 10% denaturing polyacrylamide gels.

### **3.5.7 In vitro Recombination assay**

The assay substrate (digested pBG1701) is the same as described for the *in vitro* assay in Chapter 2. Time course reactions were initiated by addition of 50 nM Cre to 0.2 nM <sup>32</sup>P end-labeled substrate in NCB buffer. At appropriate time intervals, 30 ul aliquots were removed and quenched with an equal volume of phenol/chloroform (Reactions were also commonly quenched with addition of 0.1% SDS, heat to 70 °C followed by digestion with proteinase K). The extracted aqueous reactions were analyzed on a 6% acrylamide gel in 1X TBE for 2 hours at 10 V/cm. Gels were dried and quantified on a Phosphoimager.

## **Chapter 4. E176 is part of a regulatory switch involving the general acid K201: characterization of the residues involved in general acid catalysis**

### **4.1 Introduction**

Cre E176 is conserved as an acidic residue in more than 95% of tyrosine recombinases (YRs). In Cre-DNA complex crystal structures, E176 accepts hydrogen bonds from the catalytic residues K201 and R173, which activate the O5' leaving group of the scissile phosphate during cleavage and activate the incoming 5'-hydroxyl group during ligation (Ghosh *et al.*, 2005a; Krogh & Shuman, 2000; Krogh & Shuman, 2002; Nagarajan *et al.*, 2005). Sequence alignments of YRs identified this highly conserved acidic residue many years ago (Abremski & Hoess, 1992; Nunes-Düby *et al.*, 1998), leading to the characterization of a small number of mutants at the corresponding D194 in Flp recombinase (Chen *et al.*, 1992a; Friesen & Sadowski, 1992; Lee & Jayaram, 1993). These mutants were found to be catalytically deficient, but without a clear understanding of function. The functional role of this residue in catalysis has largely been ignored, possibly because TopIB enzymes have no analogous residue and instead have a solvent molecule that interacts with K167 and R130 (Davies *et al.*, 2006; Perry *et al.*, 2010; Perry *et al.*, 2006).

In this Chapter I describe the biochemical characterization of E176, R173 and K201 Cre mutants. The data are most consistent with a structural role for E176 in positioning R173 and K201. The role of each of these residues in general acid catalysis is examined using 5'-bridging phosphorothiolate substrates. The resulting thio-effects are different than those measured in TopIB. Both K201 mutants and mutations that disrupt the position of K201 display a strong dependence on the concentration of phosphorothiolate substrate for activity. This differs from WT Cre, which has cleavage activity at all substrate concentrations. This suggests that K201 may act

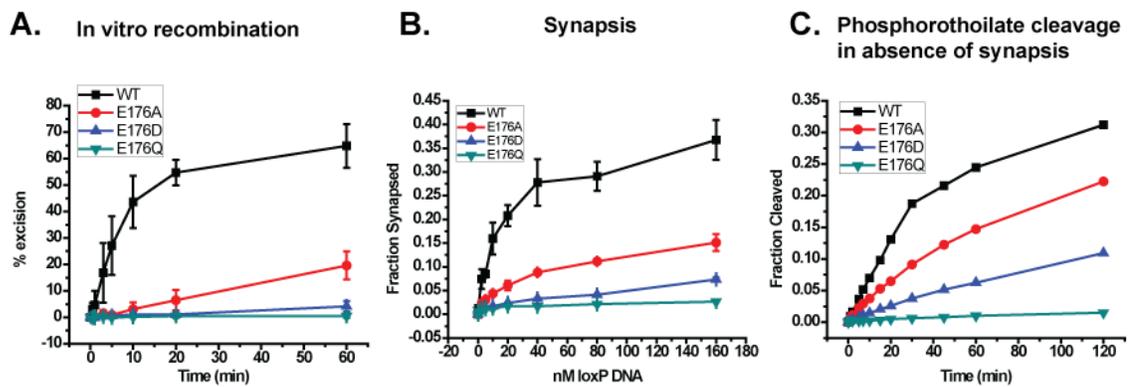
as a regulatory switch that is turned on during synapsis. In the absence of synapsis, the active site exists in a weakly active state. Synapsis positions K201 in the active site for catalysis through interactions that involve E176.

## 4.2 Characterization of E176 mutants

Despite its high conservation, fifteen substitutions at position 176 in Cre were active for recombination *in vivo* (see Chapter 2). To better understand the nature of the interactions of E176 with K201 and R173, the E176A, E176D and E176Q mutants were purified. All three mutants displayed activity *in vivo* and both E176A and E176D were active *in vitro*. However, E176Q was inactive *in vitro* with soluble lysates.

Time course experiments to probe the ability of purified E176A, E176D, and E176Q to mediate recombination *in vitro* confirmed results from the soluble lysates. Fig. 4.1A shows a graph of percent excision versus time for *in vitro* recombination. The product curves for the E176 mutants, particularly E176D and E176Q, are linear. Initial velocities were determined from the slope in order to compare the rates of recombination for the three mutants to WT Cre (see Table 4.1). E176A is the most robust of these mutants, with nearly 6% of WT recombination activity. The conservative E176D mutant is less active with 1.23% of WT activity, while E176Q has no detectable recombination activity *in vitro*. The results demonstrate that the removal of the entire E176 side chain is better tolerated by Cre than either of the more conservative substitutions.

All three E176 mutants were defective in synapsis compared to WT Cre (Fig. 4.1B). E176A is the most proficient of the mutants, followed by the E176D and E176Q mutants, which are nearly dead for synapsis. The deficiency in synapsis may explain the poor recombination activity. However, if the E176 mutants disrupt the positions of K201 or R173, then this would suggest an additional defect in phosphoryl-transfer activity.



**Figure 4.1** Catalytic activities of E176 mutants

The alanine, aspartic acid and glutamine mutants at E176 were measured for *in vitro* recombination (A), Synapsis (B), and cleavage with a suicide 5'-bridging phosphorothiolate suicide substrate (C). Cleavage was measured in the absence of synapsis with substrate at 0.2 nM.

*Table 4.1: Initial velocities of E176 mutants in in vitro recombination.*

Initial Velocities			
	% * min <sup>-1</sup>	% * sec <sup>-1</sup>	Rel:WT
WT	5.45	0.09	100
E176A	0.32	0.01	5.8
E176D	0.07	0	1.23
E176Q	N/A	N/A	N/A

Initial velocity is reported at %excised \* min<sup>-1</sup> and %excised \* sec<sup>-1</sup>. The third column reports the % of mutant activity relative to WT Cre.

E176 mutants were measured for cleavage activity to assess defects in catalysis. In order to separate cleavage activity from the defects in synapsis, cleavage was measured at concentrations of 5'-bridging phosphorothiolate substrate (0.2nM) that is 10-fold below the  $K_D$  of synapsis(Fig. 4.1C) (Ghosh *et al.*, 2007). At this low concentration of substrate, the Cre-dimer bound to *loxP* has been shown to cleave *loxP* substrate (Ghosh *et al.*, 2005a) (see Chapter 5). All three E176 mutants showed decreased cleavage of the phosphorothiolate substrate relative to WT Cre. This indicates that the E176 mutants disrupt the structure of the Cre active site as cleavage of this substrate does not require the general acid, K201 (Ghosh *et al.*, 2005a). The E176 mutants were then tested for their ability to cleave nicked suicide substrates containing phosphodiester linkages at the scissile phosphate. In the absence of synapsis (0.2 nM substrate), E176A and E176D cleave less than 5% of the substrate and E176Q does not produce any detectable cleavage product. E176A and E176D but not E176Q are rescued for cleavage at higher substrate concentrations (50 nM), favorable for synapsis formation, where they cleave as well as WT Cre (~50% cleavage product). Use of the 5'-bridging phosphorothiolate substrate under conditions favorable for synapsis is able to rescue E176Q for cleavage. The results suggest that the decreased activity observed for the E176 mutants are likely due to a structural disruption of R173 or K201.

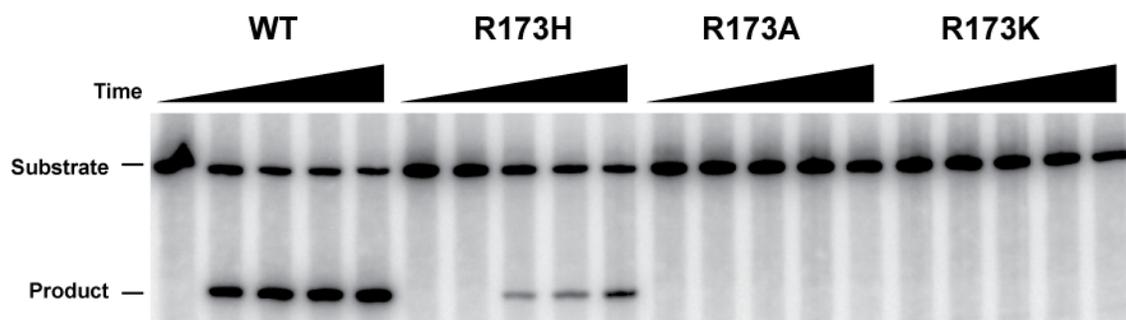
### 4.3 Characterizing the role of R173

As described in Chapter 2, the *in vivo* characterization of all nineteen substitutions at R173 revealed that R173H, R173K, and R173L mutants retained activity. The conservative R173K mutant was less active than R173H *in vivo*. This was confirmed when R173H, and not R173K, showed recombination activity *in vitro* in the presence of soluble lysates. In order to understand the role of R173 in Cre, the R173H, R173K and R173A mutants were cloned into expression vectors and purified for characterization *in vitro*.

The purified R173 mutants were tested for *in vitro* recombination activity. Portions of the reaction were removed at 5, 30, 60, and 180 minutes to allow the R173 mutants adequate time to produce a detectable excision product. As shown in Fig. 4.2, WT Cre efficiently generates excision product, reaching equilibrium after 30 minutes. R173H excises the 234 bp product with noticeably reduced activity compared to WT, consistent with the results determined using soluble lysate. After three hours of incubation, no excision products were observed for either of the R173K or R173A mutants. Because R173K displayed weak activity *in vivo*, a longer incubation may be necessary to produce detectable excision product *in vitro*. Covalent intermediate accumulated in the soluble lysate of R173H but this was not observed in the assays using purified R173H protein (Fig. 4.2). This absence of covalent intermediate was a result of quenching with phenol/chloroform causing any substrate covalently attached to Cre to partition into the organic phase.

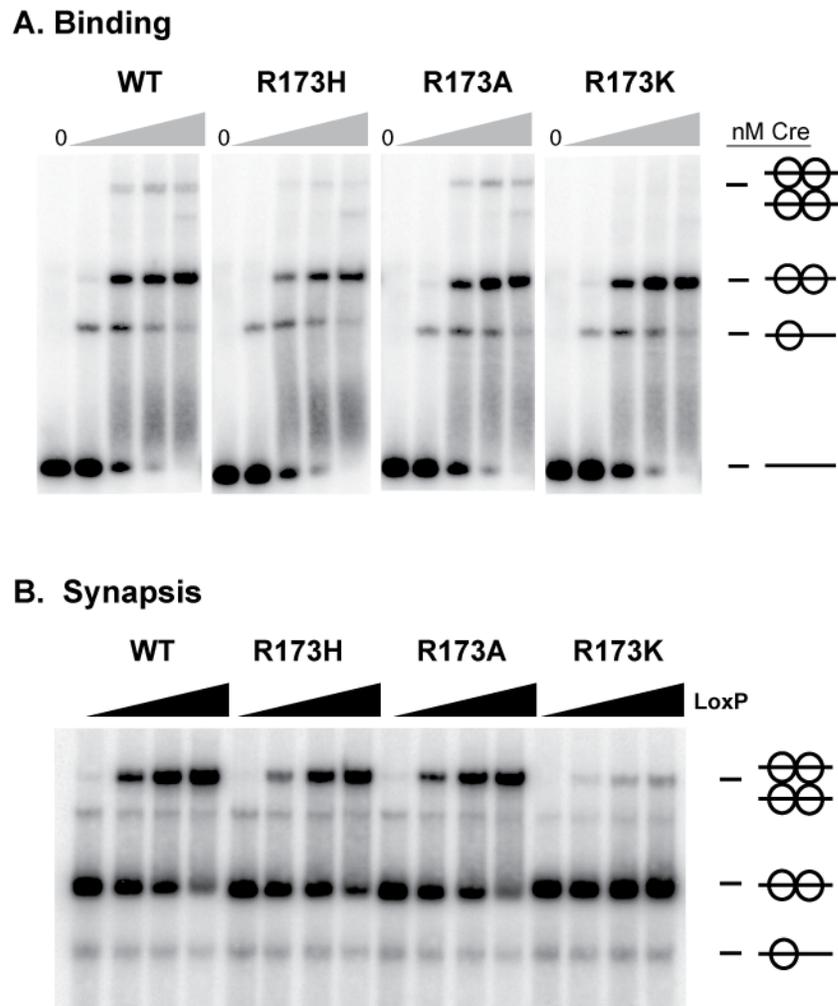
The R173K mutant in Cre has been shown to be defective in synapsis (Guo *et al.*, 1999). The mutants at R173 were therefore, tested for their ability to bind and synapse *loxP* DNA. WT Cre binds *loxP* cooperatively as a dimer with subnanomolar affinity. As shown in Fig. 4.3A, all three mutants bound *loxP* like WT Cre, where protein was titrated in ten fold intervals from 0.5 to 500 nM with 0.5 nM <sup>32</sup>P-labeled *loxP* substrate. At 500nM Cre, all Cre mutants completely bound the *loxP* substrate. During the synapsis assay (Fig. 4.3B), Cre (1 $\mu$ M) was titrated with increasing concentrations of unlabeled *loxP* substrate up to 160 nM. R173H and R173A are both competent in synapsis, while R173K is severely deficient. This suggests that the interactions made by R173 are not required for synapsis and that R173K may disrupt synapsis through a disruptive interaction. Although the defect in synapsis could explain the weak recombination activity observed for R173K, this mutant may also disrupt phosphoryl transfer as was found for the corresponding mutation in TopIB (Krogh & Shuman, 2000; Krogh & Shuman, 2002; Wittschieben & Shuman, 1997).

To better characterize the functional role of R173 in catalysis, cleavage assays were performed using both nicked suicide and intact *loxP* substrates as described in Chapter 3 (Fig. 4.4A). Cleavage reactions were performed with 50 nM total substrate to promote synapsis and optimize the cleavage activity of Cre. After 1 hour the R173H mutant showed modest cleavage activity correlating well with recombination activity. However, R173A, which is inactive for recombination *in vivo* (Chapter 2) or *in vitro*, was also able to cleave the nicked suicide substrate. A weak band corresponding to cleavage of *loxP* by R173K is visible with the nicked top strand (TS) substrate (loxSS1/2/3). The R173 mutants cleave the TS preferentially with both the TS and BS (bottom strand) nicked substrates. WT Cre and R173H accumulate more TS covalent complex, but R173A accumulates BS covalent complex with intact *loxP*. The rationale for the strand bias is unclear, but mutants at H289 also have a strand cleavage preference opposite to that of WT Cre (see Chapter 4). In the crystal structure of the Cre transition state mimic, R173 and H289 interact with the non-bridging oxygen (O2P) of the scissile phosphate of *loxP*, suggesting that this interaction is important for the strand cleavage preference in Cre (Fig. 2.2).



**Figure 4.2** *In vitro recombination assay of the Cre R173 mutants*

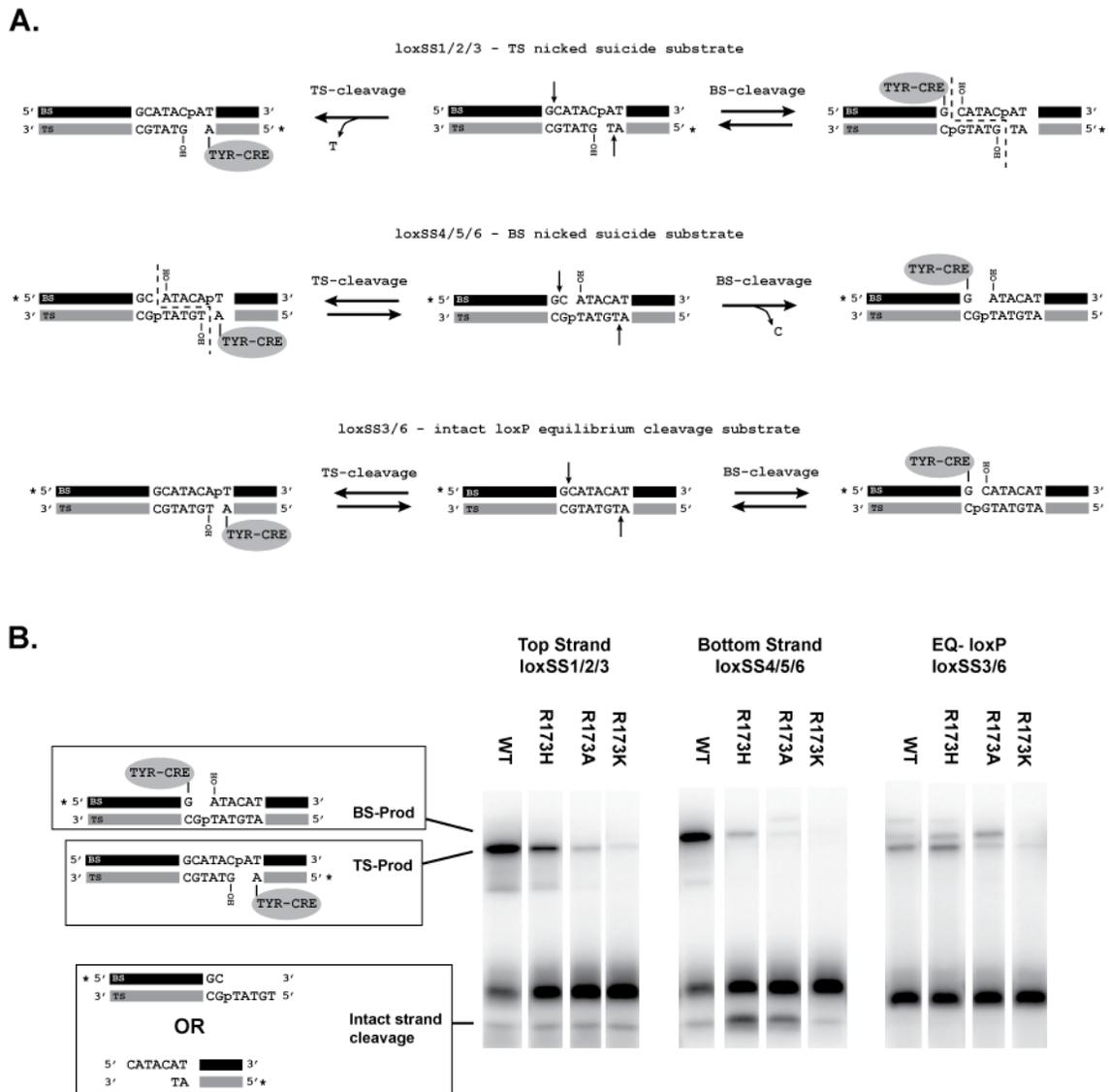
Cre excision of the 602 bp recombination substrate produces a 234 bp product. Reactions were quenched at 0, 5, 30, 60 and 180 minutes with phenol/chloroform.



**Figure 4.3** *Binding and synapsis of R173 mutants*

A. Binding of Cre at 0, 0.5, 5, 50 and 500 nM to 0.5 nM *loxP* substrate. In addition to monomer and dimer species, the synapsis band is also visible for WT Cre and the R173H and R173A mutants.

B. Synapsis of R173 mutants at 1 μM Cre is measured using 0.2, 10, 40, and 160 nM *loxP*.



**Figure 4.4 Cleavage by R173 mutants**

A. Schematic showing Cre cleavage of TS-nicked (loxSS1/2/3), BS-nicked (loxSS4/5/6) and intact *loxP* (loxSS3/6).

B. Gels of Cre R173 mutant cleavage reactions with the Ts, BS and intact *loxP* substrates. R173 mutants preferentially cleave the TS substrate (covalent product for loxSS1/2/3, and intact strand product for loxSS4/5/6). However when tested with the intact *loxP* substrate (loxSS3/6), R173H accumulates BS covalent product while R173A accumulates TS covalent product.

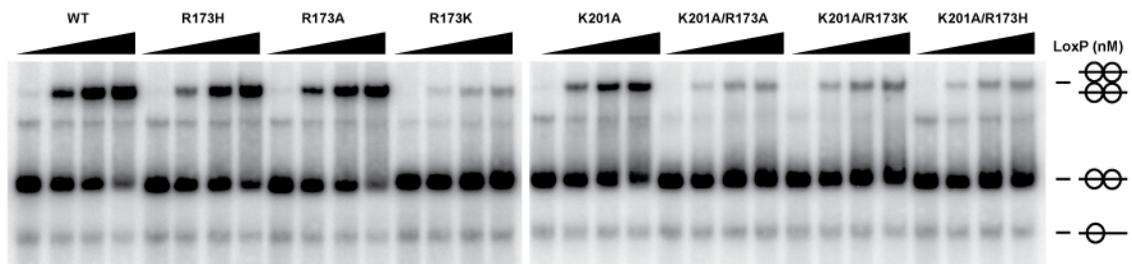
## 4.4 Thiol rescue of Cre mutants at R173, K201 and E176

Studies of vaccinia virus topoisomerase using 5'-bridging phosphorothiolate suicide substrates suggested that both K167 and R130 (Table 1.1) may function in proton donation to the 5'-hydroxyl leaving group during cleavage in TopIB (Krogh & Shuman, 2001; Krogh & Shuman, 2002). However, it remains unclear which of these two residues is the proton donor, as a second study suggests that R130 may act as the sole proton donor (Nagarajan et al., 2005). These experiments rely on the fact that the 5'-thiol leaving group has a pKa 5 units lower than the 5'-hydroxyl leaving group, and therefore cleavage does not require assistance of a general acid proton donor. Mutants of residues involved in general acid catalysis with reduced ability to cleave natural substrates containing a phosphodiester (PO) linkage at the scissile phosphate will be rescued using a 5'-bridging phosphorothiolate substrate (PS). The thio-effect is a measurement of this rescue obtained by taking the ratio of the rates of cleavage with the PS substrate and the PO substrate. A large thio-effect suggests the mutated residue is involved in general acid catalysis, which was determined for R130 and K167 in TopIB (Krogh & Shuman, 2000; Krogh & Shuman, 2002) and K201 in Cre (Ghosh et al., 2005a).

To determine the roles of K201 and R173 in general acid catalysis by Cre, the K201A/R173A, K201A/R173K and K201A/R173H double mutants were purified. Both the single and double mutants were tested for their ability to synapse *loxP* sites. As Fig. 4.5 shows, WT Cre and K201A have the same level of synapsis, but the K201A double mutants are defective in synapsis. This result suggests that an interaction from R173 or K201 with O5' of the scissile phosphate is necessary for efficient synapsis. K201 likely satisfies this requirement when R173 is mutated to alanine or histidine, and R173 is capable of fulfilling these interactions with the scissile phosphate in the absence of K201. However the R173A, R173H and presumably even R173K mutants are unable to interact with O5' and are therefore unable to efficiently mediate synapsis in the absence of K201.

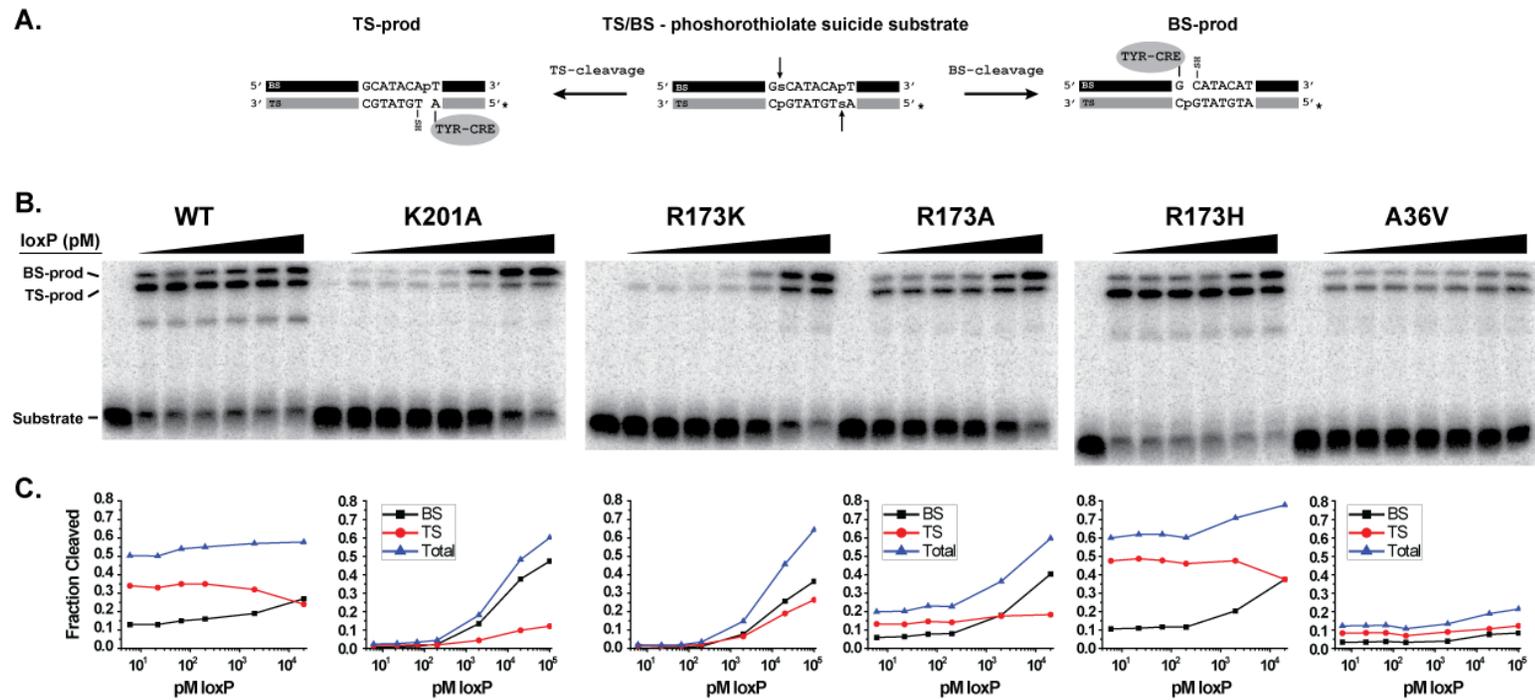
WT Cre has been shown to cleave low concentrations of the PS DNA efficiently (Ghosh et al., 2005a). At 0.1 nM substrate, which is nearly 100-fold below the  $K_D$  for synapsis, Cre exclusively cleaves the TS of the phosphorothiolate substrate. As the substrate concentration is increased, cleavage of the BS becomes preferred in an apparent correlation with formation of synapsis. To determine if the cleavage activity of K201A and the mutants at R173 are similar to that of WT Cre, cleavage of 6 pM to 100 nM PS DNA was measured. The reactions were analyzed by SDS-PAGE and quantitated for TS, BS and total cleavage (Fig. 4.6). Only R173H displayed a cleavage profile resembling that of WT Cre. K201A and R173K do not cleave the PS substrate on either strand at low concentrations of PS, but are able to cleave at higher concentration of substrate. This suggests that these mutants are unable to cleave in the absence of synapsis and the formation of synapsis rescues the cleavage defect of the Cre-loxP dimer. The R173A mutant shows reduced cleavage activity at low substrate concentrations. Based on these results, concentrations of loxP substrate greater than 20 nM are optimal for measuring thio-effects, as all mutants show cleavage of the phosphorothiolate substrate at these concentrations.

To quantify a thio-effect, the mutants must be capable of cleaving a phosphodiester (PO) linkage at the 5' position of the scissile phosphate. Since the R173 mutants were more active in cleaving the top strand of *loxP* (Fig. 4.4), the TS-nicked substrate (*loxSS1/2/3*) was used in order to give weakly active mutants such as R173K and R173A a chance to cleave the phosphodiester bond. The TS-nicked substrate contains an intact, natural BS where cleavage and ligation of this strand is in equilibrium. To compensate for this, the same natural BS (*loxSS3*) is paired with a suicide 5'-bridging phosphorothiolate TS substrate (*loxSS7-8*). Fig. 4.7 shows these substrates schematically and an example of a 60 minute cleavage reaction using both nicked and phosphorothiolate substrates for the mutants at R173 and K201. A 50 nM substrate concentration was used in these assays to promote synapsis dependent cleavage. WT Cre efficiently cleaves both substrates. However, K201A and the R173 mutants are defective in cleaving the PO substrate. The PS substrate rescues cleavage activity of all mutants with the exception of the K201A/R173A double mutant.



**Figure 4.5** *Synthesis of R173 and K201A mutants*

Synapsis of 1 $\mu$ M Cre with 0.2, 10, 40, and 160 nM *loxP*. Gel showing synapsis of WT, R173H, R173A and R173K is duplicated from Fig. 4.3 for comparison.

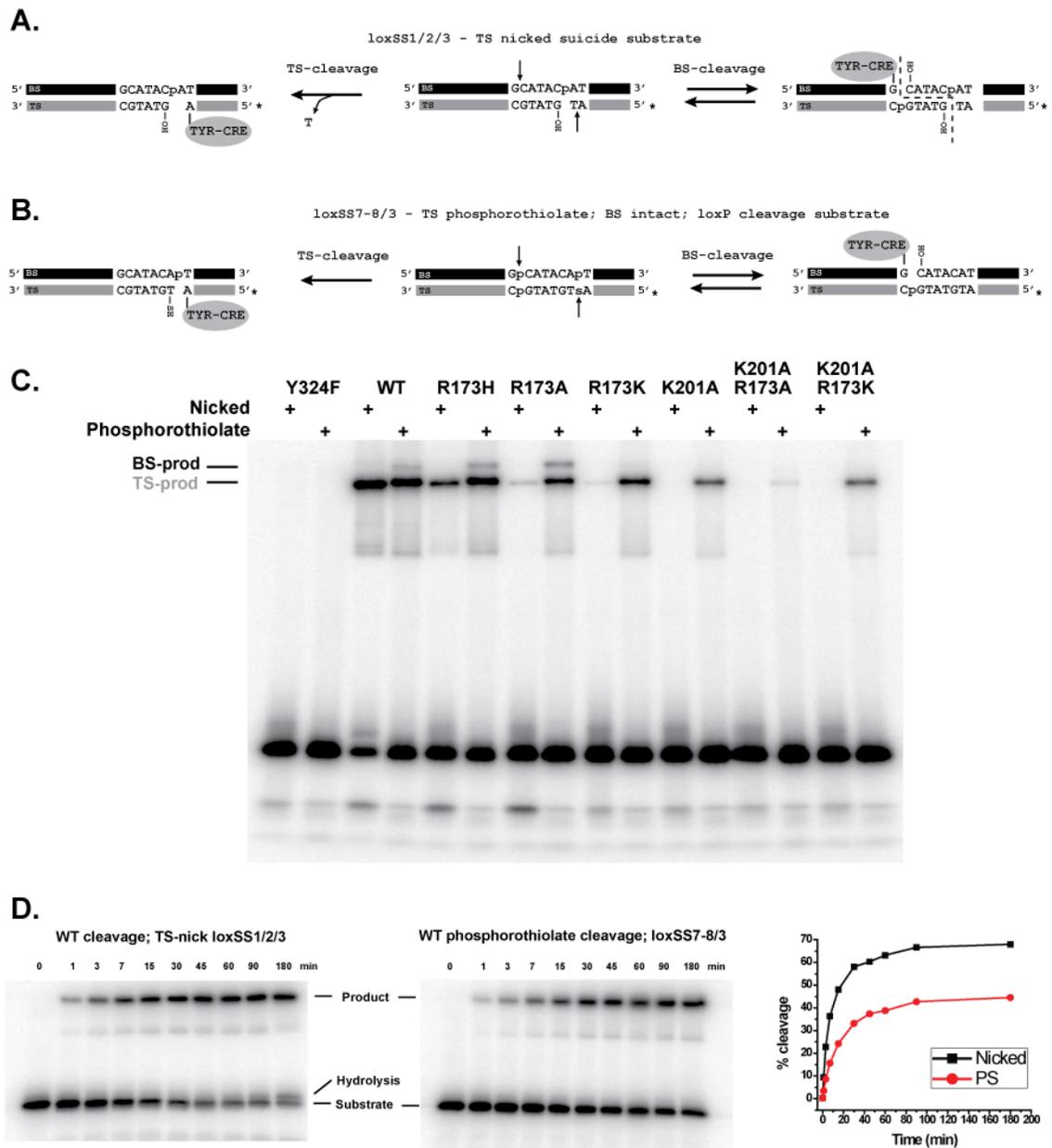


**Figure 4.6** Concentration dependence in cleavage of 5'-bridging phosphorothiolate substrates

A. Schematic of TS and BS cleavage by Cre with 5'-bridging phosphorothiolate substrates.

B. Gels of PS cleavage reaction at 1 hour. Lanes: 0, 6, 22, 66, 200, 2,000, and 20,000 pM *loxP*-PS substrate. K201A, R173K, and A36V have an additional lane containing 100,000 pM *loxP*-PS substrate.

C. Plots of the fraction of TS, BS and Total cleavage product as a function of concentration of substrate.



**Figure 4.7** *Rescue of cleavage activity with 5'-bridging phosphorothiolates*

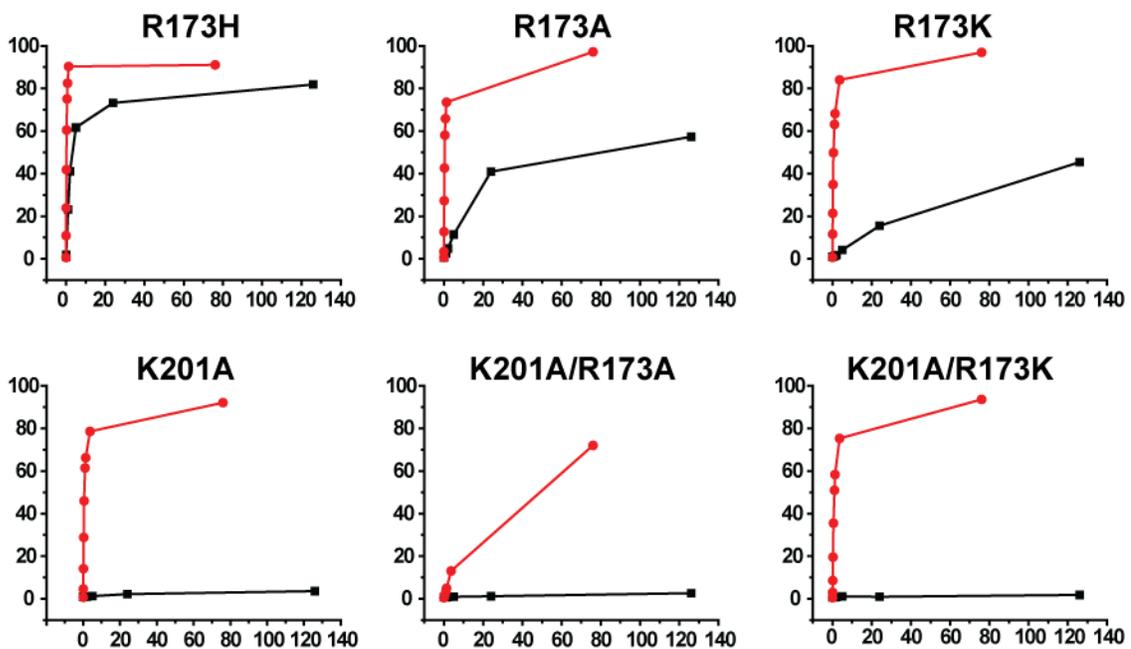
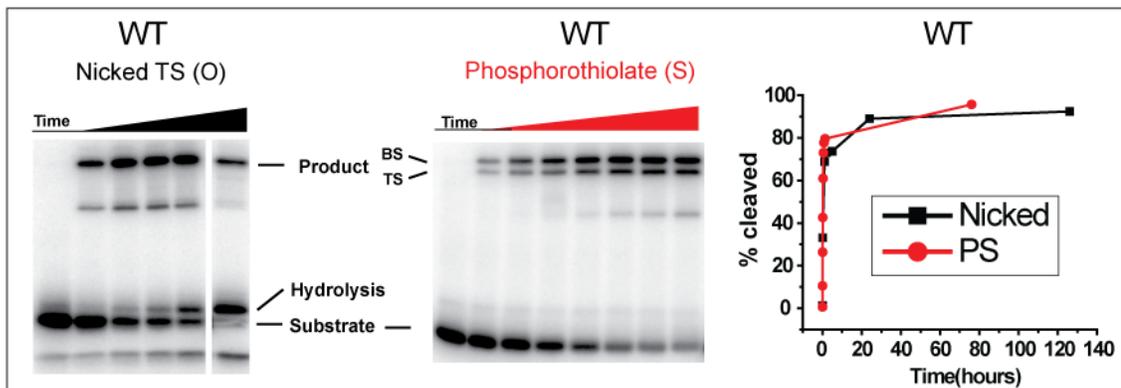
A. Schematic of TS nicked (loxSS1/2/3) cleavage by Cre. Cre cleavage at the TS nick is irreversible but cleavage on intact BS is in equilibrium. B. Schematic of Cre TS-phosphorothiolate suicide cleavage. Cre cleavage of TS-PS is irreversible but cleavage on intact natural BS substrate is in equilibrium.

C. Cleavage of nicked versus phosphorothiolate substrate after 60 minute incubation of Cre with 50 nM substrate.

D. Example of time course of cleavage with WT Cre using nicked and phosphorothiolate substrate. Cleavage of nicked DNA is about 2x faster than phosphorothiolate (PS) DNA. Thio-effect = ~0.5.

Time course experiments were performed to estimate the rate of cleavage for the Cre R173 and K201 mutants. Reactions with the nicked TS substrate were incubated for 126 hours to obtain measurable cleavage for the weakly active R173K and K201A mutants. Fig. 4.8 shows graphs of the cleavage products of TS-nicked and phosphorothiolate substrates versus time. With the exception of WT Cre, all mutants cleave the phosphorothiolate substrate faster than the nicked substrate. The K201A/R173A is less able to cleave the PS substrate than the other mutants. This demonstrates that Cre requires an interaction from K201 or R173 with either the O5' or O2P of the *loxP* scissile phosphate in order to cleave the PS substrate. K201A generates 1-3% cleavage product after 126 hours of incubation. Therefore, to determine the thio-effects for the panel of R173 and K201A mutants the initial velocities of cleavage were measured.

Initial velocities were determined in triplicate. The reactions for fast mutants such as WT Cre were quenched in under a minute. Weakly active mutants like K201A, were incubated for up to 140 hours to generate measurable cleavage product with the TS-nicked substrate. Table 4.2 shows the measured initial velocities and calculated thio-effects for Cre compared to those determined for vaccinia virus topoisomerase (Krogh & Shuman, 2002). WT Cre has a thio-effect less than 1, as it cleaves the nicked substrate faster than it does the phosphorothiolate substrate. All single mutants of R173 have thio-effects that agreed with each other within 5-fold, where R173H had the smallest thio-effect and R173K the largest. K201A showed the largest rescue with a 1300-fold thio-effect. The thio-effects for the K201A:R173A and K201A:R173K double mutants are unable cleave the PO linkage of the nicked substrate. The large thio-effects for the K201A and R173 mutants are consistent with both K201A and R173 having a role in general acid catalysis, but the larger thio-effect for K201A indicates that K201 is more likely the proton donor responsible for activating O5' of the *loxP* scissile phosphate at cleavage.



**Figure 4.8** *Cleavage time course with nicked and phosphorothiolate substrates*

Top: Gels for WT cleavage time course reactions with TS-nicked and phosphorothiolate (PS) substrates. Graph of total cleavage as a function of time (top right). Graphs of similar cleavage reactions for other mutants below and set to the same scale (% cleavage on the y-axis and time in hours on the x-axis).

Table 4.2: Thio-effects for Cre and vaccinia virus topoisomerase

<b>Cre</b>				<b>V. Topo</b>	
	<b>O</b>	<b>S</b>	<b>S/O</b>		<b>S/O</b>
WT	1.62E-01	8.37E-02	<b>0.52</b>	WT	<b>0.67</b>
R173H	3.18E-03	4.60E-02	<b>14</b>		
R173A	3.35E-04	9.63E-03	<b>28</b>	R130A	<b>190</b>
R173K	1.45E-04	1.17E-02	<b>80</b>	R130K	<b>2600</b>
K201A	2.91E-06	3.88E-03	<b>1333</b>	K167A	<b>370</b>
K201A_R173A	6.10E-07	N/A	<b>N/A</b>		
K201A_R173K	8.70E-07	N/A	<b>N/A</b>	K167A_R130K	<b>2000</b>

O: initial velocity of cleavage of the nicked TS substrate (loxSS1/2/3) as %TS\*sec<sup>-1</sup>

S: initial velocity of cleavage for the TS-phosphorothiolate substrate (loxSS7-8/3) as %TS\*sec<sup>-1</sup>

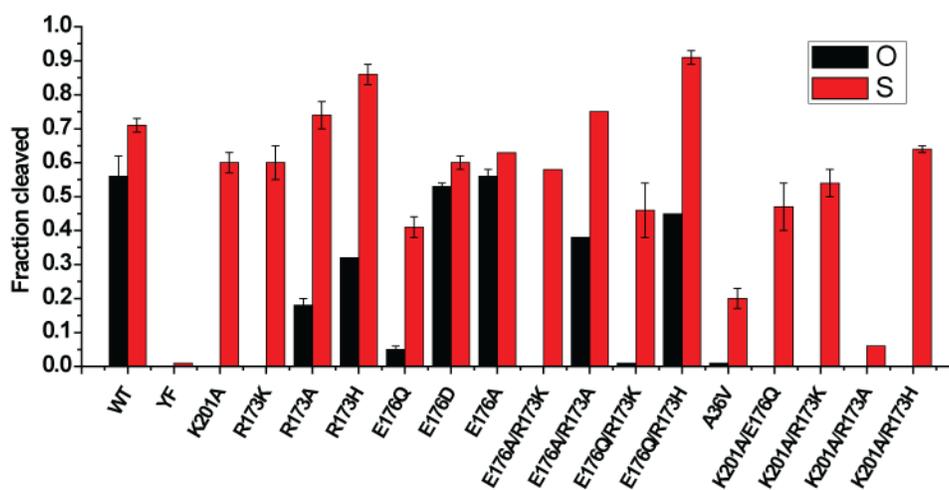
S/O: Thio-effect is ratio of initial velocities in Cre. Reported thio-effect in V. topo (Krogh & Shuman, 2002)

E176Q and R173K mutants share a similar deficiency in recombination, cleavage and synapsis, suggesting that E176Q and R173K mutants may disrupt catalysis in a similar manner. Perhaps E176Q introduces an unfavorable interaction, which forces R173 to adopt a non-catalytic conformation. To further examine the role of E176 in general acid catalysis, several R173/E176 and K201A/E176 double mutants were constructed and purified. The cleavage activity of all of the mutants was tested using nicked and phosphorothiolate suicide *loxP* substrates. One hour cleavage reactions with PO and PS substrates were measured at low (0.2 nM) and high (50 nM) concentrations to study the effect of synapsis on cleavage. The cleavage activity (plotted in Fig. 4.9) was determined from two or three independent measurements for all but the E176A/R173K and E176A/R173A double mutants. Although the rate of cleavage can not be determined based on the 1 hour cleavage assay, a qualitative assessment of the rescue activity is instructive in understanding the functional relationship between R173, K201 and E176 in general acid catalysis. The cleavage values plotted in Fig. 4.9 are consistent with results obtained from the concentration dependence of phosphorothiolate cleavage (Fig. 4.6) and experiments measuring thio-effects (Fig. 4.7 and Fig. 4.8).

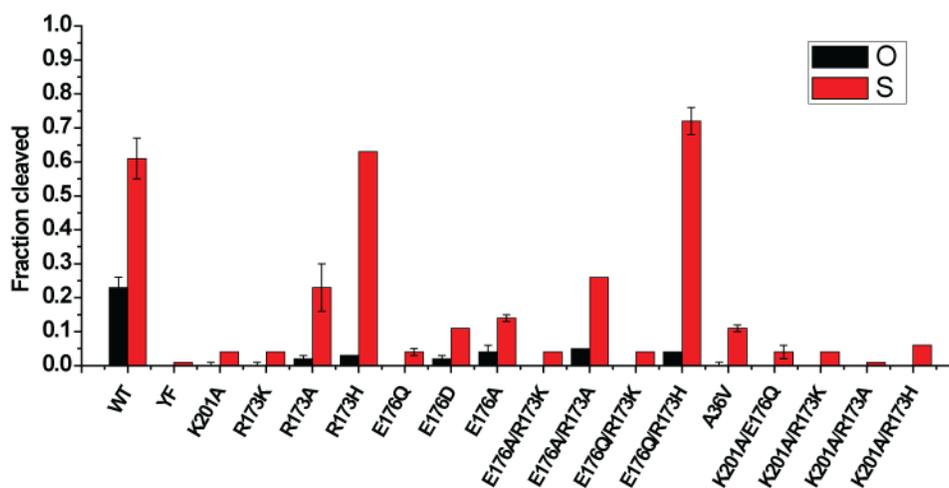
E176Q, but not E176D or E176A, has a substantial thio-effect under conditions favoring synapsis. The E176Q/R173H double mutant has cleavage activity equivalent to that of the R173H single mutant. This suggests that the mechanism of disruption by E176Q may involve Nη of R173, since R173H would not interact with E176Q in this manner (Fig. 2.2). The E176Q/R173K double mutant behaves like the R173K single mutant, showing substantial rescue with phosphorothiolate, but only under conditions favoring synapsis. The E176A/R173A double mutant is active for cleavage on the nicked substrate and displays a cleavage profile similar to that of both the E176A and the R173A single mutants. This double mutant demonstrates that the catalytic defects caused by E176Q and R173K are due to disruptive effects rather than the loss of an important interaction. The K201A/E176Q double mutant has a cleavage profile identical to that of the K201A/R173K double mutant, where a substantial rescue of cleavage activity is observed with the phosphorothiolate substrate, but only under conditions favoring synapsis. This contrasts

with the K201A/R173A mutant, which is deficient at cleavage of the phosphorothiolate substrate in conditions favoring synapsis. These results suggest that an interaction from either K201 or R173 with the scissile phosphate is required to stabilize the transition state during cleavage of the phosphorothiolate substrate. Taken together, these results from the cleavage assays provide a more complete set of data that probe the interactions made between E176, R173 and K201 as they pertain to cleavage within a Cre-*loxP* dimer and within the synaptic complex.

### A. Synapsis dependent cleavage



### B. Synapsis independent cleavage



**Figure 4.9** *Cleavage by Cre mutants on nicked and phosphorothiolate substrates*

60 minute endpoint cleavage reactions with nicked (O) and phosphorothiolate (S) substrates. Plotted cleavage is the average of two or three independent measurements except for the E176A/R173K and E176A/R173A double mutants, which were only measured once.

A. Cleavage measured in conditions favorable for synapsis formation. Substrate was at 50 nM to promote synapsis.

B. Cleavage independent of synapsis. Substrate at 0.2 nM to measure cleavage of the Cre dimer bound to *loxP* in the absence of efficient synapsis.

Table 4.3: Summary of activities of mutants at E176, R173 and K201

Name	<i>in vivo</i> activity	<i>in vitro</i> excision	Synapsis	PS no synapsis	PS + synapsis	PO no synapsis	PO + synapsis	notes
WT	+++	+++	+++	+++	+++	+++	+++	
K201A	0*	0	+++	+	+++	0	0	minimal <i>in vivo</i> activity at detection limit (Chapter 2)
R173H	++	++	+++	+++	+++	+/0	++ (*)	TS cleavage preference on nicked substrate
R173A	0	0	+++	++	+++	0	+	TS cleavage preference on nicked substrate
R173K	+	0	+	+	+++	0	+/0	minimal cleavage on nicked substrate with TS preference
E176A	+++	++	++	+	+++	+	+++	
E176D	+++	++	+	+	+++	+	+++	
E176Q	++	0	+	+	+++	0	++	
K201A/R173H			+	+	+++	0	0	
K201A/R173A			+	0	+	0	0	
K201A/R173K			+	+	+++	0	0	
E176Q/R173K				+	+++	0	0	
E176Q/R173H				+++	+++	0	++ (*)	TS cleavage preference on nicked substrate
E176Q/K201A				+	+++	0	0	
E176A/R173A				++	+++	+	++ (*)	TS cleavage preference on nicked substrate

## 4.5 Discussion

### 4.5.1 E176 in general acid catalysis

These studies demonstrate an important but non-essential structural role for E176 in positioning K201 and R173 in the active site of Cre. The E176A mutant is catalytically competent for cleavage, synapsis and recombination *in vitro*. The E176Q mutant is considered to be a conservative mutation, but was found to harbor catalytic deficiencies far greater than those of the E176A mutant. E176Q introduces a disruptive interaction, presumably with N $\eta$ 1 of R173. The phenotype of the E176Q mutant is equivalent to that of the R173K mutant, where both mutants are severely defective in synapsis and cleavage, and inactive for recombination *in vitro*.

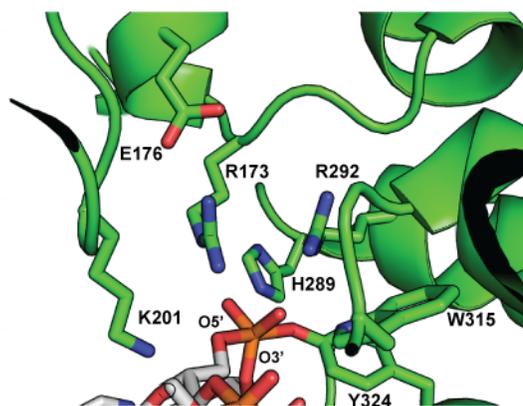
Biochemical characterization of the R173A and R173H mutants demonstrated that the interactions made by arginine to O5' and O2P of the scissile phosphate in the transition state are not required for cleavage or synapsis. Therefore, like E176Q, R173K must introduce a disruptive interaction that results in deficiencies in recombination, cleavage and synapsis.

To determine if the R173K and E176Q altered the activity via the same mechanism, a series of double mutants were constructed and studied for cleavage of 5'-bridging phosphorothiolate substrates. The results from these cleavage experiments suggested that R173K directly, and E176Q indirectly via R173, disrupts the position of K201. Indeed, studies with non-bridging phosphorothiolate substrates in TopIB suggest that in the transition state, the proximity of R130 to K167 introduces electrostatic repulsion between R130 and K167 (Nagarajan *et al.*, 2005). In the transition state mimic structure of Cre the amino group of K201 is 4.55 Å and 4.66 Å from the N $\epsilon$  and N $\eta$ 2 atoms of R173, respectively. Similar R130-K167 distances are observed in the transition state structures of trypanosomal and vaccinia virus TopIB (Davies *et al.*, 2006; Perry *et al.*, 2010). Disruption of R173 by E176Q could push the N $\eta$  of R173 closer to K201, increasing the

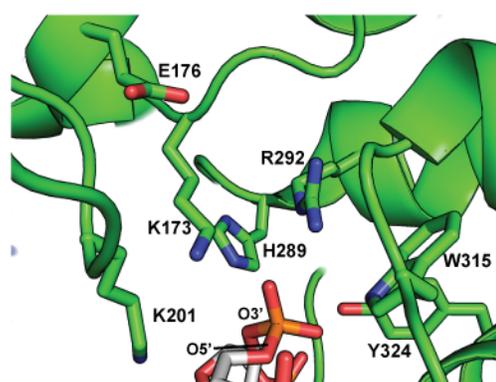
electrostatic repulsive forces felt by K201 and thereby disrupting the ability of K201 to effectively interact with O5' of the scissile phosphate and function in acid/base catalysis. Studies of R130K in TopIB with non-bridging phosphorothiolate substrates demonstrated that the lysine is too short to efficiently interact with the O5' of the scissile phosphate and too long to be able to adopt a conformation allowing for favorable bonding interactions with the non-bridging oxygen, O2P (Nagarajan *et al.*, 2005). This is likely also true for R173K in Cre given the similarity of active sites of Cre and TopIB. Because the  $\epsilon$ -amino group of R173K does not make efficient contact with the scissile phosphate, the lysine side chain likely experiences greater conformational freedom than the arginine residue and therefore disrupts K201 in a manner similar to that of E176Q. This is consistent with observations from the R173K synaptic structure where K173 is pointed directly at K201, and the  $\epsilon$ -amino group of K201 is oriented away from O5' (see Fig. 4.10) (Guo *et al.*, 1999).

E176 is well conserved in YRs as an acidic residue. Both  $\lambda$  integrase and Flp recombinase have an aspartic acid in this position. However in the crystal structures of  $\lambda$  int (Biswas *et al.*, 2005) and Flp (Chen *et al.*, 2000; Conway, Chen & Rice, 2003) DNA complexes, the aspartic acid residue interacts with the  $\beta$ 2- $\beta$ 3 region and the conserved arginine in a different manner than E176 does in Cre (discussed in Chapter 2). The glycine and asparagine mutants at D194 in Flp are defective in recombination and cleavage, which suggests that the interactions made by aspartic acid may fulfill a similar structural role as observed in Cre (Chen *et al.*, 1992a; Lee & Jayaram, 1993). E176 in Cre functions as a structural bridge that positions R173 and K201 so that their electrostatic repulsive forces are minimized. TopIBs function robustly without an equivalent residue at this position. Instead, TopIB relies on water to make a similar set of interactions with R130 and K167. This evolutionary difference between Cre and TopIB may be explained by the different regulatory mechanisms used to position the lysine and arginine in the active site.

**A. Cre transition state mimic (3MGV)**



**B. R173K synaptic structure (4CRX)**



**Figure 4.10 Disruption of K201 by R173K**

Views of the active site in the 'activated' Cre subunits of (A) the transition state mimic structure (pdb 3MGV) and (B) the synaptic structure determined with Cre mutant R173K (pdb 4CRX). K173 does not engage the scissile phosphate at either O5' or nonbridging oxygen O2P, but instead the amino group is oriented in the direction of K201. Although the shift is small, the amino group of K201 is pointed away from the active site and O5'.

Although R173 of Cre and R130 of TopIB are nearly superimposable in the transition state structures, the primary and secondary structures surrounding each arginine are very different in the two enzymes (Fig. 4.11A). In Cre, R173 is located at the N-terminus of helix H, three residues away from E176. In *variola topoisomerase*, R130 is located in a loop, which extends away from the active site. Upon DNA binding, a portion of this loop forms a short helix ( $\alpha 5$ ), which makes specific interactions in the major groove of DNA (Perry *et al.*, 2010; Perry *et al.*, 2006). The formation of  $\alpha 5$  may act as a regulatory switch that prevents R130 from effectively interacting with the scissile phosphate in the absence of DNA binding. Although  $\alpha 5$  is unique to the viral topoisomerases (Dong & Berger, 2008; Perry *et al.*, 2006), the loop region proximal to the catalytic arginine in cellular and bacterial TopIBs forms a different set of interactions that may act in a similar manner to the viral TopIBs in regulating activity through modulation of the position of the catalytic arginine (Patel, Shuman & Mondragón, 2006; Redinbo *et al.*, 1998; Stewart *et al.*, 1998).

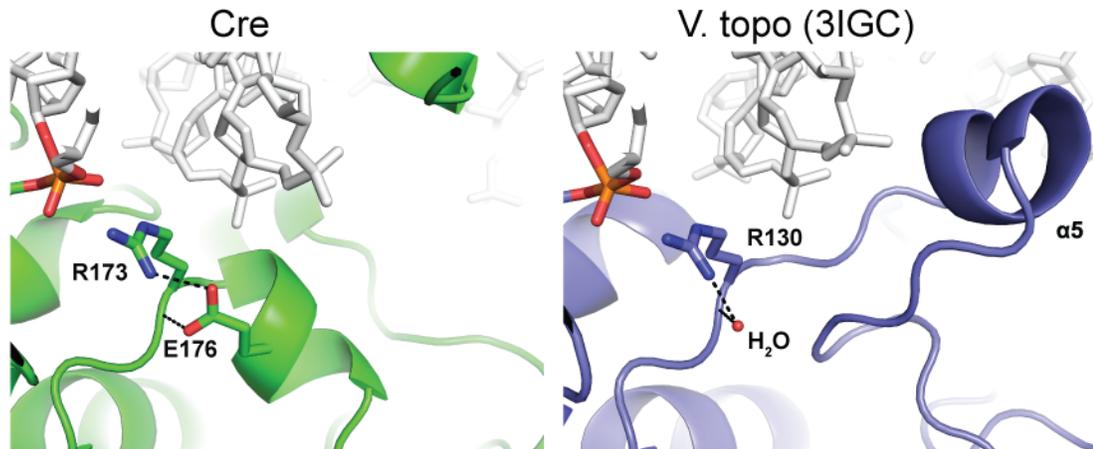
This model is consistent with the thio-effects measured in vaccinia virus topoisomerase, which indicated that mutants of R130 have a larger thio-effect than those of K167 (Krogh & Shuman, 2002). However, the thio-effects from the equivalent substitutions in Cre are different than those observed in TopIB. K201A has a much larger thio-effect than the mutants at R173, which suggests that the cleavage activity of Cre depends more on K201 than on R173. These differences in the thio-effects for arginine may be due to interactions between the  $\epsilon$ -nitrogen of R173 and the non-bridging oxygen, O2P, in the transition state. H289 also hydrogen bonds to the O2P scissile phosphate of the Cre transition state mimic. When R173 is mutated to an alanine, H289 can still stabilize the negative charge at O2P in the transition state. TopIB does not have a histidine at this position, so the stabilization of negative charge on O2P relies entirely on R130. This interaction is lost when R130 is mutated to either an alanine or lysine residue, resulting in the observed drop in cleavage activity with the natural substrate. The R130 interaction with O2P appears to be dispensable with 5'-bridging phosphorothiolate substrates, as both R130A and R130K are capable of cleavage resulting in the large thio-effect observed in TopIB.

The difference in thio-effects for the catalytic lysine in Cre and TopIB may be explained by a difference in the mechanism of positioning the lysine in the active site, as shown in Fig. 4.11B. K201 in Cre is located on the flexible  $\beta$ 2- $\beta$ 3 loop, which makes a salt bridge interaction with another Cre dimer across the synaptic interface between R199 on  $\beta$ 2- $\beta$ 3 and D126 on helix E of the synaptic partner. This suggests that synapsis has a role in positioning K201 in the active site. TopIB functions as a monomer, so the equivalent beta hairpin ( $\beta$ 7- $\beta$ 8) inserts K167 into the active site without any assistance from inter subunit interactions.

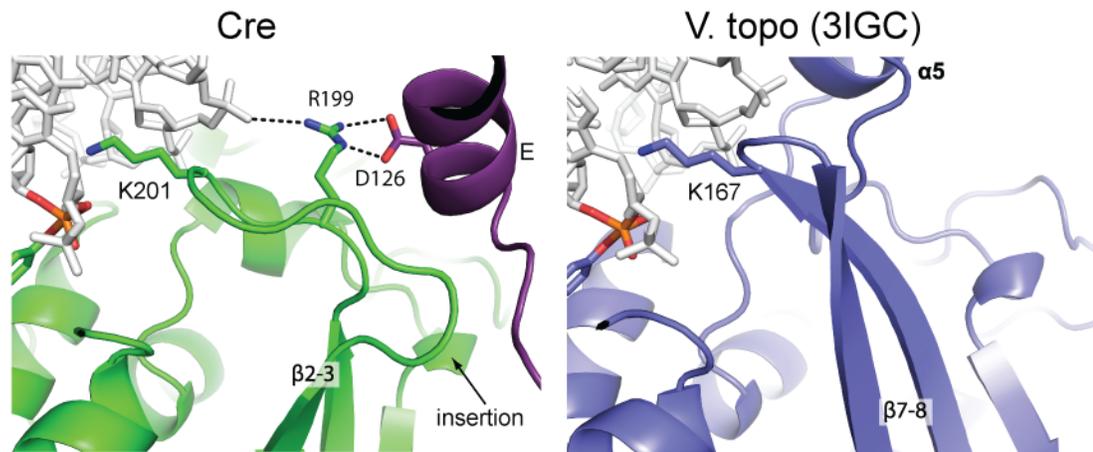
The conformational flexibility of the loop carrying K201 is best depicted by comparing the positions of K201 and the  $\beta$ 2- $\beta$ 3 hairpin in the active and inactive subunits of a Cre synaptic complex with the crystal structure of Cre monomer complex (Yuan, 2008) (Fig. 4.12). The overall structures are nearly identical with an average C $\alpha$  rmsd of 1.6 Å, but deviate at the  $\beta$ 2- $\beta$ 3 loop carrying K201. In the Cre half site complex, K201 is completely removed from the active site. R173 of the half site complex is also reoriented and does not form any interactions with the scissile phosphate. The comparison supports a mechanism of Cre regulation where synapsis formation facilitates the placement of K201 in the active site. This model of regulation agrees with the phenotypes observed for the E176Q and R173K mutants, which likely disrupt activity by the partial or full displacement of K201 from the active site via electrostatic and/or steric repulsion with R173 or K173.

The concentration dependence of phosphorothiolate substrate cleavage provides an interesting depiction of the differences between the active sites of Cre in dimeric and Cre in synaptic complexes. Synapsis is dependent on the *loxP* concentration under conditions where Cre is saturating. At concentrations of substrate well below the  $K_D$  of synapsis Cre exists almost exclusively as a dimeric complex in solution (Ghosh *et al.*, 2007). Cleavage of the phosphorothiolate substrate by WT Cre was used to show that at low *loxP* concentrations, the Cre dimer preferentially cleaves the TS, but at higher concentrations BS cleavage becomes favored correlating with the formation of the synaptic complex (Ghosh *et al.*, 2005a).

### A. Positioning catalytic arginine



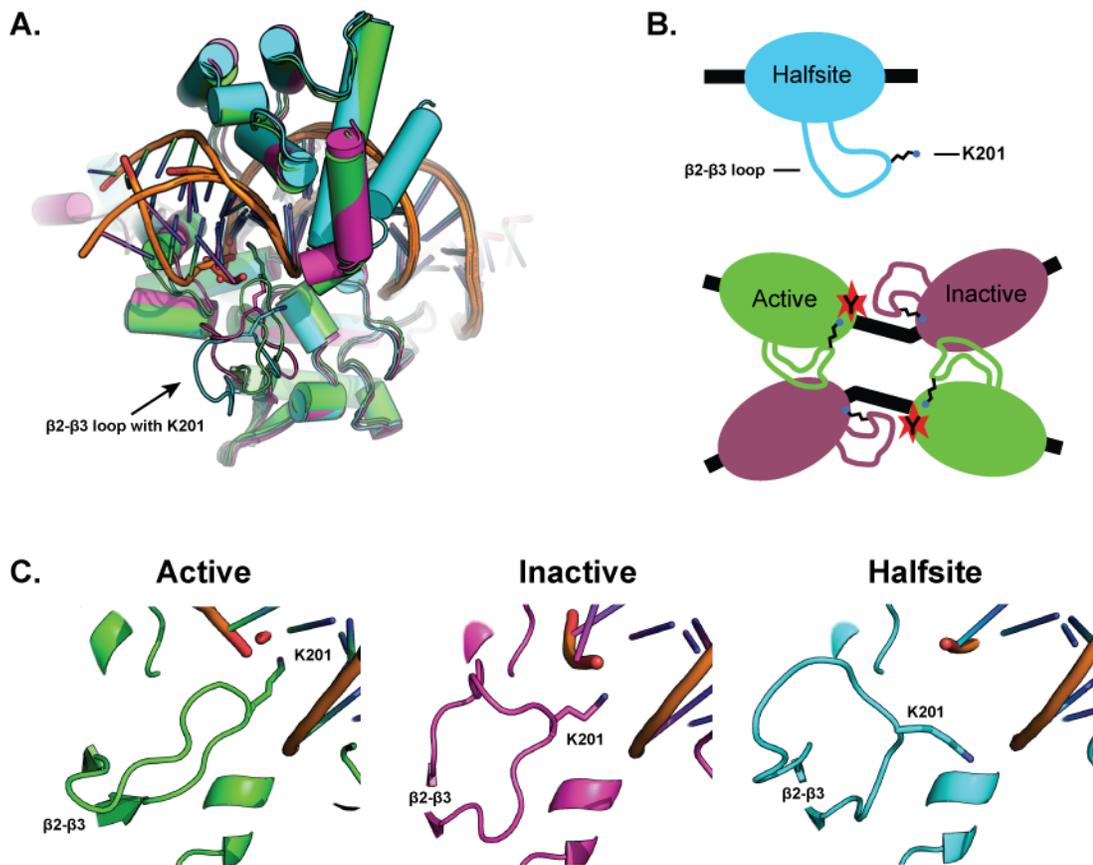
### B. Positioning catalytic lysine



**Figure 4.11** Variation in the position of catalytic arginine and lysine in Cre and TopIB

A. The position of R173 in Cre relies on E176. TopIB has no equivalent residue to E176 and instead relies on its close proximity to  $\alpha 5$ , which forms specific contacts in the major groove of DNA to dock R130.

B. The  $\beta 2$ - $\beta 3$  loop containing K201 in Cre forms interactions with  $\alpha E$  of a neighboring Cre subunit across the synaptic interface. These interactions anchor K201 in the active site of the cleaving subunit. Because TopIB functions as a monomer, K167 is more rigidly packed in the active site.



**Figure 4.12** *K201 is on a mobile structural element in Cre*

A. Overlay of the active (Green) and inactive (Purple) subunits of Cre from the transition state structure(3MGV) with the structure of the Cre half site complex (Cyan). The structures overlay well with an average RMSD of 1.6 Å, except for the region containing K201, which is on the bottom of the image and highlighted in C.

B. Cartoon representation of the position of K201 in the mobile loop. The  $\beta$ 2-3 loop of the active subunit of Cre (Green) interacts with the inactive Cre subunit across the synopsis interface, which helps to position K201 in the active site. The  $\beta$ 2- $\beta$ 3 loops of the inactive subunit (purple) and Cre monomer (cyan) do not form the same interactions resulting in the removal of the loop containing K201 from the active site.

C. Views comparing the position of K201 and the  $\beta$ 2- $\beta$ 3 loop of the active, inactive and monomeric Cre structures.

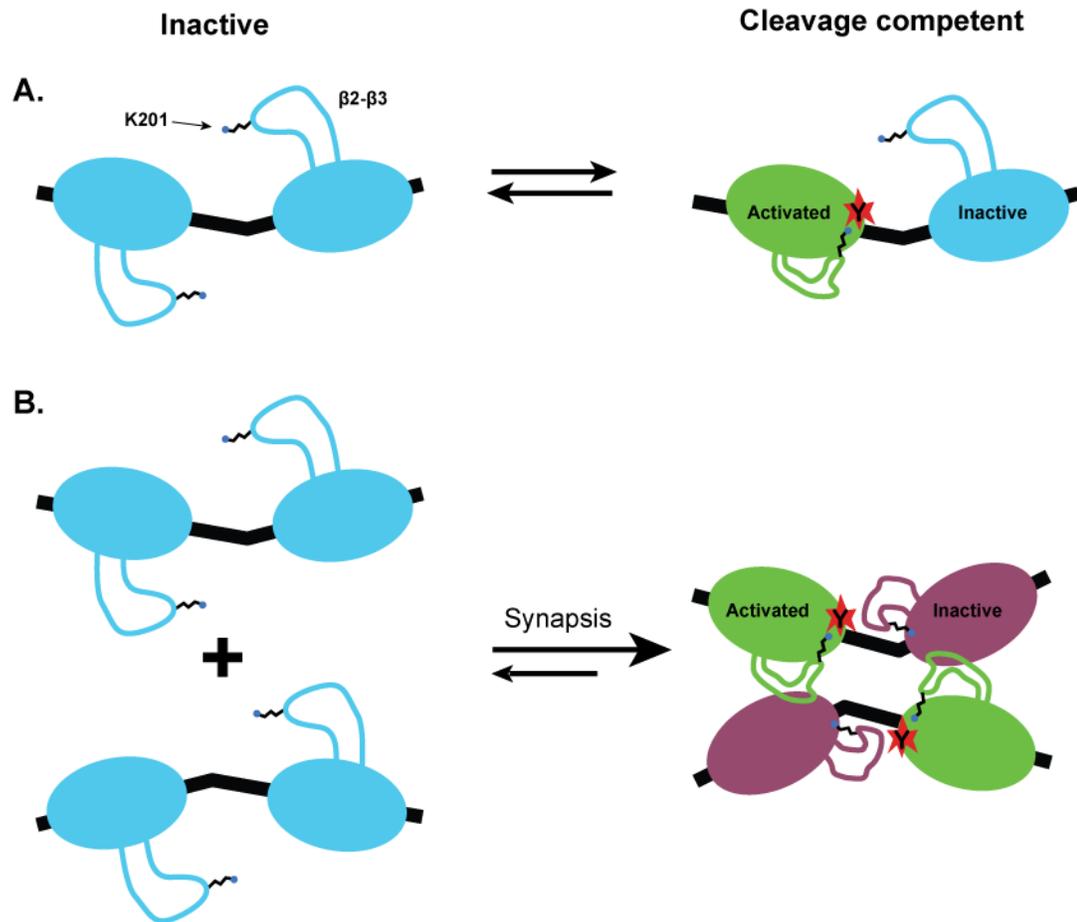
The fraction of total cleavage product remained nearly constant, even though the proportions of TS and BS cleavage changed with substrate concentration. The K201A and R173K mutants were unable to cleave the phosphorothiolate substrate at low *loxP* concentrations, indicating that these mutants are unable to cleave as a dimer. However, R173H has a cleavage profile similar to that of WT Cre, cleaving the TS efficiently at low concentrations of phosphorothiolate substrate. Cre R173A has cleavage activity at low substrate concentrations, but at a level significantly lower than that of either WT or R173H Cre. These results are perplexing in that they suggest that the general acid, K201, is essential for the cleavage activity of the Cre dimer but not for the activity of the synaptic complex. Because cleavage of the 5'-bridging phosphorothiolate substrate does not require a proton donor, K201 must have some other role in cleavage catalyzed by the Cre dimer, which becomes unnecessary upon formation of a synaptic complex.

Without a crystal structure, the positions of the active site residues of the Cre dimer in complex with DNA are unknown. However, the position of K201 within the Cre-dimer is likely to be similar to that observed in the Cre-DNA half site structure as the half site complex was crystallized in the absence of a synaptic partner. The active site of the Cre-dimer likely exists in a partially activated state due to misoriented active site residues, reduced intersubunit interactions and reduced DNA bending (Chapter 5). Perhaps K201 associates with the scissile phosphate in the Cre dimer in some manner that stabilizes the transition state enough to allow efficient cleavage of the phosphorothiolate substrate and limited cleavage of nicked DNA containing a phosphodiester linkage (Fig. 4.13A). In the absence of K201 association with the scissile phosphate, Cre is unable to cleave the phosphorothiolate substrate as a dimer. This is observed with the K201A mutant, which eliminates the lysine side chain, and the R173K and E176Q mutants, which block association of K201 with the scissile phosphate through electrostatic and/or steric repulsion. The R173H mutant is too short to disrupt the position of K201, and therefore cleaves the phosphorothiolate substrate as a dimer more efficiently than WT Cre. The partially

deficient cleavage activity of the R173A dimer suggests that the Cre active site benefits from the stabilizing interaction between N $\epsilon$  of R173 and the non-bridging oxygen, O2P, of the scissile phosphate.

Several results provide insight into how synapsis alleviates the requirement for K201 to cleave the phosphorothiolate substrate. The crystal structure of the K201A mutant in a synaptic complex shows that the removal of the lysine side chain does not affect the position of the  $\beta$ 2- $\beta$ 3 loop or the interaction of the 201 amide nitrogen with E176 (Ghosh *et al.*, 2007). The formation of the synaptic complex primes the active site for cleavage activity through DNA bending (Ghosh *et al.*, 2007; Guo *et al.*, 1999), intersubunit interactions (see Chapter 6), and possibly the interactions within the  $\beta$ 2- $\beta$ 3 loop with  $\alpha$ E via a salt bridge. In this primed state, K201 is not needed to provide any stabilizing contacts to the scissile phosphate for cleavage of the phosphorothiolate substrate. Evidence supporting this hypothesis comes from the concentration dependent cleavage assays of phosphorothiolate substrates by the E176, R173 and K201 Cre mutants. The K201A/R173A double mutant did not show significant cleavage activity with the phosphorothiolate substrate under conditions favoring synapsis, which demonstrates the importance of interactions of both R173 and K201 with O2P and O5' groups of the scissile phosphate.

The behavior of K201 as a catalytic switch (Fig. 4.13) that is turned on at synapsis highlights the importance of E176 in Cre. This switch is only effective if the lysine is able to efficiently interact with the O5' group of the scissile phosphate upon formation of a synaptic complex. E176 facilitates the orientation of K201 by serving as an anchor point via the hydrogen bonding interaction between the carboxylate sidechain of E176 with the amide nitrogen of K201, while minimizing the electrostatic repulsion between K201 and R173. The lysine switch likely evolved in Cre and other YRs as a way to restrict cleavage and possibly ligation activity to the synaptic complex. TopIB acts as a monomer, so the corresponding lysine is always placed in the active site.



**Figure 4.13** Model for regulation of Cleavage in Cre by a K201 switch

A. Cre dimer bound to *loxP*. The  $\beta 2-\beta 3$  loop containing K201 is mobile so that K201 is displaced from the active site some fraction of the time (cyan), but is also able to make productive interactions with the scissile phosphate that allow for cleavage of *loxP* in the absence of synapsis (green).

B. Under conditions favorable for synapsis, two Cre dimers bound to *loxP* (cyan) form a synaptic complex where two subunits (green) are active and two are inactive (purple). The interactions made by  $\beta 2-\beta 3$  with the synaptic partner anchor K201 in the active site of the cleaving subunit.

## 4.6 Methods

### 4.6.1 *Cloning and Cre purification*

Method details described in Chapter 3

### 4.6.2 *EMSA assays*

Binding and synapsis assays performed as described in Chapter 3.

### 4.6.3 *In vitro recombination*

R173 mutants were measured for *in vitro* recombination using an assay described in Chapter 3. The E176 mutants were measured for *in vitro* recombination activity using a different substrate than the one described in Chapter 2 and 3. For this reason, the method details are described below.

#### ***In vitro recombination assay: E176 mutants***

pGV257, a 3kB plasmid containing two *loxP* sites in direct repeat flanking a 314 bp spacer was linearized by digestion with BSRG-1, dephosphorylated and <sup>32</sup>P end-labeled with polynucleotide kinase. Reactions were carried out in NCB buffer ( 20 mM Hepes, pH 7.5, 5 mM MgCl<sub>2</sub>, 150 mM NaCl, 2 mM dithiothreitol, and 50 µg/ml sheared salmon sperm DNA) with substrate at 0.2nM and initiated by the addition of Cre to 50 nM. At appropriate time intervals, 20 ul aliquots were removed and quenched with an equal volume of phenol/chloroform. The extracted aqueous layer contains unreacted substrate, and product. The products of the reaction

are a 315 bp linearized product, which carries the 5'-<sup>32</sup>P and an unlabeled 2.8kb circle. Reactions were run on a 6% acrylamide gel in 1x TBE for 2 hours at 10V/cm. Gels were dried and analyzed with a phosphoimager.

#### **4.6.4 Concentration dependent cleavage with 5'-bridging phosphorothiolate substrate**

5'-bridging phosphorothiolate substrate preparation and cleavage assay described in Chapter 3. See Appendix B for details of loxSS7-8 and loxSS9-10 oligonucleotides.

20 ul reactions containing 0.006 nM <sup>32</sup>P end-labeled 5'-bridging phosphorothiolate substrate (loxSS7-8/9-10) and varied amounts of unlabeled but equivalent substrate (0 – 100 nM) in 1X NCB were initiated by the addition of 500 nM Cre and incubated for 60 minutes at 37 °C. The reactions were quenched by addition of 5ul of 5X SDS loading buffer and run on 12.5% SDS-Page for 1 hour at 20 V/cm. The dried gels were quantified with a phosphoimager.

#### **4.6.5 Cleavage assays for thio-effect**

Cleavage reactions with nicked TS (loxSS1/2/3) and 5'-bridging phosphorothiolate substrate performed as described in Chapter 3 with following modifications.

Reactions containing 0.2 nM <sup>32</sup>P-end labeled substrate and 50 nM unlabeled substrate were incubated by addition of Cre to 500 nM and incubated at 37 °C. At appropriate time intervals, 20 ul aliquots were removed and quenched in 5X SDS loading buffer.

##### ***Endpoint-cleavage assays***

Cleavage was measured with either TS nicked (loxSS1/2/3) or 5'-bridging phosphorothiolate (loxSS7-8/9-10) substrates.

Reactions in the absence of synapsis contained 0.2 nM  $^{32}\text{P}$  end labeled substrate in 1X NCB and were initiated by the addition of 500 nM Cre. After 60 minutes at 37 °C, the reactions were quenched with 5X SDS loading buffer, and run on a 12.5% SDS-Page gel at 20 V/cm for 1 hour.

Reactions performed in conditions favoring synapsis were identical to those described above except the reaction was spiked with 50 nM unlabeled substrate to promote synapsis.

## Chapter 5. Allosteric Regulation of Cleavage in Cre

### 5.1 Introduction

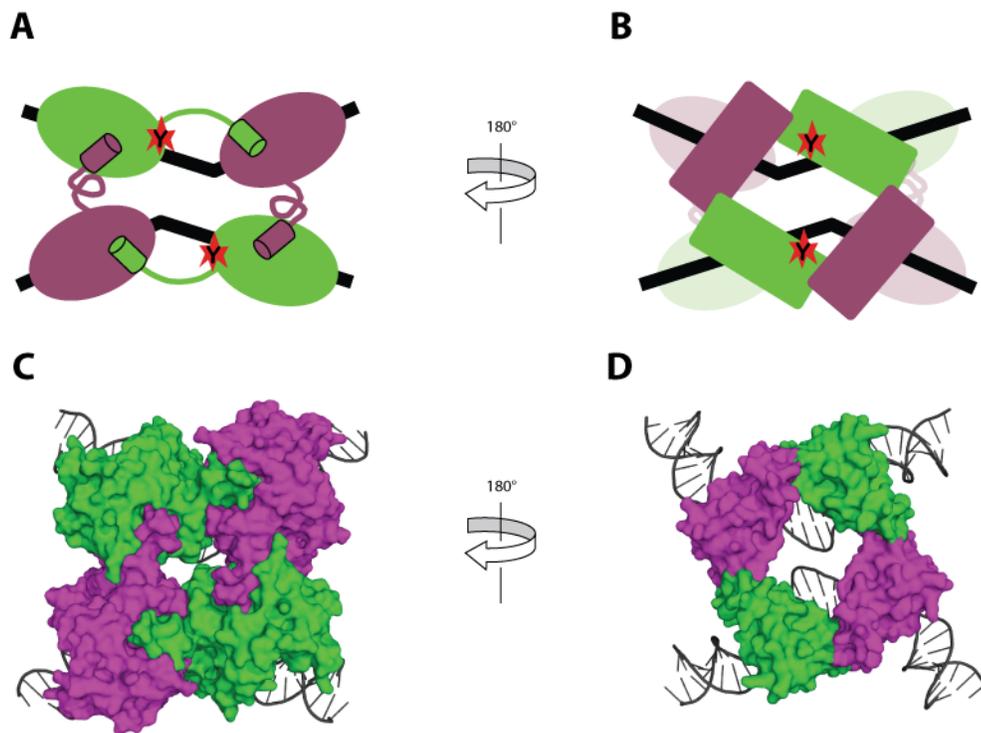
Crystal structures of Cre bound to DNA in tetrameric complexes revealed extensive protein:protein contacts at the dimer and synaptic interfaces (Van Duyne, 2001). The N-terminal and C-terminal domains of each Cre subunit bury nearly 3500 Å<sup>2</sup> of surface in the tetramer (Fig. 5.1). The packing interactions formed in the Cre tetrameric complex may be necessary to activate Cre for cleavage of DNA. Large protein interfaces are also observed in crystal structures of tetrameric complexes of λ-int and Flp, however the interfaces are less extensive, burying less surface area than Cre (Fig. 1.8) (Biswas *et al.*, 2005; Chen & Rice, 2003c).

A striking feature of the Cre tetramer is that the C-terminal tail forms an alpha helix (hN) that exchanges *in trans* to bind in a hydrophobic pocket of a neighboring subunit (Fig. 5.1). hN consists of the last twelve amino acids (E331 to D343). A short 5 amino acid linker from R326 to E331 connects hN to the preceding helix (hM), which carries the catalytic tyrosine (Y324). The exchange of hN with partner subunits and the proximity to Y324 has led to the suggestion that this region is involved in regulating recombination activity in Cre (Gopaul *et al.*, 1998; Van Duyne, 2001). The sequential exchange of the C-terminal domain is believed to be a common feature of most tyrosine recombinases despite the lack of sequence and structural conservation (Fig. 5.2) (Hallet *et al.*, 1999; Kazmierczak *et al.*, 2002; Spiers & Sherratt, 1999; Tekle *et al.*, 2002). Unlike Cre, the C-terminus of λ-int forms a β-strand that binds a neighbor subunit *in trans* (Fig. 5.3) (Biswas *et al.*, 2005). Flp recombinase differs from other tyrosine recombinases in that the helix containing the catalytic tyrosine is exchanged with the partner *in trans* and the C-terminal helix returns to bind *in cis* (Chen & Rice, 2003b). Though no equivalent tetrameric structure of XerC/D bound to DNA has been determined, sequence and structural comparisons between the unbound

XerD structure to various Cre tetrameric structures suggest that the C-termini of XerC and XerD should resemble Cre more than  $\lambda$ -int (Fig. 5.2)(Hallet *et al.*, 1999; Spiers & Sherratt, 1999; Subramanya *et al.*, 1997).

Cre is a monomer in solution and forms stable complexes as a monomer and dimer bound to *loxP* (Ghosh *et al.*, 2007). The structures of tetrameric Cre-DNA complexes do not accurately represent the structure of Cre in any of these pre-synaptic states because the protein:protein interfaces have not formed. Recently, a crystal structure of a Cre monomer bound to a *loxP* was determined (Yuan, 2008). This is the first structure that captures Cre in a pre-synaptic state. The DNA substrate used in the crystallization was designed to force formation of a crystal lattice where adjacent Cre monomers would be 180° out of phase, to prevent protein:protein interactions like those observed in the tetrameric structures. Despite the inability to form native protein:protein interactions, the core of the Cre monomer in the half site structure superimposes well with the core of Cre in the tetrameric complex structures. Interestingly, the largest differences are at the N-terminus and the C-terminus. Helix A (hA) forms part of the dimer and tetrameric interface in tetrameric structures, but is shifted by nearly 90° in the half site structure. Interestingly, hN from the half site structure manages to bind in a hydrophobic pocket of a neighboring Cre subunit as a crystal packing artifact. Although the conformation of hN bound in the pocket is similar, the linker connecting hM and hN (hM-hN) adopts a radically different conformation to allow hN to reach the pocket of an adjacent molecule in the crystal. The half site structure suggests that the Cre monomer bound to *loxP* is poised to bind hN *in trans*, but does not provide information about the position and structure of hN in the Cre monomer in solution.

Modeling has suggested that hN may be able to bind in the hydrophobic pocket *in cis*. Crystal structures have been determined for  $\lambda$ -int as an unbound monomer, monomer bound to DNA, and a tetrameric complex bound to DNA.



**Figure 5.1** *Intersubunit interactions formed by Cre in the tetrameric-DNA complex*

Cartoon and surface representation of Cre tetramer bound to loxP viewed from the C-terminal, catalytic domain (A and C, and the N-terminal domain (B and D). The N-terminal and C-terminal domains both make extensive interactions in the tetramer at both the dimer and synaptic interfaces.

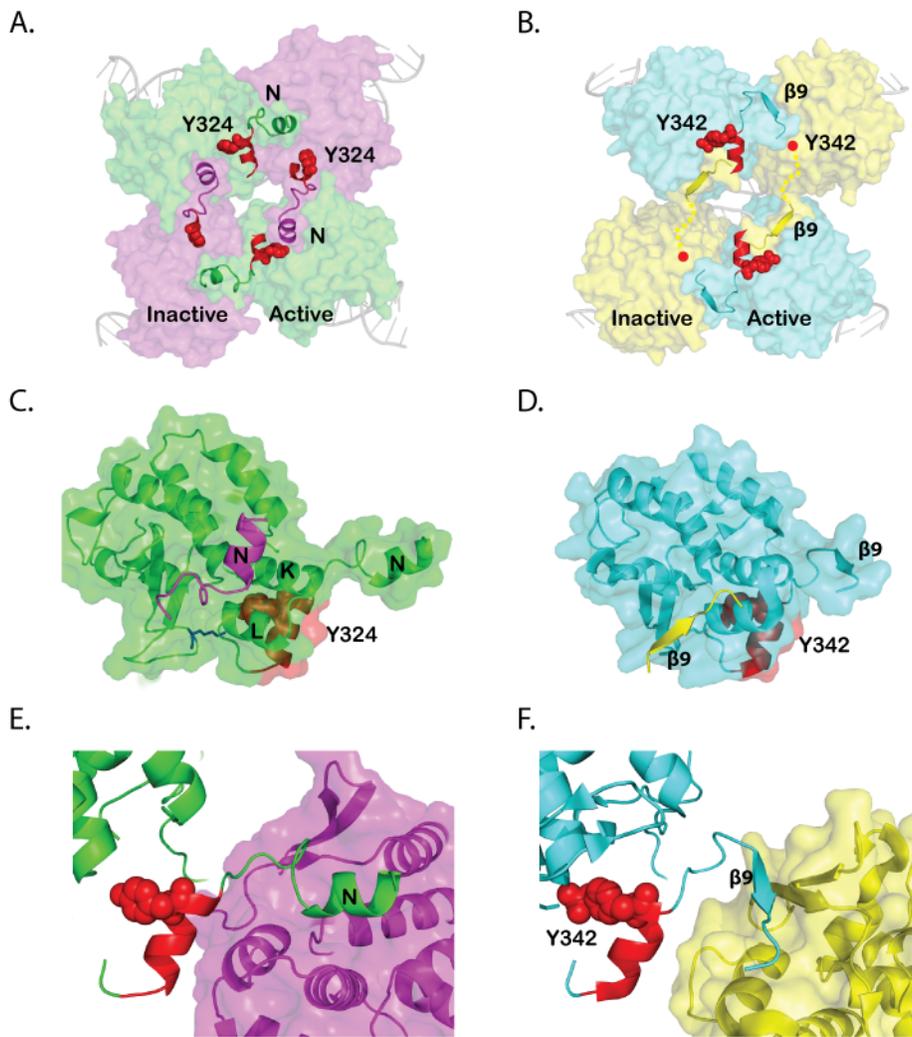


**Figure 5.2** *Sequence alignment of tyrosine recombinase C-terminal tails*

C-terminal sequence alignment. Active site residues are in boldface, and secondary structure represented below sequence. Sequences are highly conserved through catalytic tyrosine but diverge in sequence and structure in the region C-terminal to the tyrosine. Note: Flp not included because C-terminal region does not align.

In all three structures the C-terminal  $\beta$ -strand binds in a similar orientation against conserved  $\beta$ -strands that contain the catalytic lysine, except that the binding is *in cis* for the unbound structure and *in trans* for the DNA bound complexes (Aihara *et al.*, 2003b; Biswas *et al.*, 2005; Kwon, Tirumalai, Landy & Ellenberger, 1997) (see Fig. 5.3). Interestingly, the C-terminal  $\beta$ -strand of  $\lambda$ -int in the monomer complex binds in trans through a crystallographic neighbor. To determine if hN of a Cre monomer is able to bind the hydrophobic pocket *in cis*, a cysteine was engineered into both hN and the binding pocket to measure hN binding through crosslinking. Although this mutant readily formed crosslinks in the presence of *loxP* corresponding to dimers, trimers and tetramers of Cre, no crosslinks corresponding to unbound monomer or bound monomer were detected (Yuan, 2008). If hN does not bind Cre *in cis*, perhaps the C-terminus adopts a different structure like that observed for XerD where the C-terminal tail forms a larger helix that docks against the catalytic domain (Subramanya *et al.*, 1997). This idea is appealing because Cre binds *loxP* cooperatively to form a dimeric species, suggesting the protein:protein contacts formed at this interface are strongly favored. Presumably, hN from one Cre subunit is capable of binding in the hydrophobic pocket of the second subunit in the dimer, although it seems unlikely that hN of the second subunit will be able to bind in the pocket of the first without the protein unfolding.

In addition to structural questions pertaining to assembly of the tetrameric complex, the activity of Cre in pre-synaptic states is also unknown. Given the evolutionary relationship to topoisomerases, the Cre monomer may have cleavage activity. Indeed, studies have demonstrated that  $\lambda$ -int, XerC and XerD monomers have topoisomerase activity on supercoiled plasmid DNA containing recognition sites (Cornet *et al.*, 1997; Craig & Nash, 1983; Kikuchi & Nash, 1979). Further work has shown that the  $\lambda$ -int monomer is capable of cleaving a linear substrate in the absence of the regulatory N-domain (Lee, Whang, Lee & Jayaram, 1994). The trans cleavage mechanism used in FliI precludes it from acting as a monomer under normal conditions, but studies have shown that providing the nucleophile exogenously as tyramine (a tyrosine mimic) or as a peptide that contains the tyrosine allows FliI to cleave DNA as a monomer and relax DNA supercoils (Whiteson *et al.*, 2007; Xu *et al.*, 1998).



**Figure 5.3** *C-terminal elements and binding compared in Cre and  $\lambda$ -integrase*

Comparison of the tetrameric structures of Cre (A) and  $\lambda$ -int (B). Isolated view of active subunit, showing docking of hN in Cre (C) and  $\beta 9$  in  $\lambda$ -int (D). Close-up view of linker region of an active subunit connecting the catalytic tyrosine to hN of Cre (E) and  $\beta 9$  of  $\lambda$ -int (F). NOTE: hK and hL in Cre (and equivalent helices in  $\lambda$ -int) hidden from view in (E and F). Active subunits colored in green, blue; inactive subunits colored in magenta, yellow; active site tyrosine and its helix colored in red.

Topoisomerase activity in Cre has not been reported under natural conditions, although studies have demonstrated that Cre can exhibit relaxation activity on a plasmid substrate containing a wildtype *loxP* site and a mutant *loxP* site carrying a deletion in the crossover region (Abremski, Wierzbicki, Frommer & Hoess, 1986). This activity is more likely a consequence of a failed recombination reaction involving the Cre tetramer rather than cleavage activity of Cre as a monomer.

Structures of Cre have suggested that hN has a role in regulating recombination activity. Mutants of Cre at G333 on hN are defective in recombination, but retain the ability to bind *loxP* (Wierzbicki, Kendall, Abremski & Hoess, 1987). Mutations or deletions in the C-terminal regions of  $\lambda$ -int and XerCD disrupt recombination activity but can increase their topoisomerase activity relative to wildtype protein (Hallet *et al.*, 1999; Kazmierczak *et al.*, 2002; Spiers & Sherratt, 1997; Spiers & Sherratt, 1999; Tekle *et al.*, 2002). This suggests a regulatory role for hN in preventing cleavage prior to tetramer formation. In Cre, Y324 is separated by a short linker connecting Y324 and hN (hM-hN linker). In tetrameric structures of Cre, the linker adopts an extended conformation in the active subunit relative to the linker in the inactive subunit suggesting a role for hN in positioning Y324 in an active conformation (Fig. 5.1). The binding pocket for hN in Cre is formed by several helices that contain active site residues H289, R292 and W315. Binding of hN may be necessary to stabilize the active site.

In this chapter the requirements for the formation of an active Cre-DNA complex are investigated with a focus on the regulatory role and function of hN. Studies of pre-synaptic complexes show that Cre is unable to cleave DNA as a monomer but the protein interfaces formed in the Cre-*loxP* dimeric complex are sufficient for a low level of activity in the absence of synapsis. Deletion of hN results in Cre, which is incapable of recombination, cleavage or synapsis. Investigation into how hN regulates Cre cleavage activity suggests that having hN bound in the hydrophobic pocket is insufficient to activate Cre for cleavage and that the length of the hM-hN linker is essential for synapsis and positioning Y324.

## 5.2 Minimal complex for cleavage of *loxP*

Synapsis stimulates two of the four Cre subunits for cleavage of *loxP* (Ghosh *et al.*, 2005a). The protein:protein interactions formed in this tetrameric complex likely facilitate the cleavage activity (Ghosh *et al.*, 2007), but it is not known if these contacts are required for Cre to cleave DNA.

The cleavage activity of the Cre monomer bound to *loxP* half site was studied by Peng Yuan, a former graduate student. At high concentrations (4 $\mu$ M complex), Cre was capable of cleaving the half site, trapping approximately 50% of Cre in a covalent intermediate (Yuan, 2008). Although Cre was capable of this cleavage in solution, no evidence for covalent intermediate formation was observed in the crystal structure. At such high concentrations, extensive protein:protein interactions are likely formed, thereby activating Cre for cleavage. The cleavage activity was measured as a function of Cre-*loxP* complex concentration. Cleavage increased with complex concentration, indicating that the observed cleavage was likely due to protein:protein interactions, and not to the Cre monomer acting on an isolated *loxP* substrate. Additional experiments with a Cre mutant (A36V) that is known to be partially defective in synapsis (Hoess *et al.*, 1987; Wierzbicki *et al.*, 1987), further supported the conclusion that Cre requires intersubunit interactions for cleavage.

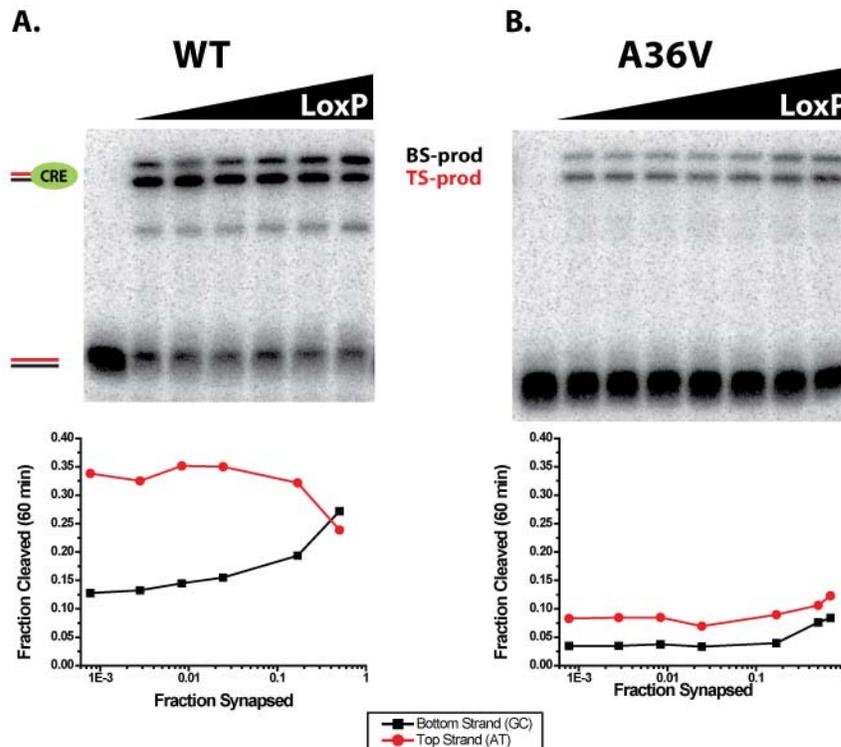
The dimeric Cre-*loxP* complex might be better suited to have cleavage activity considering the Cre dimer forms protein:protein interactions, and hN from one or both subunits will be able to bind the hydrophobic pocket of the partner subunit. The formation of synapsis has been shown to accelerate cleavage (Ghosh *et al.*, 2005a; Guo *et al.*, 1999; Hoess, R H, Wierzbicki A, Abremski KE, 1990), so in order to be able to determine if Cre is cleaving as a dimer, the reaction must be performed under conditions where synapsis does not occur. Assuming Cre is at saturating concentrations, the  $K_D$  of synapsis of Cre with *loxP* is approximately 10nM (Ghosh *et al.*, 2007). Previous work examined cleavage using 5'-bridging phosphorothiolate suicide substrates at

concentrations of substrate nearly 100 fold below the  $K_D$  of synapsis, and found that cleavage of the bottom strand (GC half site) approaches zero, but the cleavage of the top strand (AT half site) is robust even at the lowest substrate concentrations (Ghosh *et al.*, 2005a). The results suggest that the Cre dimer is capable of cleaving the top strand of *loxP*, but cleavage of the bottom strand appears to require synapsis. This set of experiments was repeated with Cre A36V, a synapsis defective mutant, to determine if a low level of synapsis is responsible for the previously measured cleavage activity (Fig. 5.4). Wildtype Cre robustly cleaved the top strand of the phosphorothiolate substrate. Surprisingly, cleavage of the bottom strand does not go to zero, but instead reaches a plateau of about 12% at concentrations 1000 fold below the  $K_D$  for synapsis. Interestingly, A36V displays a steady basal level of cleavage for almost the entire concentration range tested, which supports the idea that the Cre dimer is active for cleavage. However, the phosphorothiolate substrate is an unnatural substrate that alleviates the need for a general acid to assist the leaving group during cleavage, therefore the requirements to cleave phosphorothiolate DNA are less stringent than those for a natural substrate.

To eliminate the possibility of the that cleavage by the Cre-*loxP* dimer is an artifact of the artificial phosphorothiolate substrates, cleavage was measured with a suicide substrate that contained a natural phosphodiester linkage at the scissile phosphate. Cleavage was assayed at a concentration of *loxP* ranging from 2 pM to 2000 pM. The lowest concentration was 5000 fold below the  $K_D$  of synapsis (Ghosh *et al.*, 2007). Reactions with wild type Cre were incubated for 1 hour at 37 °C. At the lowest concentration, cleavage was measured to be 10% of total counts and remained nearly constant over a sixty fold increase in substrate concentration before increasing significantly at concentrations close to the  $K_D$  of synapsis. The trend was the same for cleavage of both top and bottom strands. The increase in cleavage near the end of the measured concentration range of the assay is consistent with stimulation from the formation of synapsis. The concentration dependence of synapsis in Cre has been studied and modeled (Ghosh *et al.*, 2007). Using this model, the approximate amount of synapsis formed can be calculated for the concentrations of substrate used in the cleavage experiment (Fig. 5.4). At 2pM substrate, less

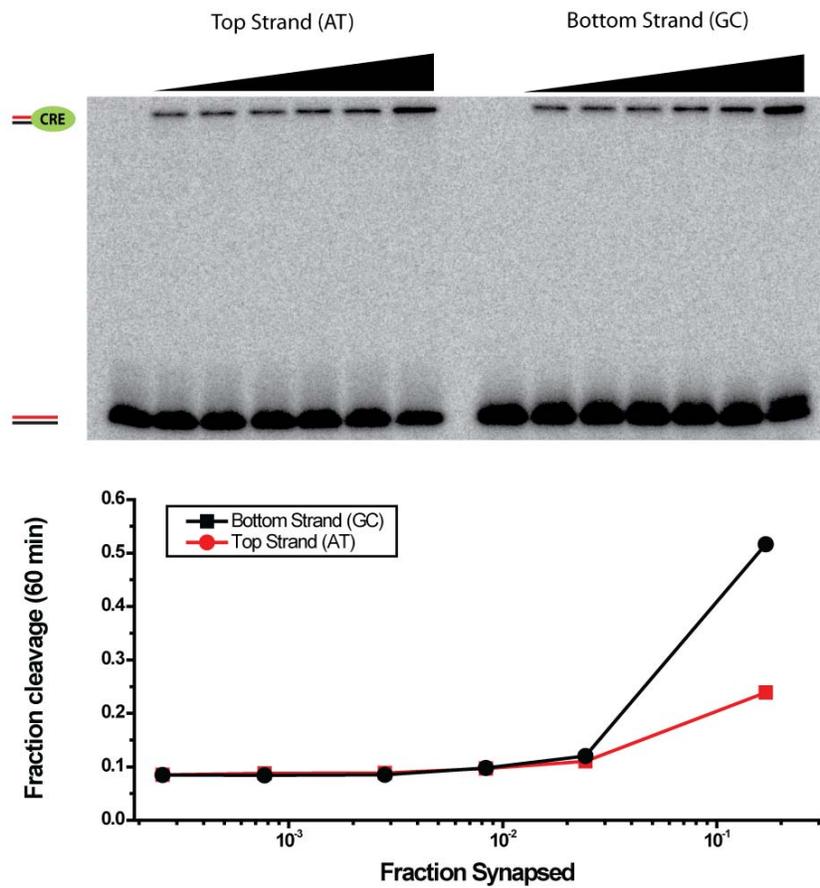
than 0.03% Cre bound *loxP* is in a synaptic complex, and while this amount will increase with concentration, the cleavage product does not. This indicates that the observed cleavage is independent of synapsis suggesting that observed cleavage is from the Cre-*loxP* dimer complex (Fig. 5.5).

The ability of the dimeric Cre-*loxP* complex to cleave but not the Cre monomer to cleave DNA substrate suggests that intersubunit protein contacts must be made to activate Cre for cleavage. It is unclear, however, if one or both Cre subunits in the dimer are capable of cleavage. Although the use of nicked suicide substrates allows cleavage to be measured at a normal phosphodiester linkage, the presence of the nick may influence bending of the *loxP* site, which is known to be important in regulating the activity of Cre in the synaptic complex (Ghosh *et al.*, 2005a; Guo *et al.*, 1997; Guo *et al.*, 1999). Without structural information for the Cre-dimer, it is unknown what role DNA bending plays in the cleavage activity of this complex.



**Figure 5.4** *WT Cre cleaves the TS of loxP preferentially as a dimer.*

Cre cleavage as a function of loxP concentration for wildtype (A) and A36V (B). Use of a suicide phosphorothiolate substrate was used to trap Cre-loxP in a covalent complex. Reactions were separated by SDS-PAGE after 60 mins. Concentrations of substrate used were, 6, 22, 66, 200, 2000, 20,000, and 100,000(A36V only) pM A. Bottom strand product (BS) runs with slower mobility than top strand covalent product (TS). Graphs on bottom show fraction cleaved vs. fraction synapsed. The fraction synapsed was calculated using a published model for Cre (Ghosh, Guo & Van Duynne, 2007). The x-axis is plotted on a log scale to show the entire concentration range clearly.



**Figure 5.5** *WT Cre cleaves nicked substrates as a dimer bound to loxP*

Cre cleavage as a function of loxP concentration for substrates with a nick on the top strand or bottom strand. Reactions were separated by SDS-PAGE after 60mins. Concentrations of substrate used were, 2, 6, 22, 66, 200, and 2000 pM. Bottom. Graph of fraction cleaved vs. fraction synapsed. The fraction synapsed was calculated using a published model for Cre (Ghosh *et al.*, 2007). X-axis plotted on a log scale to show entire concentration range clearly.

### 5.3 Characterization of the Cre C-terminal deletion mutant

To determine if hN has an important function in Cre, a deletion mutant lacking the last 12 amino acids in Cre was studied. This mutant ( $\Delta C$ ) lacks hN but retains the hN-hM linker. Cre $\Delta C$  was found to be inactive for recombination by both *in vivo* and *in vitro* assays (Table 5.1). This result clearly demonstrates that hN is essential for Cre activity. In order to understand this defect, assays to assess various activities at specific steps in the recombination pathway were performed.

*loxP* binding assays of purified Cre $\Delta C$  showed that  $\Delta C$  is partially defective. Unlike WT Cre, which shows strong cooperative binding to populate the Cre-*loxP* dimer complex, Cre $\Delta C$  formed the dimer complex in an uncooperative manner (Fig. 5.6A). This result demonstrates that hN is important for cooperative binding, and suggests that hN participates in the Cre dimer interface as observed in tetrameric DNA complex structure of Cre (Ghosh *et al.*, 2007; Guo *et al.*, 1997; Guo *et al.*, 1999). Cre $\Delta C$  was unable to form measurable synaptic complex in the EMSA assay (Fig. 5.6B). This is consistent with the crystal structure of the Cre synaptic complex, which shows that hN forms extensive interactions at the synaptic interface (Ghosh *et al.*, 2007; Guo *et al.*, 1999). Interestingly, the dimeric complex formed by  $\Delta C$  displayed slower mobility than the WT Cre-*loxP* dimer complex. This could be caused by a defect in bending *loxP*, resulting in an elongated complex that would migrate more slowly in a native gel relative to a smaller complex formed from a sharply bent *loxP* observed in the Cre tetrameric crystal structures. This is the opposite of what would be expected with bending experiments using long DNA substrates.

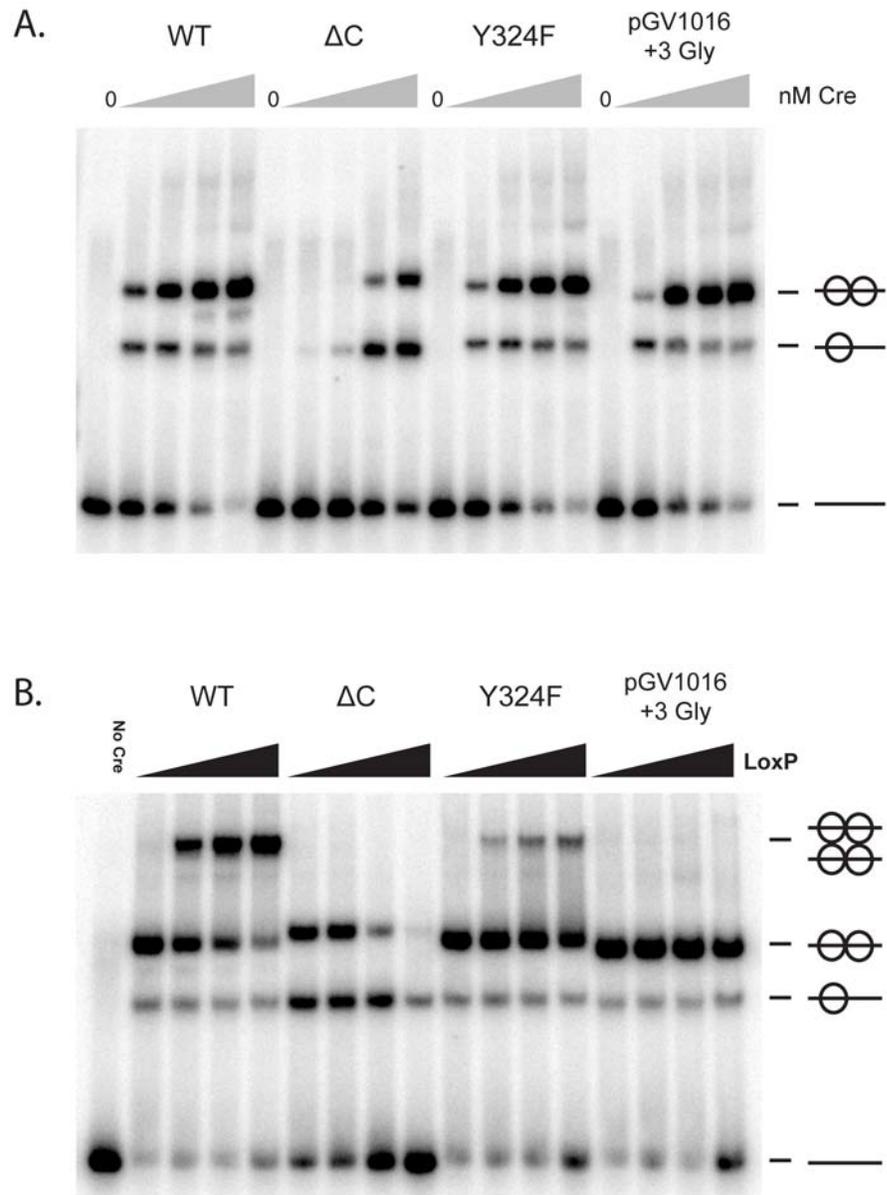
To determine if Cre $\Delta C$  is capable of cleaving *loxP*, cleavage assays with both a nicked and a phosphorothiolate *loxP* suicide substrates were performed. Cre $\Delta C$  was incapable of cleaving a natural *loxP* site after 60 minutes at 37 °C, but did have weak activity on a phosphorothiolate substrate. To be certain that  $\Delta C$  was inactive in cleavage, a second set of reactions was incubated for 140 hours at 37 °C. Cre $\Delta C$  accumulated no covalent product on the nicked suicide

substrate but did build up nearly 50% total covalent product with the 5'-bridging phosphorothiolate substrate with a strong bias for cleavage of the top strand as would be expected for a mutant defective in synapsis (Ghosh *et al.*, 2005a) (Fig. 5.7).

Requirements for cleavage of the 5'-bridging phosphorothiolate substrate are less stringent than the natural *loxP* site (Ghosh *et al.*, 2005a; Krogh & Shuman, 2000). As discussed in Chapter 4, the general acid lysine appears to be displaced from the active site of the Cre-*loxP* dimer. The increased mobility of Y324 in Cre $\Delta$ C means that the nucleophile and general acid will be less likely to occupy positions compatible with cleavage activity at the same time. Since the 5'-bridging phosphorothiolate substrate alleviates the need for the general acid, K201, cleavage only depends on the ability of Y324 to attack the scissile phosphate. Cleavage of the phosphorothiolate substrate by  $\Delta$ C indicates that Y324 is able to attack the scissile phosphate without hN. This cleavage activity is weak, which indicates that Y324 of  $\Delta$ C does not efficiently attack the scissile phosphate.

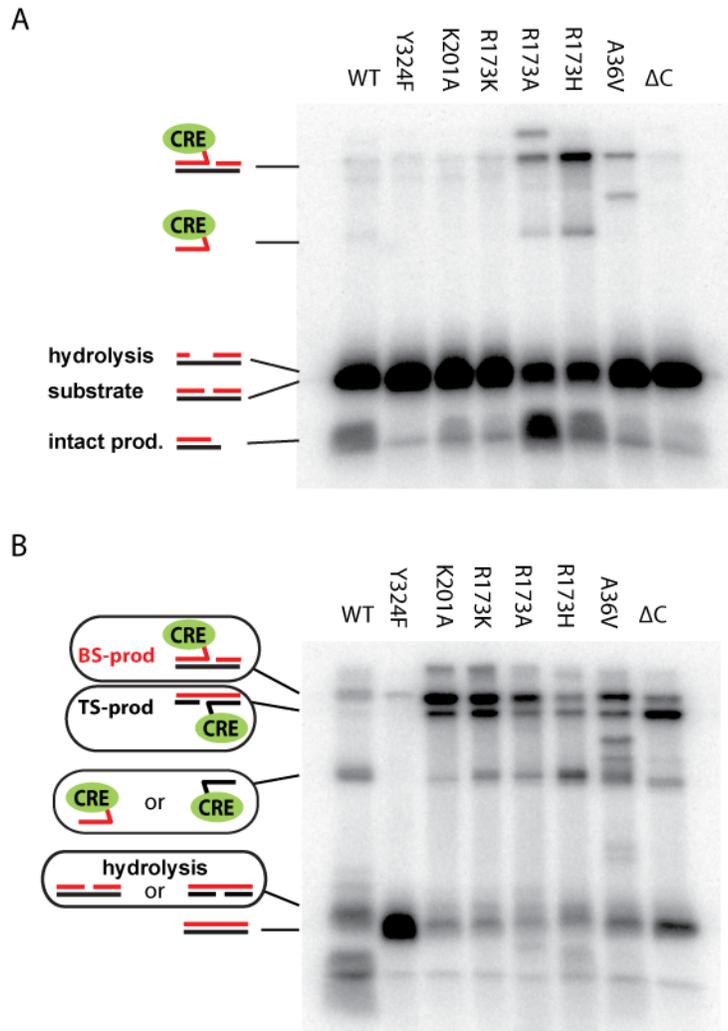
## 5.4 Mechanism of regulation by hN

The C-terminal helix of Cre is essential for recombination, synapsis and cleavage. Two possible mechanisms for regulation were tested. hN binds in a hydrophobic pocket near active site residues W315, R292 and H289 and may support a competent active site. In this model, a Cre monomer must have a neighbors' hN docked in the hydrophobic pocket *in trans* to be capable of cleavage. Alternatively, hN may regulate Cre by acting as a tether between monomers, where the docking of hN into a neighboring pocket causes a tension on Y324 of the donor just seven amino acids away from the beginning of hN, and locks it in a position favorable for attack on the scissile phosphate.



**Figure 5.6** *Binding and Synapsis of Cre $\Delta C$*

A. Binding of Cre to 200pM 64 bp loxP (loxSS3/6, see appendix B). Cre concentrations tested were, 0, 0.5, 5, 50, 500 nM. B. Synapsis by 500 nM Cre as a function of loxP (loxSS3/6) LoxP concentrations tested were 0.2, 10, 40, 160 nM.



**Figure 5.7** *CreΔC cleaves phosphorothiolate substrates.*

Cleavage of nicked bottom strand (A) or phosphorothiolate (B) loxP substrate after 140hour incubation at 37 °C. ΔC is incapable of cleaving nicked substrate containing natural phosphodiester linkage, but is able to cleavage the unnatural phosphorothiolate substrate. Note: A36V has low activity on both substrates, and WT hydrolyzes the phosphotyrosine linkage of both substrates.

### **5.4.1 Helix-N binding pocket occupancy is insufficient for Cre activity**

If Cre requires binding of hN in the hydrophobic pocket to stabilize the active site, Cre $\Delta$ C ought to be capable of cleavage when hN is docked. To test this idea, a two sets of experiments were performed. The first experiment tested if Cre $\Delta$ C could be complemented by active site mutants K201A and Y324F, which are incapable of cleavage but contain an intact C-terminus. Cre $\Delta$ C was mixed at varied ratios with K201A or Y324F and incubated with either a nicked *loxP* suicide substrate or a phosphorothiolate suicide substrate. After one hour at 37 °C, no cleavage product was visible for any protein mixture using the nicked substrate. However Cre $\Delta$ C with K201A was capable of cleaving the phosphorothiolate substrate. The K201A mutant is capable of cleaving the phosphorothiolate substrate alone (Chapter 4) (Ghosh *et al.*, 2005a), so it is impossible to determine if Cre $\Delta$ C had any role in the observed cleavage (Fig. 5.7). Much weaker cleavage activity was observed with the Cre $\Delta$ C/Y324F mixture on the phosphorothiolate substrate. This was also not surprising since Cre $\Delta$ C has weak activity on this substrate (Fig. 5.8). Although the effect was small, the amount of cleavage product decreased as the ratio of Cre $\Delta$ C increased. Since Y324F is unable to cleave this substrate alone, Y324F must stimulate Cre $\Delta$ C activity. The maximum stimulation was observed when the ratio of Cre $\Delta$ C:Y324F was 1:3. This suggests that Y324F may stimulate the formation of synapsis where the active complex is a tetramer of Cre consisting of three Y324F monomers and one Cre $\Delta$ C monomer (Fig. 5.8A model VI). The stimulatory effect is small because Y324F is partially defective in synapsis. If the formation of synapsis is the cause of observed increase in Cre $\Delta$ C activity, then it is difficult to identify the role of hN in this experiment.

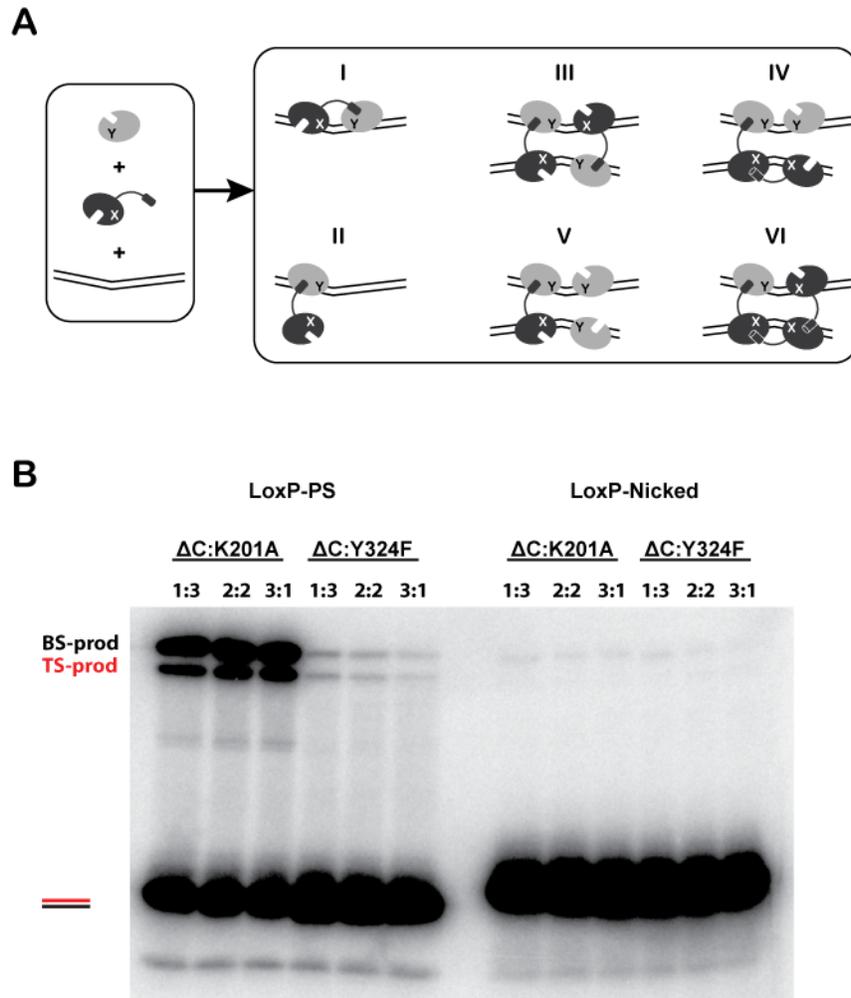
To eliminate some of the issues of the complementation experiment and isolate the effect of hN docking into the hydrophobic pocket, a peptide of hN was used in place of active site mutants. If hN binding in the pocket is required for the formation of a competent active site, then it should

be possible to provide the helix *in trans* as a peptide, assuming that the affinity is high enough for binding to occur. Since this 13 amino acid peptide contains no aromatic residues, the 83% pure dry peptide was dissolved in water the concentration was measured using absorbance at 214 nm (Kuipers & Gruppen, 2007). The peptide was added in 200 fold excess (>100uM) to a solution of 0.5  $\mu$ M Cre $\Delta$ C and then measured for cleavage activity. The high concentration of peptide was used to promote binding, though binding of the peptide to Cre was not experimentally measured at this concentration. No stimulation of cleavage was observed when the hN peptide was added to reactions with either nicked or phosphorothiolate DNA suicide substrates (Fig. 5.10). The inability of Y324F or K201A to complement  $\Delta$ C on the nicked substrate shows that hN binding to the  $\Delta$ C pocket is not sufficient to activate  $\Delta$ C subunit.

#### **5.4.2 Helix-N as a tether for Y324**

Crystal structures of Cre tetrameric complexes bound to *loxP* show that hN binding in the hydrophobic pocket is nearly identical for active and inactive subunits, but the orientation of hM and the linker connecting hM and hN are quite different (Ghosh *et al.*, 2007; Guo *et al.*, 1997; Guo *et al.*, 1999). In active subunits, hN exchanges across the dimer interface and the hM-hN linker region forms an extended conformation with a N327-E331 C $\alpha$  distance of 11 Å. The inactive subunits of Cre exchange hN as part of the synaptic interface and have a compressed linker with a N327-E331 C $\alpha$  distance of 7 Å. Y324 and hM are also in different positions for active and inactive subunits (Fig. 5.11).

If Cre uses a tethering mechanism to control activity through the orientation of Y324, then changing the length of the hM-hN linker between Y324 and hN should disrupt the tether and have an affect on activity. A series of Cre mutants containing one, two or three glycine insertions before hN at E331 were constructed and purified (pGV1014, pGV1015, and pGV1015; see table 5.1).



**Figure 5.8** *ΔC* complementation assay

$\Delta$ C(Gray) containing an intact active site is mixed at varied ratios with active site mutants Y324F and K201A (Black) that are incapable of cleavage but carry hN.

A) Cartoon depicts possible interactions formed in solution.

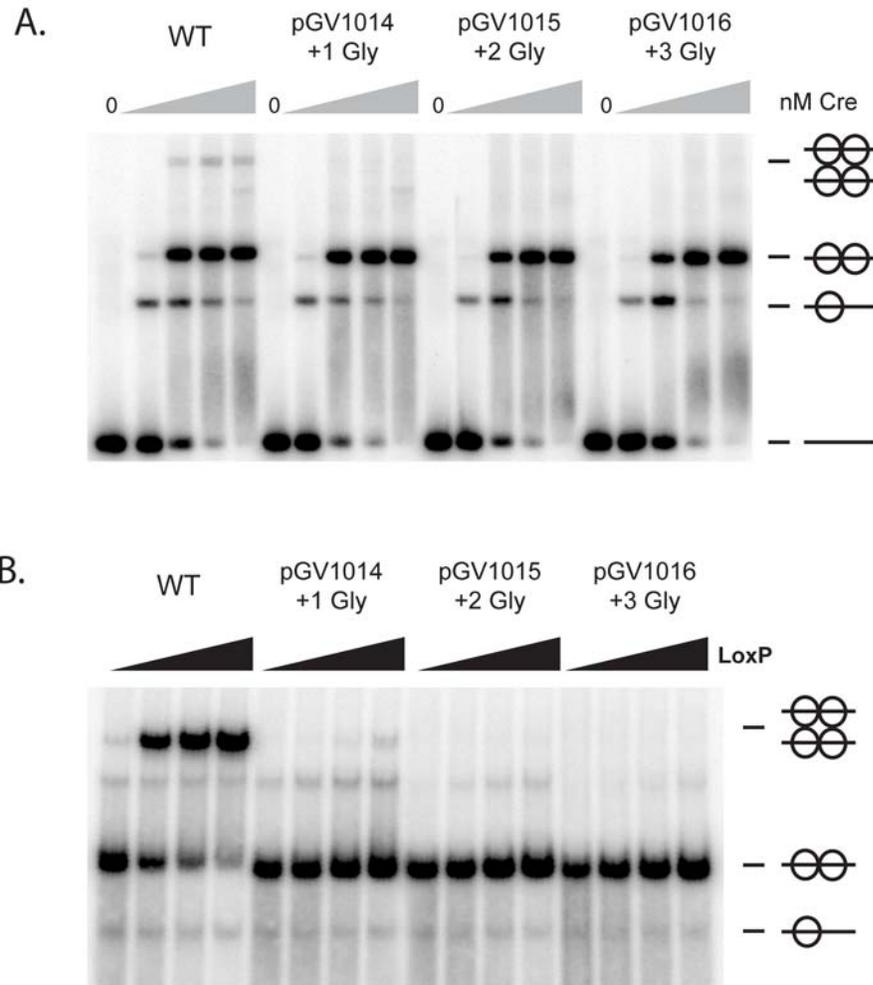
B) Cleavage reactions were incubated for 60 min at concentrations favorable for synapsis (500 nM Cre/50 nM DNA substrate) with either phosphorothiolate substrate(PS) (left) or nicked suicide substrate (right). Note: K201A has near WT cleavage activity on PS substrate. Cleavage product for  $\Delta$ C:Y324F on PS-substrate decreases as a function of Y324F indicating that a tetrameric complex similar to VI (A) may form and have greater activity than others. No cleavage was observed on nicked suicide substrate. hN donated from K201A or Y324F to Cre $\Delta$ C is unable to rescue the  $\Delta$ C cleavage defect.

The glycine insertion mutants were all found to be inactive for recombination *in vivo* (data not shown). In order to determine which step in recombination is disrupted by the glycine insertions, the mutants were tested for binding, synapsis and cleavage.

The Cre glycine insertion mutants are able to bind *loxP* cooperatively, whereas Cre $\Delta$ C is defective in cooperative binding to *loxP*. A minor difference in cooperativity is observed for the glycine insertions that increases with the length of the hM-hN linker (Fig. 5.9A). All three glycine insertion mutants are severely defective in synapsis formation. A minimal amount of product is visible for the Cre mutant containing one glycine, but no synapsis is visible for the Cre mutants with two or three glycines (Fig. 5.9B). These results suggest that the longer linker between hM and hN is more disruptive to the interactions formed by Cre at the synaptic interface than at the dimeric interface.

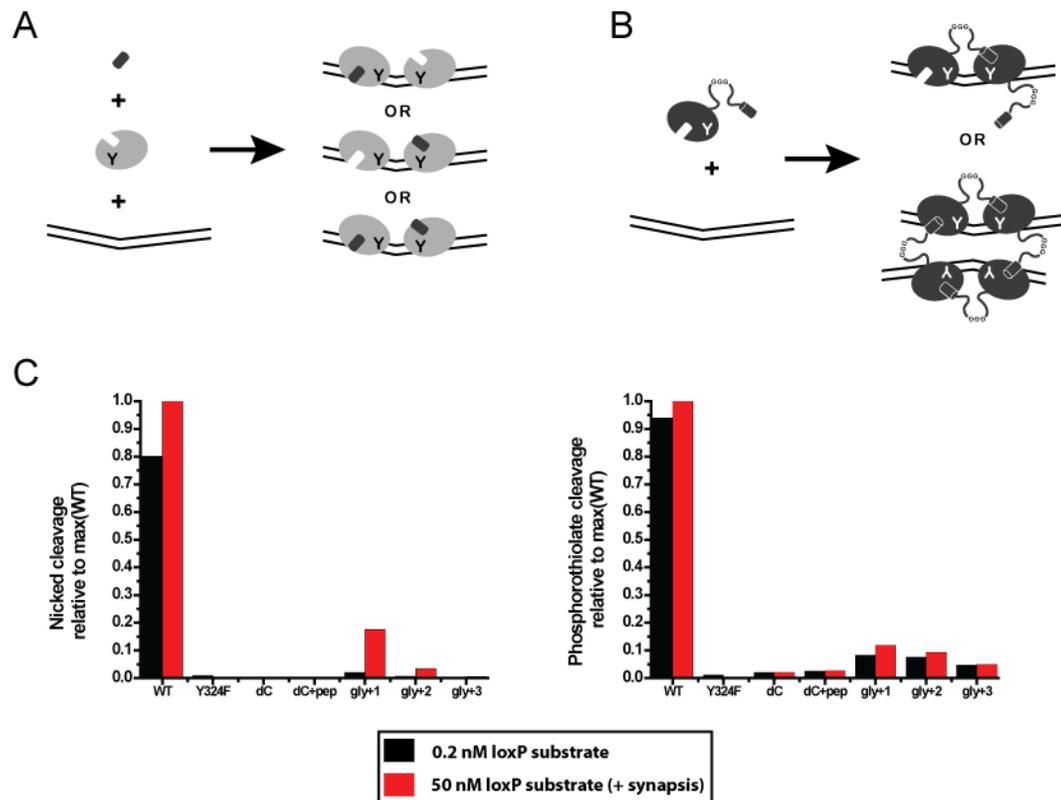
The Cre glycine mutants were then tested for cleavage activity (Fig. 5.10). Using a nicked suicide substrate, the mutants show cleavage activity that decreases with increasing linker length. At conditions unfavorable for synapsis, the cleavage product is barely detectable for the +2 and +3 glycine mutants. As expected, these mutants show little stimulation when cleavage is performed under conditions favorable for synapsis. The +1 glycine mutant has a modest increase in cleavage activity followed by a nearly undetectable increase in cleavage activity of the +2 glycine mutant with no increase observed for the +3 glycine mutant. When cleavage was measured using the phosphorothiolate suicide substrate, all glycine insertion mutants were deficient in cleavage compared with WT Cre, with all three glycine insertion mutants displaying a similar extent of cleavage and little to no stimulation when assayed in conditions favorable for synapsis. All three glycine insertion mutants were better able to cleave than Cre $\Delta$ C. In the absence of synapsis, the glycine insertion mutants are severely deficient in cleavage, directly demonstrating a role for hN through the hM-hN linker in positioning Y324 for cleavage.

The glycine insertions disrupt synapsis formation without having a significant effect on the ability to form the dimeric complex. This contrasts with  $\Delta C$ , which is defective in both cooperative binding and synapsis. In the WT Cre-DNA tetramer, the linker at the synaptic interface is compressed relative to the linker at the dimer interface (Fig. 5.11). The compressed linker at the synaptic interface may not be able to accommodate the additional amino acids, or may prevent necessary synapsis contacts from forming.



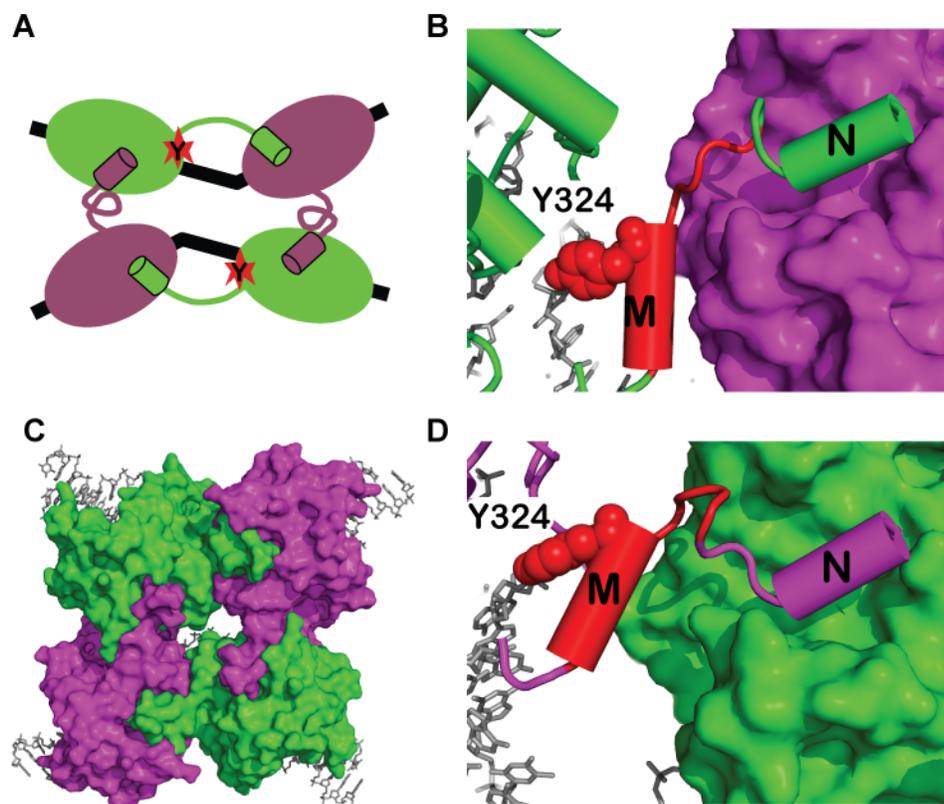
**Figure 5.9** *Binding and Synapsis of Glycine Insertion Mutants*

A. Binding of Cre to 200pM 64 bp loxP (loxSS3/6). Cre concentrations tested, 0, 0.5, 5, 50, 500 nM. B. Synapsis by Cre as a function of loxP (loxSS3/6) at constant Cre concentration of 500nM. LoxP concentrations tested were, 0.2, 10, 40, 160 nM.



**Figure 5.10 Cleavage activity of glycine insertion mutants**

Cleavage activity after 60 mins of  $\Delta C$  with hN peptide and glycine insertion mutants. Assays performed using nicked suicide substrate (left) and phosphorothiolate suicide substrate (right). Reactions performed at 0.2nM substrate and 50nM substrate to measure effects of synapsis on cleavage. Activity reported as the ratio of mutant cleavage to WT at 50nM for each substrate. hN peptide is unable to stimulate  $\Delta C$  for cleavage. The glycine insertion mutants are severely defective in cleavage with a trend of gly+1>gly+2>gly+3. Synapsis is partially able to rescue gly+1 cleavage defect, most notable on the nicked suicide substrate. The severe deficiency of all these mutants indicate a role for hN in positioning Y324 for activity.



**Figure 5.11** *The hM-hN linker conformation in active and inactive Cre subunits*

A) A cartoon representation of Cre-DNA tetramer highlighting hN. B) Surface representation of Cre-DNA tetramer viewed from the C-terminal domain. hN of active and inactive subunits bind the receptor pocket in a similar manner, however the linker of active subunits of Cre (green, shown in C), is in an extended conformation relative to the linker of the inactive subunit (purple, shown in D). The conformation of the linker alters the position of hM (red) containing the tyrosine nucleophile, Y324 (red spheres).

*Table 5.1: Summary of activities for mutants characterized in study of hN*

<b>Description</b>	<b>ID</b>	<b>Binding</b>	<b>Cooperativity</b>	<b>Synapsis</b>	<b>Cleavage</b>	<b>Recombination</b>
WT	pGV15	+++	+++	+++	+++	+++
Y324F	pGV31	+++	+++	++	0	0
$\Delta$ C		++	+	0	0	0
+1 Gly	pGV1014	+++	+++	+	+	0
+2 Gly	pGV1015	+++	+++	0	+	0
+3 Gly	pGV1016	+++	++	0	+	0

## 5.5 Discussion

### 5.5.1 *Assembly of an active Cre-loxP complex*

Crystal structures of the Cre-loxP synaptic complex (Ghosh *et al.*, 2007; Guo *et al.*, 1999), covalent intermediate (Guo *et al.*, 1997), HJ intermediate (Gopaul *et al.*, 1998) and the transition state mimic (this work, Chapter 2) suggest that the C-terminus of Cre has a role in the assembly of the synaptic complex and the regulation of activity. The studies presented here shed further light on these roles.

Determination of the Cre half site structure captured the first view of Cre in a pre-synaptic state (Yuan, 2008). The structure and position of the C-terminus was a primary interest in capturing this structure. Despite attempts to prevent intermolecular exchange of hN, Cre monomers were able to exchange hN *in trans* with a neighboring Cre monomer in the crystal lattice. The interactions formed by hN in the half site structure were similar to those observed in the tetrameric complexes. However, the hM-hN linker was in an entirely different conformation in the half site structure to allow this docking of hN, which demonstrated the flexibility of the linker. The structure is consistent with crosslinking studies which were unable to detect hN of the Cre monomer binding *in cis* (Yuan, 2008). This suggests that upon Cre binding to DNA, hN is poised to bind in the hydrophobic pocket of a neighboring Cre monomer. However, the structure does not explain how hN of the Cre monomer is oriented or where it is binding in solution. The Cre monomer bound to DNA is unable to satisfy intermolecular interactions. The demonstration that Cre is unable to cleave linear substrates as a monomer indicates that these intermolecular interactions are required for activity. This result is consistent with the observation that Cre does not function as a type I topoisomerase under normal conditions (Abremski *et al.*, 1986).

Cre cooperatively binds *loxP* to form the dimeric complex. Since there are no crystal structures of the Cre dimer bound to DNA, structural information about this state is inferred from the Cre-DNA structures of tetrameric complexes and the Cre half site structure (Ghosh *et al.*, 2007; Guo *et al.*, 1999; Yuan, 2008). The synaptic complex is two-fold symmetric where each half corresponds to a Cre dimer bound to *loxP*. The Cre dimer interface is composed of extensive interactions between N-terminal and C-terminal domains, which includes hN of the active subunit exchanged *in trans* with the inactive subunit. hN of the inactive subunit is donated across the synaptic interface into a hydrophobic pocket of an active subunit of the second Cre bound dimer. In solution, an isolated Cre dimeric complex is unlikely to have this precise organization. One possibility is that both Cre subunits in the dimer participate in a reciprocal exchange of hN. Modeling suggests that the hM-hN linker is of insufficient length to accommodate such an interaction without partial unfolding of hM. Synapsis also likely aids in the distinctive asymmetric bend of *loxP* that is observed in the Cre tetrameric-DNA complex structures. If so, then the Cre dimeric complex will likely experience greater conformational freedom in bending of *loxP* where the average state is less bent than observed in the crystal structure of the synaptic complex. Although significant differences are likely to exist between the dimeric complex observed as half of the Cre-DNA tetrameric complex crystal structure and the dimer complex in solution, the cooperative binding of Cre to *loxP* signifies that contacts responsible for the Cre dimer interface in the synaptic complex are also important for formation of the dimer.

The intermolecular contacts formed by the dimeric complex are sufficient to activate Cre for cleavage on an isolated *loxP* site. Cleavage product was detected at *loxP* concentrations 5000x below the  $K_D$  for synapsis (10nM) and remained nearly constant over a 100 fold increase in *loxP* concentration before increasing in a manner consistent synapsis formation. At saturating concentrations of Cre, synapsis is a function of *loxP* concentration. Using a previously published model that describes this relationship (Ghosh *et al.*, 2007), the fraction of synapsis formed in the

cleavage reactions at *loxP* concentrations below 10 pM is less than 0.1%. Data from assays using nicked suicide and phosphorothiolate suicide substrates agree well with a conclusion that Cre is capable of cleaving *loxP* as a dimer in the absence of synapsis.

Unexpectedly, wildtype Cre displayed greater activity on the top strand (AT half site) of the *loxP* phosphorothiolate substrate than on the top strand of the nicked *loxP* substrate at concentrations of *loxP* well below the  $K_D$  of synapsis. This difference was only observed on the top strand, as bottom strand (GC half site) cleavage products were similar for both substrates. This effect appears to be isolated to cleavage at low concentrations of *loxP*, since wildtype Cre displays a faster rate of cleavage on nicked *loxP* substrates than the phosphorothiolate substrates at conditions favorable for synapsis (see Chapter 4: thio-effect experiments). The phosphorothiolate substrate is unique because it abrogates the need for a general acid during cleavage. This suggests that the active site of the cleaving subunit in the dimeric complex is not the same as the active site of the tetrameric complex. Cre active site residues K201 and R173 are involved in general acid catalysis and slight variations in their position will have a significant impact on the ability of Cre to cleave the phosphodiester linkage of the nicked substrate but less impact on the phosphorothiolate substrates because this substrate does not require a general acid (See Chapter 4). The effect was observed only for top strand cleavage, which suggests that in the absence of synapsis, the active site is primed for cleavage on the top strand but lacks the general acid to catalyze the cleavage reaction.

The Cre A36V mutant was included in these experiments as a control for synapsis. A36 is located on a loop at both the dimeric and synaptic interfaces in the tetrameric Cre structures. The A36V mutation caused a severe defect in synapsis without affecting binding. Since A36V does not synapse, it offers an excellent control for synapsis independent cleavage. Indeed, cleavage by A36V remained at a low level for both top and bottom strands of phosphorothiolate substrate over the entire concentration range studied, indicating that no stimulation from synapsis was occurring. However, cleavage of the top strand by Cre A36V was much lower than cleavage of

the top strand by WT Cre at the lowest concentrations measured. This might suggest that synapsis is the cause of the high level of top strand cleavage for WT Cre. Alternatively, A36V may disrupt the dimeric interface in ways that are not detectable in a binding experiment. From the structure, A36 is at the center of the N-terminal intermolecular interface. Mutation to valine will certainly disrupt this interaction potentially causing a defect in the dimeric complex as well as the documented defect of synapsis (Hoess *et al.*, 1987; Wierzbicki *et al.*, 1987). The active site of the cleaving subunit of wild type Cre in the dimeric complex is already partially compromised (See Chapter 4), so any disruption of the dimeric interface will likely further reduce activity. Therefore, A36V may not be an appropriate control for synapsis independent cleavage if this mutant also disrupts the dimer interface that has already been shown to be crucial for Cre activity.

All tyrosine recombinases mediate recombination through a tetrameric complex similar to that of Cre, but the assembly and activation of the recombinase tetramers differ. For simple systems such as Cre, tetramer formation through synapsis results in activation of Cre for recombination, and depends solely on Cre and *loxP* (Van Duyne, 2008). More complex systems such as  $\lambda$ -integrase and XerCD employ accessory domains or rely on interactions with additional proteins or regions of DNA for activation and control of recombination. Although recombination is tightly regulated in these complex systems, the cleavage activity of pre-tetrameric states is not. For example,  $\lambda$ -integrase, XerC and XerD have been shown to cleave their recognition sites as monomers and were found to have topoisomerase I like activity (Cornet *et al.*, 1997; Craig & Nash, 1983; Kazmierczak *et al.*, 2002; Kikuchi & Nash, 1979; Lee, Aihara, Ellenberger & Landy, 2004; Tekle *et al.*, 2002). Flp recombinase is unable to cleave DNA as a monomer under normal conditions, but this is in large part due to the trans cleavage mechanism that requires two subunits to construct a complete active site (Lee *et al.*, 1994).

Simple recombinase systems such as Cre appear to more tightly regulate cleavage activity than more complex recombinase systems such as  $\lambda$ -integrase and XerC and D (Azaro & Landy, 2002; Barre & Sherratt, 2002; Van Duyne, 2008). The complex systems have evolved a

dependence on accessory regulatory elements to control recombination activity. Cre and Flp do not require accessory regulatory elements in recombination, so all regulatory control of activity is provided by the protein and recombination site. Attempts to build  $\lambda$ -like regulation into Cre have been successful, but only after Cre binding was disrupted by altering the Cre binding element in *loxP* and adding an extra base in the crossover region to make *loxP* more like the  $\lambda$ -integrase recognition site, *attP* (Warren, Laxmikanthan & Landy, 2008). These data support a model where tyrosine recombinases with extra regulatory elements such as  $\lambda$ -int and XerCD, evolved from a common precursor like Cre. If such an evolutionary path did exist, perhaps recombinases such as  $\lambda$ -int and Xer were originally more tightly controlled to prevent topoisomerase activity as monomers similar to Cre. Incorporation of accessory regulatory elements comes at a price for these systems. In order to become sensitive to regulatory cofactors, the recombinases have weakened protein:protein interfaces and as well as binding to their recognitions sites causing them to become somewhat leaky.

### 5.5.2 Regulation of Cre activity by the C-terminal helix

Biochemical studies of Cre  $\Delta$ C showed that this mutant is defective in cooperative binding, synapsis and cleavage. This is consistent with results showing that mutation of G333 of hN to tryptophan, arginine or glutamic acid, caused severe defects in recombination without disrupting binding (Wierzbicki *et al.*, 1987). The C-terminal domains of tyrosine recombinases are not well conserved (Fig. 5.2), and not surprisingly the C-terminal deletions of other tyrosine recombinases behave differently than those in Cre.

The C-terminal truncation mutant in  $\lambda$ -int, W250ter, removed the last six amino acids and has greatly reduced recombination activity but enhanced cleavage and topoisomerase activities (Kazmierczak *et al.*, 2002; Lee *et al.*, 2004; Tekle *et al.*, 2002).  $\lambda$ -int is unique in that it is the only system for which structures have been determined for the unbound protein (Kwon *et al.*, 1997), bound monomer (Aihara *et al.*, 2003b), and tetrameric complexes (Biswas *et al.*, 2005).

The C-terminus adopts a different conformation in each of these structures. The equivalent region of hN from Cre forms a beta strand ( $\beta 9$ ) in  $\lambda$ -int. The beta strand binds against the conserved  $\beta$ -sheet structure containing the catalytic lysine in tyrosine recombinases. This is quite different in Cre, where hN binds in a primarily  $\alpha$ -helical hydrophobic pocket. The tetrameric structures reveal that the C-terminal  $\beta$ -strand is exchanged with a partner subunit *in trans* similar to Cre and Flp (Biswas *et al.*, 2005; Chen & Rice, 2003b; Van Duyne, 2001) (Fig. 5.3). The structure of unbound  $\lambda$ -int reveals that  $\beta 9$  binds the  $\beta$ -sheet *in cis* (Kwon *et al.*, 1997), which differs from crosslinking studies in Cre that were unable to detect hN binding *in cis* (Yuan, 2008).  $\beta 9$  in the half site structure binds *in trans* through a crystallographic neighbor much like the Cre half site structure (Aihara *et al.*, 2003b). Mutations in the C-terminal  $\beta$ -strand (I353M and I353A) or at the receptor strand (A241V) of  $\lambda$ -int disrupt the interactions formed by the C-terminus and display a phenotype similar to the C-terminal deletion mutant, W350ter (Kazmierczak *et al.*, 2002; Tekle *et al.*, 2002). The structures and biochemical characterization of the W350ter, I353M and A241V mutants suggest that the C-terminal region of  $\lambda$ -int functions as a conformational switch that acts to prevent cleavage in the absence of a tetramer and stimulates strand exchange and Holiday Junction isomerization in the context of the tetrameric complex during recombination (Aihara *et al.*, 2003b). However, the mechanism of regulation by the C-terminus of  $\lambda$ -int is likely different from Cre (Kazmierczak *et al.*, 2002).

Sequence and structural comparisons between the unbound XerD structure to various Cre tetrameric structures suggest that the C-termini of XerC and XerD should resemble Cre more than  $\lambda$ -int (Hallet *et al.*, 1999; Spiers & Sherratt, 1999; Subramanya *et al.*, 1997). The C-terminal region of XerC is five amino acids longer than XerD, with an appended sequence 'KRGK' (Fig. 5.2). C-terminal deletion mutants lacking the last 5 or 10 amino acids of XerD are generally inactive for recombination but retain topoisomerase activity and cooperative binding with XerC (Spiers & Sherratt, 1997). The XerC $\Delta 5$  mutant (which has an equivalent C-terminal length as XerD) was capable of recombination, while XerC $\Delta 10$  was inactive for recombination. Both XerC deletion mutants retain topoisomerase activity, but are diminished in cooperative binding with

XerD (Ferreira *et al.*, 2003; Spiers & Sherratt, 1999). The results here suggest a role for the C-terminal ends of XerC and XerD in recombination and that the C-terminus of XerC has a larger role in cooperative interactions with XerD to form the dimer than does the C-terminus of XerD with XerC. Interestingly, the interaction with regulatory protein FtsK has been mapped to the C-terminus of XerD where it stimulates XerD for cleavage at recombination site *dif* (Yates *et al.*, 2006).

Examination of the tetrameric Cre structures offered two models for functional control of Cre activity by hN. Although hN binds the pocket of active and inactive subunits in a similar manner, the hM-hN linker has an altered conformation. In the active subunit this hM-hN linker adopts an extended conformation, while the linker in the inactive subunit adopts a compressed conformation. This has led to the suggestion that hN binding in the active subunit communicates with hM through the linker to position Y324 in an active conformation (Gopaul *et al.*, 1998; Van Duynes, 2001). Alternatively, hN binding may stabilize the active site of the cleaving subunit. The hydrophobic pocket of Cre that hN binds in consists primarily of  $\alpha$ -helices. H289 and R292 are located on  $\alpha$ K with W315 in a loop between  $\alpha$ L and  $\alpha$ M. The interactions formed by hN binding may be necessary to form a competent active site.

To examine if hN is necessary for a competent active site, a series of experiments were performed with Cre $\Delta$ C where hN was provided as a peptide or from an active site mutant incapable of cleavage (K201A or Y324F). In both cases, hN was unable to stimulate  $\Delta$ C for cleavage suggesting that hN binding the hydrophobic pocket was insufficient to activate Cre for cleavage.

Glycine insertions in the hM-hN linker were constructed to determine if hN binding communicates with Y324 to adopt an active position. All three glycine insertion mutants were competent for cooperative binding, but defective for synapsis. This contrasts with Cre  $\Delta$ C, which is defective in both cooperative binding and synapsis, suggesting that exchange of hN is critical in cooperative binding at the dimer interface and is tolerant of the insertions in the linker. The

synaptic interface requires hN as well, but the interface is not able to tolerate the insertions in the linker. Examination of Cre synaptic structure reveals that hM-hN linker adopts an extended conformation across the dimer interface while forced to compress at the synaptic interface (Ghosh *et al.*, 2007; Gopaul *et al.*, 1998; Guo *et al.*, 1997; Guo *et al.*, 1999). Since the hM-hN linker is compressed at the synaptic interface, perhaps the insertions in the hM-hN linker are unable to be accommodated due to potential steric clashes with a synaptic partner.

Increasing the hM-hN linker by a single glycine results in more than a five fold drop in cleavage activity. Extending the linker by two or three glycines results in mutants that are essentially inactive for cleavage. Therefore, the hM-hN linker has evolved a specific length to function as a sensor. hN binding in the hydrophobic pocket of a neighboring subunit on the same *loxP* site communicates a change in position of Y324 on hM that is permissive for cleavage activity. This demonstrates a dual role for hN and the associated linker by directly affecting cleavage in the active site through modulation of the position of Y324 as well as being essential in the formation of synapsis to permit recombination.

The regulation of hN in Cre may be similar to the C-terminal regions of XerC and XerD. Studies in XerC and XerD relied on Cre structures to aid in the identification of regions responsible for a C-terminal interaction in a Cre-like hydrophobic pocket (Hallet *et al.*, 1999; Spiers & Sherratt, 1999). The C-terminal regions and receptor pockets in XerC and XerD were found to exhibit recombinase specificity where the C-terminal region of XerC will only bind in the pocket of XerD and vice versa (Ferreira *et al.*, 2003; Hallet *et al.*, 1999). On DNA substrates where XerC is favored for cleavage, a C-terminal deletion mutant of XerC( $\Delta$ 289-298) disrupts cleavage by XerC with an observable stimulation of cleavage by XerD. However, when wildtype XerC is mixed with a C-terminal deletion of XerD ( $\Delta$ 294-298) on the same substrate no change in cleavage is observed (Ferreira *et al.*, 2003). This suggests that like Cre, the donation of the C-terminal region in the cleaving subunit is more important than receiving a C-terminal region to stabilize the active site. When the equivalent of the hM-hN linker region was shortened in either

XerC or XerD, the recombinase with an unperturbed C-terminal region was activated for cleavage (Ferreira *et al.*, 2003). This demonstrated that like Cre, the length of the linker region connecting the catalytic tyrosine nucleophile with the C-terminal helix has evolved a specific length and is critical for activity.

The C-terminal regions of tyrosine recombinases are poorly conserved despite close proximity to the catalytic tyrosine. However, structures of tyrosine recombinases demonstrate that the C-terminal regions are intimately involved in protein:protein interactions that form the dimer and tetramer complexes and are important for coordinating subunit activity in recombination. Flp recombinase is unique from most tyrosine recombinases in that the helix containing the catalytic tyrosine is exchanged with the partner *in trans* and the C-terminus returns to bind *in cis*. For other tyrosine recombinases, the functional consequence of mutation or deletion in the C-terminal region vary considerably, suggesting that the mechanism of regulation in recombination has not been conserved in the family. In Cre, hN is an essential regulatory switch that mediates formation of the dimer and tetramer and directly impacts catalytic activity by positioning the catalytic tyrosine for cleavage. The importance of hN in regulating Cre activity relative to equivalent C-terminal domains in  $\lambda$ -int and XerCD agrees well with the previous conclusion that complex tyrosine recombinase systems reliant on accessory regulatory elements have reduced built-in regulation that mediate activities such as cleavage and cooperative binding in favor of regulating recombination.

## **5.6 Chapter specific methods**

### **5.6.1 *Cloning and Cre purification***

Method details described in Chapter 3

### **5.6.2 *EMSA assays***

Binding and synapsis assays performed as described in Chapter 3.

### **5.6.3 *Concentration dependent cleavage as dimer***

Assay as described in Chapter 3 except the concentration of radiolabeled substrate was held fixed at a low concentration 2 pM (nicked suicide substrate; loxSS4/5/6 and loxSS1/2/3) or 6 pM (5'-bridging phosphorothiolate substrate; loxSS7-8/9-10) and spiked with unlabeled substrate to fill out the concentration range.

### **5.6.4 *Complementation cleavage assay***

$\Delta C$  was premixed with either K201A or Y324F at ratios of 1:3, 2:2, and 3:1. The cleavage reaction was initiated by addition of a master mix containing assay buffer, radiolabeled and cold DNA substrate. Final assay concentrations were 1  $\mu M$  Cre, 0.2nM radiolabeled substrate, and 50 nM cold substrate. The procedure was performed with both nicked and phosphorothiolate DNA substrates (loxSS4/5/6 and loxSS7-8/9-10).

### **5.6.5 *hN peptide cleavage***

The C-terminal 13 amino acid peptide corresponding to hN was ordered from Genscript and determined to be 83% pure by mass spectral analysis. 500 mg of hN peptide was dissolved in 200ul of dH<sub>2</sub>O and confirmed soluble by measurement of backbone absorbance at 214 nm (Kuipers & Gruppen, 2007). hN peptide was added to a standard one hour cleavage assay reaction with  $\Delta C$  at a final concentration of approximately 120  $\mu M$ . Reactions performed with both nicked BS and phosphorothiolate substrates, with and without extra cold DNA to stimulate formation of synapsis.

## Chapter 6. Conclusions and Perspectives

### 6.1 Summary of conclusions

Decades of genetic, biochemical and structural studies have advanced the understanding of site-specific recombination mediated by YRs and the mechanistic relationship to TopIB enzymes. Cre and a handful of other YRs have been at the forefront of most of this research (Van Duyne, 2008). The relative simplicity of Cre-*loxP* recombination compared to the more complicated YRs systems such as  $\lambda$ -integrase and XerC/D, has made it an excellent model for understanding the mechanism of recombination (Sauer, 2002). The simplicity of the system has proved extraordinarily useful the application of Cre as a tool in genomic manipulations *in vitro* and *in vivo*. However, the simplicity of the system also exposes at least three weaknesses when used in this context. First, Cre has a predilection for the excision reaction rather than integration. Therefore, Cre is less efficient when used to introduce genes rather than catalyzing their removal. Secondly, Cre-*loxP* recombination is a reversible reaction that rarely converts 100% of substrate to product creating a mosaic effect. Lastly, Cre is as a single protein that recognizes a specific sequence, restricting the number of simultaneous DNA manipulations that can be performed (Buchholz & Stewart, 2001). Therefore, the generation of novel Cre-like recombinases with unique specificity, improved directionality and altered reaction preference will have a tremendous impact on the application of site-specific recombinases as genomic tools. Although the engineering of Cre has been partially successful (Buchholz & Stewart, 2001; Livet *et al.*, 2007; Sarkar, Hauber, Hauber & Buchholz, 2007), a fundamental understanding of the requirements for catalysis and the regulatory mechanisms that control activity is a prerequisite for any rational protein design. The work described in this dissertation builds on the fundamental understanding of the requirements for Cre activity.

## 6.2 Tolerance to change in the active site

The construction and characterization of the library of active site mutants discussed in Chapter 2 is the most comprehensive compilation of activities of any YR or TopIB system. The *in vivo* recombination assay proved to be a powerful tool to characterize Cre mutant excision activity. Of the seven conserved active site residues in YRs, H289 and E176 are the most tolerant to change in Cre. These residues have no corresponding residue in TopIBs. The understanding of these results was facilitated by the determination of a Cre-*loxP*-vanadate transition state mimic complex structure, which is analogous to similar intermediates captured in TopIBs (Davies *et al.*, 2006; Perry *et al.*, 2010). The active sites of Cre and TopIB in the transition state are essentially superimposable, with the lone exceptions of H289 and E176. In this structure H289, interacts with the phenolic oxygen of Y324 and the non-bridging oxygen of O2P (Fig. 2.2), which had led to speculation that the histidine functions as a base responsible for activation of the tyrosine nucleophile (Van Duyne, 2001; Whiteson *et al.*, 2007). Fourteen different substitutions at 289 in Cre were active, which argues against H289 having an essential catalytic role. E176 accepts hydrogen bonds from K201 and R173, which are two residues involved in general acid catalysis. In TopIB, the corresponding spaces are filled by two well ordered water molecules that mediate a similar set of interactions. The similarity in position of the other shared active site residues, highlights the evolutionary relationship in phosphoryl transfer mechanism between YRs and TopIB.

## 6.3 H289 accelerates water mediated hydrolysis

YR-DNA complex structures and biochemical studies of several point mutants in F1p implicated the conserved histidine in general base catalysis that activates the tyrosine for nucleophilic attack on the scissile phosphate (Pan, Luetke & Sadowski, 1993a; Whiteson *et al.*, 2007). The results from the *in vivo* characterization of mutants at H289 in Cre argues against

such a role. All nineteen single site mutants at H289 were purified and tested for *in vitro* recombination, cleavage, binding and synapsis. Most mutants with activity *in vivo* had correlating activities *in vitro*. The H289A mutant, which is the structural equivalent of TopIB has decreased activity. The conservative H289Q mutant was the most active with nearly 50% of WT Cre activity. In addition to H289Q, the H289L and H289M mutants accumulated covalent intermediates at levels higher than WT Cre. This suggested a possible deficiency in ligation, but the accumulation of covalent intermediate was not a general property of mutants at H289. Perhaps water is able to functionally replace the histidine in ligation as observed in TopIB. The H289L, H289M and H289Q mutants are excellent structural mimics for histidine that would exclude water from being able to effectively participate in ligation. Smaller amino acids at position 289 would allow water to access the scissile phosphate in a manner similar to that observed in TopIB and the H289A structure. If water is acting in place of these mutants, then they should produce a measurable solvent kinetic isotope effect with D<sub>2</sub>O.

The pH rate profile experiments on cleavage activity demonstrated that H289 is not required to function as a catalytic base to activate Y324 for nucleophilic attack on the scissile phosphate. Surprisingly, these pH experiments revealed that WT Cre efficiently hydrolyzes the 3'-phosphotyrosine covalent intermediate as a function of pH, but H289Q does not. In fact, none of the H289 mutants have any measurable hydrolysis activity even after 24 hours, where WT Cre had converted 100% of the substrate to hydrolysis product. The rate of hydrolysis by WT Cre was more than 600 times faster than TopIB, suggesting that H289 may be conserved to accelerate hydrolysis of the covalent intermediate. Although hydrolysis of the covalent intermediate would appear to be counterproductive, this activity may be a rescue mechanism for stalled recombination reactions. The resulting 3'-phosphate containing nick is then able to be corrected with host enzymes. The recombination mechanism is far more complex than relaxation, so YRs are more likely to stall as a covalent intermediate during recombination.

The use of substrates containing a methylphosphonate at the scissile position to study the rate of hydrolysis may be instructive to understanding the evolutionary relationship between Cre and TopIB. This substrate alleviates the needs for charge stabilization in the transition state at the methyl group, partially alleviating the need for Cre Arg292. In Cre and Flp, the rate of hydrolysis with the methylphosphonate substrates does not change, however TopIB experiences at  $10^3$  increase in the rate of hydrolysis (Ma *et al.*, 2009a; Ma *et al.*, 2007; Ma *et al.*, 2009b; Tian *et al.*, 2003). This effect may be explained in part by H289 in Cre, and the corresponding H305 in Flp. Without a corresponding histidine, Arg223 in TopIB could adopt a conformation similar to the histidine in Cre, capable by catalyzing hydrolysis. The histidine in Cre and Flp prevents the corresponding arginine from adopting a similar conformation. Cre H289A is unable to hydrolyze covalent intermediate of substrates containing normal phosphate (Chapter 3). However H289A should be capable of hydrolyzing covalent intermediate with the methylphosphonate substrate by a similar mechanism to that proposed for TopIB via the arginine. Likewise, the incorporation of histidine into the active site of TopIB by mutating K220 of *Vaccinia* topoisomerase, may accelerate the rate of hydrolysis of covalent intermediate using natural substrates.

## 6.4 E176 plays a structural role

To investigate the functional role the conserved Glu/Asp residue, I studied three mutants at E176 in Cre and found that E176A is better tolerated than either of the more conservative mutants E176Q and E176D. The biochemical properties of the E176Q mutant were identical to the catalytically defective R173K mutant. By measuring cleavage activity with 5'-bridging phosphorothiolate substrates for mutants at E176, R173 and K201, I proposed a model where E176Q disrupts the orientation of R173 such that the arginine displaces K201 from the active site through electrostatic and/or steric repulsion. The characteristics of the R173K mutant suggest a similar mode of disruption. This is supported based on modeling that suggests that the lysine is

too short to interact with O5' and too long to interact with O2P of the scissile phosphate of the transition state. This is also supported by the crystal structure of the R173K synaptic complex, where the  $\epsilon$ -nitrogen of K201 is oriented away from O5'.

TopIB has no equivalent Glu/Asp residue, but two well ordered water molecules form a similar set of interactions with K167 and R130. Investigation of TopIB structures shows that R130 is located on a loop that may play a role in orienting R130 in the active site. A Glu/Asp residue in a similar position in TopIB would likely help anchor R173 and limit the effectiveness of any regulatory role that the loop may have.

## 6.5 K201 as a regulatory catalytic switch

The concentration dependence of cleavage by mutants at E176, R173 and K201 with 5'-bridging phosphorothiolate substrates demonstrated that K201 has an additional role in stabilizing the scissile phosphate of the Cre-dimer bound to *loxP*. When cleavage was measured under conditions favoring synapsis, cleavage no longer required K201. The loop containing K201 in Cre forms synapsis dependent interactions and is displaced from the active site in the absence of these interactions. These data demonstrated a model for lysine as a regulatory switch that limits cleavage and possibly ligation activity in Cre until the synaptic complex forms. This highlights the structural role of E176, which will help orient the mobile lysine in the active site through the carboxylate hydrogen bonding interaction with the amide nitrogen of K201 and by restricting mobility of the R173 sidechain, which minimizes the electrostatic and/or steric repulsion at the  $\epsilon$ -nitrogen of K201.

This form of regulation is unique to Cre and possibly other YR systems because TopIB does not form higher order complexes. This region of YRs is variable, which suggests that other YR systems may rely a set of different interactions, possibly with accessory proteins to orient K201 in the active site. Disruption of the  $\beta$ 2- $\beta$ 3 loop containing K201 in Cre could help confirm this model by disrupting the interactions formed at synapsis.

## 6.6 Cre cleaves *loxP* as a dimer

Cre forms extensive protein:protein interactions in the synaptic complex burying nearly 3500 Å<sup>2</sup> of surface area (Van Duyne, 2001). Other YRs systems such as λ-int and Flp form less extensive interactions in the recombinase tetramer-DNA complex structures (Grindley *et al.*, 2006). Synapsis had been shown to have a stimulatory effect on cleavage activity in Cre, but it was unclear if synapsis was required for cleavage. Work performed with 5'-bridging phosphorothiolate substrates suggested that Cre could cleave as a dimer bound to *loxP*, however this substrate has reduced requirements, which may not reflect Cre activity on a natural *loxP* substrate. I demonstrated that Cre could cleave a natural phosphodiester linkage at substrate concentrations more than 1000-fold below the K<sub>D</sub> of synapsis. This demonstrated that Cre could cleave as a dimer bound to *loxP*. Cleavage as a dimer indicated that the protein:protein contacts formed by the Cre dimer are enough to support a low level of catalysis and indicates that K201 is able to populate the active site long enough to catalyze cleavage by the Cre-dimer without the interactions formed through synapsis.

## 6.7 The C-terminus of Cre regulates activity through Y324

The C-terminal helix (hN) of Cre is swapped to a neighboring monomer in the tetrameric-DNA complex structures of Cre. I demonstrated that hN is essential for Cre activity in cooperative binding, synapsis and cleavage. In the Cre dimer, hN of at least one monomer is able to bind the partner, which could explain the cleavage results described above. By altering the length of the hM-hN linker, I showed that hN regulates Cre activity by acting as a tether, which modulates the position of Y324. These insertions maintained cooperative binding at the dimer interface, but were defective in synapsis, which demonstrated the sensitivity of the linker length to the synaptic interface. An interesting question is whether or not Cre is able to cleave *loxP* as a monomer. Evidence suggests that Cre does not cleave as a monomer (Abremski *et al.*, 1986), which

contrasts with  $\lambda$ -integrase (Craig & Nash, 1983; Kikuchi & Nash, 1979) and XerCD (Cornet *et al.*, 1997), which were shown to cleave as monomers and have topoisomerase activity. This led to a model where Cre, and other 'simple' recombinases have tighter built-in regulatory mechanisms to control activity than these other complex systems. The complex YR systems rely on accessory proteins and DNA binding domains to regulate activity of the tetramer (Azaro, M A, Landy, A, 2002; Barre & Sherratt, 2002). In order to be able to respond to these accessory regulatory proteins, the 'complex' YR systems had to become less tightly regulated as monomers and dimers. This meant that the built in mechanisms like those described in Cre were not needed and likely in competition to the accessory factors.

## Appendix A. Cre Mutant Reference Table

Cre Res	System	Mutant	In vivo activity	In vitro activity	Binding	Cleavage	Ligation	HJ resolution	Notes (†)	References
<b>176</b>	P2(E197)	K	+							28
	Flp(D194)	G		0	0	0		0		26
	Flp(D194)	N		0	+	0	+	0	Poor binding; defective ligation	26,58
	R (Z.rouxii) (D194)	G		0	+	+		+	Defective HJ res_strand transfer rxn; WT binding	26
	Flp (D194)	Y	+	+	+					19,28
	Topo(G132)	D		0		+	+		Relaxation $10^{-2}$ / $kcl$ $10^{-3}$ / $klig$ $10^{-1}$	1
	Topo(G132)	S		+	+	+	+		$kcl$ $10^{-2}$ / $klig$ $10^{-2}$	1,2,3
Topo(G132)	A			+	+	+	+		2	
<b>173</b>	$\lambda$ int (R212)	Q	0	0	+	0		0		25,20,21
	Cre (R173)	C/K	0	+	+	0			0.2 WT recombination	24,27,34
	Flp(R191)	K	+	0	+	+	+	0	accum cov int./weak cleavage/ligation deficient	26,19,23,22,52
	Flp(R191)	Q/E		0	+	+	0	0	weak cleavage	26,58
	Flp + R (R191)	S/P		0	+	0	0	0		26,58
	Xer(R148)	K				+		+	Minimal cov formed on HJ sub	55
	Xer(R148)	C						+		55
	Topo(R130)	A		0	+	+	+		$kcl$ $10^{-5}$ / $klig$ $10^{-5}$	2,8,15,4
Topo(R130)	K		0	+	+	+		$kcl$ $10^{-4}$ / $klig$ $10^{-5}$	2,8,15,4	
<b>201</b>	Flp (K223)	A		0						43
	Flp (K223)	R		+					defective recombination	43
	Xer (K172)	A	+		+	0			defective strand exchange/ no relaxation	35,29
	Cre(K201)	A		0	+	0		0		31,32,33
	Topo(K167)	A		+	+	+	+		relaxation $10^{-2}$ / $kcl$ $10^{-4}$ / $klig$ $10^{-5}$	15,2,7,8,14
	Topo(K167)	R		+	+	+	+		relaxation 1/3 / $kcl$ $10^{-2}$ / $klig$ $10^{-2}$	2,7
	Topo(K167)	D		+	+	+	+		relaxation $10^{-1}$ / binding $10^{-1}$	5
Topo(K167)	E		+	+	0			relaxation $10^{-2}$	7	
<b>289</b>	$\lambda$ int (H308)	L	0	0	+					25
	Cre (H289)	Y	+	+	+				recombination ( <i>in vitro</i> $10^{-3}$ ; <i>in vivo</i> 20%)	24
	Flp (H305)	A				0				37
	Flp (H305)	Q	+	+	+	+			ligation defective	36,41,37
	Flp (H305)	L/P	0	0	+	+	+	+	defective ligation	36,44,40,22,41,52,59,60,58,37
	Cre (H289)	A,N		+				+	accum HJ, opp strand bias	38,39
	Xer(H244L)	L				+			buildup cov intermediate	55,56
	Tnpl(H234)	L		+		+			buildup cov intermediate	57
	Topo(K220)	A		+	+	+			<i>in vitro</i> relaxation 20% / cleavage 10%	9
	Topo(K220)	D		0	+	0				6
Topo(K220)	I,N		+	+	+				6	

## Appendix A. Continued.

Cre Res	System	Mutant	In vivo activity	In vitro activity	Binding	Cleavage	Ligation	HJ resolution	Notes (†)	References
<b>292</b>	λint (R311)	C/H	0	0	+	0		0		25,20,48,49
	Cre (R292)	C	0	+	+				recombination <i>in vitro</i> 10 <sup>-3</sup> / defective binding	24
	P2 (R272)	K	+	+						28
	Flp (R308)	G/P/Q	0	0	+	+	0	0		36,40,41,42,58
	Flp (R308)	K		0	+	+	+			22,42,52
	XerC(R243)/XerD(R247)	Q	0		+	+		+		30,55
	XerC(R243)/XerD(R247)	K				+		+	cov formed on HJ sub	55
	Topo(R223)	K	+	+	+	+			Kcl(eq) = >60% / relaxation and binding 10%/ cleavage 2%	9,10,17,15,14,18
	Topo(R223)	A		0	+	+	+		Kcl(eq) = 1/ kcl 10 <sup>-4</sup> / klig 10 <sup>-5</sup>	9,15,18
	Topo(R223)	E,G		0	+	0				6
Topo(R223)	Q			0					3	
<b>315</b>	Flp (W330)	A	0	0	+	0	+			43,54
	Flp (W330)	Q		0		+			kcl 10 <sup>-2</sup>	44
	Flp (W330)	H		+	+	+	+		defective recombination / kcl 10 <sup>-3</sup>	45,54
	Flp (W330)	F		+	+	+	+		WT recombination; cleavage 5x less	46,54
	Cre(W315)	H		+						54
	Cre(W315)	F		+						54
	Cre(W315)	A		0						54
	Xer(H270)	C				+				55
	Topo(H265)	A		+		+	+		Kcl(eq) = 5xWT / kcl 10 <sup>-2</sup> / klig 10 <sup>-2</sup>	10,17,15,14
	Topo(H265)	N,Q		+		+	+		Kcl(eq) = 5xWT	10
<b>324</b>	λint (Y342)	F	0	0	+	0		0		50, 51 49, 25
	Cre (Y324)	C	0	+	+					24
	Flp (Y343)	F	0	0	+	0		0		44,22,41,46,58
	Flp (Y343)	S	0	0	+	0		0		44,46,40,41,22
	XerC(Y275)/XerD(Y279)	F	0		+					30
	Topo(Y274)	A		0	+	0				4
	Topo(Y274)	F,K		0	+	0				11,12,15,16,13
	Topo(Y274)	C		0	+	+			Endo cleavage	13
	Topo(Y274)	H,E		0	+				Endo cleavage	13

† Activities relative to wildtype enzyme

## Appendix A. Continued.

System	Num	Ref
Topo	1	Morham SG, Shuman S. <i>Genes Dev.</i> 1990 Apr;4(4):515-24. PubMed PMID: 2163340.
Topo	2	Wittschieben J, Shuman S. <i>Nucleic Acids Res.</i> 1997 Aug 1;25(15):3001-8.
Topo	3	Morham SG, Shuman S. <i>J Biol Chem.</i> 1992 Aug 5;267(22):15984-92. PubMed PMID: 1322412.
Topo	4	Wittschieben J, Shuman S. <i>J Biol Chem.</i> 1994 Nov 25;269(47):29978-83. PubMed PMID: 7961997.
Topo	5	Gupta M, Zhu CX, Tse-Dinh YC. <i>J Biol Chem.</i> 1994 Jan 7;269(1):573-8. PubMed PMID: 8276853.
Topo	6	Klemperer N, Trakman P. <i>J Biol Chem.</i> 1993 Jul 25;268(21):15887-99. PubMed PMID: 8393454.
Topo	7	Petersen BO, Wittschieben J, Shuman S. <i>J Mol Biol.</i> 1996 Oct 25;263(2):181-95. PubMed PMID: 8913300.
Topo	8	Kroggh BO, Shuman S. <i>J Biol Chem.</i> 2002 Feb 22;277(8):5711-4. Epub 2001 Dec 27. PubMed PMID: 11756402.
Topo	9	Cheng C, Wang LK, Sekiguchi J, Shuman S. <i>J Biol Chem.</i> 1997 Mar 28;272(13):8263-9. PubMed PMID: 9079646.
Topo	10	Petersen BO, Shuman S. <i>J Biol Chem.</i> 1997 Feb 14;272(7):3891-6. PubMed PMID: 9020090.
Topo	11	Shuman S, Kane EM, Morham SG. <i>Proc Natl Acad Sci U S A.</i> 1989 Dec;86(24):9793-7.
Topo	12	Stivers JT, Shuman S, Mildvan AS. 1994 Jan 11;33(1):327-39. PubMed PMID: 8286354.
Topo	13	Wittschieben J, Petersen BO, Shuman S. <i>Nucleic Acids Res.</i> 1998 Jan 15;26(2):490-6.
Topo	14	Kroggh BO, Shuman S. <i>Mol Cell.</i> 2000 Jun;5(6):1035-41. PubMed PMID: 10911997.
Topo	15	Nagarajan R, Kwon K, Nawrot B, Stec WJ, Stivers JT. 2005 Aug 30;44(34):11476-85. PubMed PMID: 16114884.
Topo	16	Kwon K, Jiang YL, Song F, Stivers JT. <i>J Biol Chem.</i> 2002 Jan 4;277(1):353-8. Epub 2001 Oct 31. PubMed PMID: 11689573.
Topo	17	Tian L, Claeboe CD, Hecht SM, Shuman S. <i>Mol Cell.</i> 2003 Jul;12(1):199-208. PubMed PMID: 12887905.
Topo	18	Tian L, Claeboe CD, Hecht SM, Shuman S. <i>Structure.</i> 2005 Apr;13(4):513-20. PubMed PMID: 15837190.
Recombinase	19	Friesen,H. and Sadowski,P.D. (1992) <i>J. Mol. Biol.</i> , 225, 313–326.
Recombinase	20	Han,Y.W., Gumpport,R.I. and Gardner,J.F. (1994) <i>J. Mol. Biol.</i> , 235, 908–925
Recombinase	21	Segall,A.M. and Nash,H.A. (1996) <i>Genes Cells</i> , 1, 453–463
Recombinase	22	Pan,G., Luetke,K. and Sadowski,P.D. (1993) <i>Mol. Cell. Biol.</i> , 13, 3167–3175.
Recombinase	23	Kulpa,J., Dixon,J.E., Pan,G. and Sadowski,P.D. (1993) <i>J. Biol. Chem.</i> , 268, 1101–1108.
Recombinase	24	Wierzbicki,A., Kendall,M., Abremski,K. and Hoess,R. (1987) <i>J. Mol. Biol.</i> ,195, 785–794.
Recombinase	25	MacWilliams,M.P., Gumpport,R.I. and Gardner,J.F. (1996) <i>Genetics</i> , 143, 1069–1079
Recombinase	26	Chen,J., Evans,B.R., Yang,S., Araki,H., Oshima,Y. and Jayaram,M. (1992) <i>Mol. Cell. Biol.</i> , 12, 3757–3765.
Recombinase	27	Abremski,K.E. and Hoess,R.H. (1992) <i>Protein Engng</i> , 5, 87–91
Recombinase	28	Nunes-Düby SE, Kwon HJ, Tirumalai RS, Ellenberger T, Landy A. <i>Nucleic Acids Res.</i> 1998 Jan 15;26(2):391-406.
Recombinase	29	Cao Y, Hallet B, Sherratt DJ, Hayes F. <i>J Mol Biol.</i> 1997 Nov 21;274(1):39-53. PubMed PMID: 9398514.
Recombinase	30	Blakely G, et al. <i>Cell.</i> 1993 Oct 22;75(2):351-61. PubMed PMID: 8402918.
Recombinase	31	Lee L, Sadowski PD. <i>J Biol Chem.</i> 2003 Sep 19;278(38):36905-15. Epub 2003 Jul 8. PubMed PMID: 12851389.
Recombinase	32	Ghosh K, Lau CK, Gupta K, Van Duyne GD. <i>Nat Chem Biol.</i> 2005 Oct;1(5):275-82. Epub 2005 Sep 11.
Recombinase	33	Ghosh K, Guo F, Van Duyne GD. <i>J Biol Chem.</i> 2007 Aug 17;282(33):24004-16. Epub 2007 Jun 15. PubMed PMID: 17573343.
Recombinase	34	Guo F, Gopaul DN, Van Duyne GD. <i>Proc Natl Acad Sci U S A.</i> 1999 Jun 22;96(13):7143-8.
Recombinase	35	Cao Y, Hayes F. <i>J Mol Biol.</i> 1999 Jun 11;289(3):517-27. PubMed PMID: 10356326.
Recombinase	36	Parsons,R.L., Prasad,P.V., Harshey,R.M. and Jayaram,M. (1988) <i>Mol. Cell. Biol.</i> , 8, 3303–3310
Recombinase	37	Whiteson KL, Chen Y, Chopra N, Raymond AC, Rice PA. <i>Chem Biol.</i> 2007 Feb;14(2):121-9.
Recombinase	38	Martin SS, Pulido E, Chu VC, Lechner TS, Baldwin EP. <i>J Mol Biol.</i> 2002 May 24;319(1):107-27. PubMed PMID: 12051940.
Recombinase	39	Gelato KA, Martin SS, Baldwin EP. <i>J Mol Biol.</i> 2005 Nov 25;354(2):233-45. Epub 2005 Oct 5. PubMed PMID: 16242714.
Recombinase	40	Serre,M.C. and Jayaram,M. (1992) <i>J. Mol. Biol.</i> , 225, 643–649
Recombinase	41	Amin,A.A. and Sadowski,P.D. (1989) <i>Mol. Cell. Biol.</i> , 9, 1987–1995
Recombinase	42	Parsons,R.L., Evans,B.R., Zheng,L. and Jayaram,M. (1990) <i>J. Biol. Chem.</i> , 265, 4527–4533.
Recombinase	43	Chen Y, Rice PA. <i>J Biol Chem.</i> 2003 Jul 4;278(27):24800-7. Epub 2003 Apr 27. PubMed PMID: 12716882.
Recombinase	44	Jayaram,M., Crain,K.L., Parsons,R.L. and Harshey,R.M. (1988) <i>Proc. Natl. Acad. Sci. USA.</i> 85, 7902–7906
Recombinase	45	Nunes-Düby,S.E., Tirumalai,R.S., Dorgai,L., Yagil,R., Weisberg,R. and Landy,A. (1994) <i>EMBO J.</i> , 13, 4421–4430
Recombinase	46	Lebreton,B., Prasad,P.V., Jayaram,M. and Youderian,P. (1988) <i>Genetics</i> , 118, 393–400
Recombinase	47	Pargellis,C.A., Nunes-Düby,S.E., Moitoso de Vargas,L. and Landy,A.(1988) <i>J. Biol. Chem.</i> , 263, 7678–7685.
Recombinase	48	Wu,Z., Gumpport,R.I. and Gardner,J.F. (1997) <i>J. Bacteriol.</i> , 179, 4030–4038
Recombinase	49	Han,Y.W., Gumpport,R.I. and Gardner,J.F. (1993) <i>EMBO J.</i> , 12, 4577–4584
Recombinase	50	Pargellis,C.A., Nunes-Düby,S.E., Moitoso de Vargas,L. and Landy,A.(1988) <i>J. Biol. Chem.</i> , 263, 7678–7685.
Recombinase	51	Nunes-Düby,S.E., Tirumalai,R.S., Dorgai,L., Yagil,R., Weisberg,R. and Landy,A. (1994) <i>EMBO J.</i> , 13, 4421–4430
Recombinase	52	Zhu XD, Sadowski PD. <i>J Biol Chem.</i> 1995 Sep 29;270(39):23044-54. PubMed PMID: 7559444.
Recombinase	53	Cornet F, Hallet B, Sherratt DJ. <i>J Biol Chem.</i> 1997 Aug 29;272(35):21927-31. PubMed PMID: 9268326.
Recombinase	54	Ma CH, Kwiatek A, Bolusani S, Vozizyanov Y, Jayaram M. <i>J Mol Biol.</i> 2007 Apr 20;368(1):183-96. PubMed PMID: 17367810;
Recombinase	55	Arciszewska LK, Baker RA, Hallet B, Sherratt DJ. <i>J Mol Biol.</i> 2000 Jun 2;299(2):391-403. PubMed PMID: 10860747.
Recombinase	56	Cornet F, Hallet B, Sherratt DJ. <i>J Biol Chem.</i> 1997 Aug 29;272(35):21927-31. PubMed PMID: 9268326.
Recombinase	57	Vanhooff V, Normand C, Galloy C, Segall AM, Hallet B. <i>Nucleic Acids Res.</i> 2009 Dec 30. PMID: 20044348.
Recombinase	58	Lee J, Jayaram M. <i>J Biol Chem.</i> 1993 Aug 15;268(23):17564-70. PubMed PMID: 8349636.
Recombinase	59	Tribble G, Ahn YT, Lee J, Dandekar T, Jayaram M. <i>J Biol Chem.</i> 2000 Jul 21;275(29):22255-67. PubMed PMID: 10748094.
Recombinase	60	Lee J, Tonzuka T, Jayaram M. <i>Genes Dev.</i> 1997 Nov 15;11(22):3061-71. PubMed PMID:9367987

## Appendix B. Oligonucleotides and annealed substrates used in Cre assays

Table 6.1: Single stranded oligos that are annealed to form assay substrates

### Single stranded oligo substrates

Name	Length	Sequence (5'-3')	Description
loxSS1	27	TTCGAGTGGCCGATAACTTCGTATAAT	TS-nicked cleavage substrate
loxSS2	39	GTATGCTATACGAAGTTATGCAGATCTCAGCGGCGTGAC	TS-nicked cleavage substrate
loxSS3	64	GTCACGCCGCTGAGATCTGCATAACTTCGTATAGCATACATTATACGAAGTTATCGGCCAC'	intact BS
loxSS4	37	TTGTCACGCCGCTGAGATCTGCATAACTTCGTATAGC	BS-nicked cleavage substrate
loxSS5	29	ATACATTATACGAAGTTATCGGCCACTCG	BS-nicked cleavage substrate
loxSS6	64	CGAGTGGCCGATAACTTCGTATAATGTATGCTATACGAAGTTATGCAGATCTCAGCGGCGTGAC	intact TS
loxSS7-8	66	TTCGAGTGGCCGATAACTTCGTATAATGTATGCTATACGAAGTTATGCAGATCTCAGCGGCGTGAC	PS-TS cleavage substrate
loxSS9-10	66	TTGTCACGCCGCTGAGATCTGCATAACTTCGTATAGCATACATTATACGAAGTTATCGGCCACTCG	PS-BS cleavage substrate
loxSS1-pNP	26	TTCGAGTGGCCGATAACTTCGTATAA	TS-pNP modified ligation substrate
loxSS2c	40	TGTATGCTATACGAAGTTATGCAGATCTCAGCGGCGTGAC	TS-ligation substrate
loxSS4-pNP	36	TTGTCACGCCGCTGAGATCTGCATAACTTCGTATAG	BS-pNP modified ligation substrate
loxSS5c	30	CATACATTATACGAAGTTATCGGCCACTCG	BS-ligation substrate
loxS2AL	58	TTTTGTCACGCCGATAACTTCGTATAATGTACATTATACGAAGTTATCGGCCACTCG	Symmetric intact TS
loxS2BL	58	TTTTCGAGTGGCCGATAACTTCGTATAATGTACATTATACGAAGTTATGCGGCGTGAC	Symmetric intact TS

## Appendix B. continued.

Table 6.2: Annealed DNA substrates for Cre assays constructed from single stranded oligonucleotides

Annealed assay substrates		
Name	Purpose	Sequence
loxSS1/2/3	TS-nicked suicide cleavage	GTCACGCCGCTGAGATCTGCATAACTTCGTATAGC-ATAC-ATTATACGAAGTTATCGGCCACTCG CAGTGC GCGACTCTAGACGTATTGAAGCATATCG-TATG <b>TAATATGCTTCAATAGCCGGTGAGCTT</b>
loxSS4/5/6	BS-nicked suicide cleavage	TTGTCACGCCGCTGAGATCTGCATAACTTCGTATAGC ATAC-ATTATACGAAGTTATCGGCCACTCG CAGTGC GCGACTCTAGACGTATTGAAGCATATCG-TATG- <b>TAATATGCTTCAATAGCCGGTGAGC</b>
loxSS3/6	Eq-cleavage, binding and synapsis	GTCACGCCGCTGAGATCTGCATAACTTCGTATAGC-ATAC-ATTATACGAAGTTATCGGCCACTCG CAGTGC GCGACTCTAGACGTATTGAAGCATATCG-TATG- <b>TAATATGCTTCAATAGCCGGTGAGC</b>
loxS2	Eq-cleavage	TTTTGTCACGCCGCTGAGATCTGCATAACTTCGTATA <b>AT</b> GTACATTATACGAAGTTATCGGCCACTCG CAGTGC GCGCTATTGAAGCATATTACATG <b>TAATATGCTTCAATAGCCGGTGAGCTTTT</b>
loxSS7-8/9-10	TS/BS- Phosphorothioate cleavage	TTGTCACGCCGCTGAGATCTGCATAACTTCGTATAG <b>p</b> CATACAoTTATACGAAGTTATCGGCCACTCG CAGTGC GCGACTCTAGACGTATTGAAGCATATCoGTATGT <b>p</b> <b>TAATATGCTTCAATAGCCGGTGAGCTT</b>
loxSS7-8/3	TS – phosphorothioate cleavage	GTCACGCCGctGAGATCTGCATAACTTCGTATAGoCATACAoTTATACGAAGTTATCGGCCACTCG CAGTGC GCGACTCTAGACGTATTGAAGCATATCoGTATGT <b>p</b> <b>TAATATGCTTCAATAGCCGGTGAGCTT</b>
TS-pNP	TS-ligation	GTCACGCCGctGAGATCTGCATAACTTCGTATAGC <b>CATACA</b> -----TTATACGAAGTTATCGGCCACTCG CAGTGC GCGGaCTCTAGACGTATTGAAGCATATCGTATGT <b>Pnp-AATATGCTTCAATAGCCGGTGAGctt</b>
BS-pNP	BS-ligation	ttGTCACGCCGCTGAGATCTGCATAACTTCGTATAG- <b>pNP</b> CATACATTATACGAAGTTATCGGCCACTCG CAGTGC GCGACTCTAGACGTATTGAAGCATATC-----GTATGT <b>TAATATGCTTCAATAGCCGGTGAGC</b>

## References

- Davies, D. R., Mushtaq, A., Interthal, H., Champoux, J. J. & Hol, W.G.J. (2006). The structure of the transition state of the heterodimeric topoisomerase I of *Leishmania donovani* as a vanadate complex with nicked DNA. *Journal of Molecular Biology*, 357, pp. 1202-1210.
- Karimi-Busheri, F., Lee, J., Tomkinson, A. E. & Weinfeld, M. (1998). Repair of DNA strand gaps and nicks containing 3'-phosphate and 5'-hydroxyl termini by purified mammalian enzymes. *Nucleic Acids Research*, 26, pp. 4395-4400.
- Perry, K., Hwang, Y., Bushman, F. D. & Van Duyne, G.D. (2010). Insights from the Structure of a Smallpox Virus Topoisomerase-DNA Transition State Mimic. *Structure (London, England : 1993)*, 18, pp. 127-137.
- Abremski K, Wierzbicki A, Frommer B & Hoess RH (1986). Bacteriophage p1 cre-loxp site-specific recombination. site-specific dna topoisomerase activity of the cre recombination protein. *J. Biol. Chem.*, 261, pp. 391-396.
- Abremski KE & Hoess RH (1992). Evidence for a second conserved arginine residue in the integrase family of recombination proteins. *Protein Eng.*, 5, pp. 87-91.
- Adams DE, Bliska JB & Cozzarelli NR (1992). Cre-lox recombination in escherichia coli cells. mechanistic differences from the in vitro reaction. *J. Mol. Biol.*, 226, pp. 661-673.
- Aihara H, Kwon HJ, Nunes-Duby SE, Landy A & Ellenberger T (2003a). A conformational switch controls the dna cleavage activity of lambda integrase. *Mol. Cell*, 12, pp. 187-198.
- Aihara H, Kwon HJ, Nunes-Düby SE, Landy A & Ellenberger T (2003b). A conformational switch controls the dna cleavage activity of lambda integrase. *Mol. Cell*, 12, pp. 187-198.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997). Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic Acids Res.*, 25, pp. 3389-3402.
- Arciszewska LK, Baker RA, Hallet B & Sherratt DJ (2000). Coordinated control of xerc and xerd catalytic activities during holliday junction resolution. *J. Mol. Biol.*, 299, pp. 391-403.
- Azaro M & Landy A (2002).  $\Lambda$  integrase and the  $\lambda$  int family. In Craig N (Ed.), *Mobile dna ii*. : ASM Press, Washington, D.C. pp. 118-148.

Azaro, M A, Landy, A (2002). L integrase and the Int family. In Craig N (Ed.), *Mobile dna ii.* : ASM Press, Washington D. C. pp. 118-148.

Barre FX & Sherratt DJ (2002). Xer site-specific recombination: promoting chromosome segregation. In Craig N (Ed.), *Mobile dna ii.* : ASM Press, Washington, D.C. pp. 149-161.

Birling M, Gofflot F & Warot X (2009). Site-specific recombinases for manipulation of the mouse genome. *Methods Mol. Biol.*, 561, pp. 245-263.

Biswas T, Aihara H, Radman-Livaja M, Filman D, Landy A & Ellenberger T (2005). A structural basis for allosteric control of dna recombination by lambda integrase. *Nature*, 435, pp. 1059-1066.

Brosius J (1984). Toxicity of an overproduced foreign gene product in escherichia coli and its use in plasmid vectors for the selection of transcription terminators. *Gene*, 27, pp. 161-172.

Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T & Warren GL (1998). Crystallography & nmr system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, 54, pp. 905-921.

Buchholz F & Stewart AF (2001). Alteration of cre recombinase site specificity by substrate-linked protein evolution. *Nat. Biotechnol.*, 19, pp. 1047-1052.

Cao Y & Hayes F (1999). A newly identified, essential catalytic residue in a critical secondary structure element in the integrase family of site-specific recombinases is conserved in a similar element in eucaryotic type I topoisomerases. *J. Mol. Biol.*, 289, pp. 517-527.

Cassano AG, Anderson VE & Harris ME (2004). Understanding the transition states of phosphodiester bond cleavage: insights from heavy atom isotope effects. *Biopolymers*, 73, pp. 110-129.

Champoux JJ (2001). Dna topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem.*, 70, pp. 369-413.

Chen JW, Evans BR, Yang SH, Araki H, Oshima Y & Jayaram M (1992a). Functional analysis of box I mutations in yeast site-specific recombinases flp and r: pairwise complementation with recombinase variants lacking the active-site tyrosine. *Mol. Cell. Biol.*, 12, pp. 3757-3765.

Chen JW, Lee J & Jayaram M (1992b). Dna cleavage in trans by the active site tyrosine during flp recombination: switching protein partners before exchanging strands. *Cell*, 69, pp. 647-658.

Chen Y & Rice PA (2003a). The role of the conserved Trp330 in flp-mediated recombination. functional and structural analysis. *J. Biol. Chem.*, 278, pp. 24800-24807.

- Chen Y & Rice PA (2003b). New insight into site-specific recombination from flp recombinase-dna structures. *Annu Rev Biophys Biomol Struct*, 32, pp. 135-159.
- Chen Y & Rice PA (2003c). New insight into site-specific recombination from flp recombinase-dna structures. *Annu Rev Biophys Biomol Struct*, 32, pp. 135-159.
- Chen Y, Narendra U, Iype LE, Cox MM & Rice PA (2000). Crystal structure of a flp recombinase-holliday junction complex: assembly of an active oligomer by helix swapping. *Mol. Cell*, 6, pp. 885-897.
- Cheng C, Kussie P, Pavletich N & Shuman S (1998). Conservation of structure and mechanism between eukaryotic topoisomerase i and site-specific recombinases. *Cell*, 92, pp. 841-850.
- Cheng C, Wang LK, Sekiguchi J & Shuman S (1997). Mutational analysis of 39 residues of vaccinia dna topoisomerase identifies lys-220, arg-223, and asn-228 as important for covalent catalysis. *J. Biol. Chem.*, 272, pp. 8263-8269.
- Christiansen K, Knudsen BR & Westergaard O (1994). The covalent eukaryotic topoisomerase i-dna intermediate catalyzes ph-dependent hydrolysis and alcoholysis. *J. Biol. Chem.*, 269, pp. 11367-11373.
- Conway AB, Chen Y & Rice PA (2003). Structural plasticity of the flp-holliday junction complex. *J. Mol. Biol.*, 326, pp. 425-434.
- Cornet F, Hallet B & Sherratt DJ (1997). Xer recombination in escherichia coli. site-specific dna topoisomerase activity of the xerc and xerd recombinases. *J. Biol. Chem.*, 272, pp. 21927-21931.
- Craig NL & Nash HA (1983). The mechanism of phage lambda site-specific recombination: site-specific breakage of dna by int topoisomerase. *Cell*, 35, pp. 795-803.
- Craig NL, Gragie R, Gellert M & Lambowitz AM (2002). *Mobile dna ii.* : ASM Press, Washington D.C..
- Dale EC & Ow DW (1990). Intra- and intermolecular site-specific recombination in plant cells mediated by bacteriophage p1 recombinase. *Gene*, 91, pp. 79-85.
- Davies DR, Mushtaq A, Interthal H, Champoux JJ & Hol WGJ (2006). The structure of the transition state of the heterodimeric topoisomerase i of leishmania donovani as a vanadate complex with nicked dna. *J. Mol. Biol.*, 357, pp. 1202-1210.
- Dmitrova M, Younès-Cauet G, Oertel-Buchheit P, Porte D, Schnarr M & Granger-Schnarr M (1998). A new lexa-based genetic system for monitoring and analyzing protein heterodimerization in escherichia coli. *Mol. Gen. Genet.*, 257, pp. 205-212.

- Dong KC & Berger JM (2008). Structure and function of dna topoisomerases. In Rice P & Correll C (Eds.), *Protein-nucleic acid interactions*. : Royal Society of Chemistry. pp. 234-269.
- Emsley P & Cowtan K (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, 60, pp. 2126-2132.
- Ferreira H, Butler-Cole B, Burgin A, Baker R, Sherratt DJ & Arciszewska LK (2003). Functional analysis of the c-terminal domains of the site-specific recombinases xerc and xerd. *J. Mol. Biol.*, 330, pp. 15-27.
- Friesen H & Sadowski PD (1992). Mutagenesis of a conserved region of the gene encoding the flp recombinase of *saccharomyces cerevisiae*. a role for arginine 191 in binding and ligation. *J. Mol. Biol.*, 225, pp. 313-326.
- Gage PJ, Sauer B, Levine M & Glorioso JC (1992). A cell-free recombination system for site-specific integration of multigenic shuttle plasmids into the herpes simplex virus type 1 genome. *J. Virol.*, 66, pp. 5509-5515.
- Gao R, Zhang Y, Choudhury AK, Dedkova LM & Hecht SM (2005). Analogues of vaccinia virus dna topoisomerase i modified at the active site tyrosine. *J. Am. Chem. Soc.*, 127, pp. 3321-3331.
- Gelato KA, Martin SS & Baldwin EP (2005). Reversed dna strand cleavage specificity in initiation of cre-loxp recombination induced by the his289ala active-site substitution. *J. Mol. Biol.*, 354, pp. 233-245.
- Ghosh K & Van Duyne GD (2002). Cre-loxp biochemistry. *Methods*, 28, pp. 374-383.
- Ghosh K, Guo F & Van Duyne GD (2007). Synapsis of loxp sites by cre recombinase. *J. Biol. Chem.*, 282, pp. 24004-24016.
- Ghosh K, Lau C, Gupta K & Van Duyne GD (2005a). Preferential synapsis of loxp sites drives ordered strand exchange in cre-loxp site-specific recombination. *Nat. Chem. Biol.*, 1, pp. 275-282.
- Ghosh K, Lau CK, Guo F, Segall AM & Van Duyne GD (2005b). Peptide trapping of the holliday junction intermediate in cre-loxp site-specific recombination. *J. Biol. Chem.*, 280, pp. 8290-8299.
- Gopaul DN, Guo F & Van Duyne GD (1998). Structure of the holliday junction intermediate in cre-loxp site-specific recombination. *EMBO J.*, 17, pp. 4175-4187.
- Gordon JA (1991). Use of vanadate as protein-phosphotyrosine phosphatase inhibitor. *Meth. Enzymol.*, 201, pp. 477-482.
- Grindley ND, Whiteson KL & Rice PA (2006). Mechanisms of site-specific recombination. *Annu. Rev. Biochem.*, , .

Guo F, Gopaul DN & van Duyne GD (1997). Structure of cre recombinase complexed with dna in a site-specific recombination synapse. *Nature*, 389, pp. 40-46.

Guo F, Gopaul DN & Van Duyne GD (1999). Asymmetric dna bending in the cre-loxp site-specific recombination synapse. *Proc. Natl. Acad. Sci. U.S.A.*, 96, pp. 7143-7148.

Hallet B, Arciszewska LK & Sherratt DJ (1999). Reciprocal control of catalysis by the tyrosine recombinases xerc and xerd: an enzymatic switch in site-specific recombination. *Mol. Cell*, 4, pp. 949-959.

Hoess R, Wierzbicki A & Abremski K (1987). Isolation and characterization of intermediates in site-specific recombination. *Proc. Natl. Acad. Sci. U.S.A.*, 84, pp. 6840-6844.

Hoess RH & Abremski K (1984). Interaction of the bacteriophage p1 recombinase cre with the recombining site loxp. *Proc. Natl. Acad. Sci. U.S.A.*, 81, pp. 1026-1029.

Hoess RH, Ziese M & Sternberg N (1982). P1 site-specific recombination: nucleotide sequence of the recombining sites. *Proc. Natl. Acad. Sci. U.S.A.*, 79, pp. 3398-3402.

Hoess, R H, Wierzbicki A, Abremski KE (1990). Synapsis in the cre-lox site-specific recombination system. , , pp. 203-213.

Jayaram M, Grainge I & Tribble G (2002). Site-specific recombination by the flp protein of *saccharomyces cerevisiae*. In Craig N (Ed.), *Mobile dna ii*. : ASM Press, Washington, D.C. pp. 192-218.

Kazmierczak RA, Swalla BM, Burgin AB, Gumport RI & Gardner JF (2002). Regulation of site-specific recombination by the c-terminus of lambda integrase. *Nucleic Acids Res.*, 30, pp. 5193-5204.

Kikuchi Y & Nash HA (1979). Nicking-closing activity associated with bacteriophage lambda int gene product. *Proc. Natl. Acad. Sci. U.S.A.*, 76, pp. 3760-3764.

Kilby NJ, Snaith MR & Murray JA (1993). Site-specific recombinases: tools for genome engineering. *Trends Genet.*, 9, pp. 413-421.

Kirby AJ & Younas M (1970a). Reactivity of phosphate esters. diester hydrolysis. *J. Chem. Soc. B.*, , p. 510.

Kirby AJ & Younas M (1970b). The reactivity of phosphate esters. reactions of diesters with nucleophiles. *J. Chem. Soc. B*, , p. 1165.

- Klemperer N & Traktman P (1993). Biochemical analysis of mutant alleles of the vaccinia virus topoisomerase I carrying targeted substitutions in a highly conserved domain. *J. Biol. Chem.*, 268, pp. 15887-15899.
- Krogh BO & Shuman S (2000). Catalytic mechanism of DNA topoisomerase I. *Mol. Cell*, 5, pp. 1035-1041.
- Krogh BO & Shuman S (2001). Vaccinia topoisomerase mutants illuminate conformational changes during closure of the protein clamp and assembly of a functional active site. *J. Biol. Chem.*, 276, pp. 36091-36099.
- Krogh BO & Shuman S (2002). Proton relay mechanism of general acid catalysis by DNA topoisomerase I. *J. Biol. Chem.*, 277, pp. 5711-5714.
- Kuipers BJH & Gruppen H (2007). Prediction of molar extinction coefficients of proteins and peptides using UV absorption of the constituent amino acids at 214 nm to enable quantitative reverse phase high-performance liquid chromatography-mass spectrometry analysis. *J. Agric. Food Chem.*, 55, pp. 5445-5451.
- Kwon HJ, Tirumalai R, Landy A & Ellenberger T (1997). Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science*, 276, pp. 126-131.
- Lee J & Jayaram M (1993). Mechanism of site-specific recombination. Logic of assembling recombinase catalytic site from fractional active sites. *J. Biol. Chem.*, 268, pp. 17564-17570.
- Lee J, Whang I, Lee J & Jayaram M (1994). Directed protein replacement in recombination full sites reveals trans-horizontal DNA cleavage by flp recombinase. *EMBO J.*, 13, pp. 5346-5354.
- Lee L & Sadowski PD (2003). Identification of Cre residues involved in synapsis, isomerization, and catalysis. *J. Biol. Chem.*, 278, pp. 36905-36915.
- Lee L, Chu LC & Sadowski PD (2003). Cre induces an asymmetric DNA bend in its target loxP site. *J. Biol. Chem.*, 278, pp. 23118-23129.
- Lee SY, Aihara H, Ellenberger T & Landy A (2004). Two structural features of lambda integrase that are critical for DNA cleavage by multimers but not by monomers. *Proc. Natl. Acad. Sci. U.S.A.*, 101, pp. 2770-2775.
- Liu Q, Li MZ, Leibham D, Cortez D & Elledge SJ (1998). The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes. *Curr. Biol.*, 8, pp. 1300-1309.
- Liu Q, Li MZ, Liu D & Elledge SJ (2000). Rapid construction of recombinant DNA by the univector plasmid-fusion system. *Meth. Enzymol.*, 328, pp. 530-549.

- Livet J, Weissman TA, Kang H, Draft RW, Lu J, Bennis RA, Sanes JR & Lichtman JW (2007). Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature*, 450, pp. 56-62.
- Ma C, Kachroo AH, Macieszak A, Chen T, Guga P & Jayaram M (2009a). Reactions of cre with methylphosphonate dna: similarities and contrasts with flp and vaccinia topoisomerase. *PLoS ONE*, 4, p. e7248.
- Ma C, Kwiatek A, Bolusani S, Voziyanov Y & Jayaram M (2007). Unveiling hidden catalytic contributions of the conserved his/trp-iii in tyrosine recombinases: assembly of a novel active site in flp recombinase harboring alanine at this position. *J. Mol. Biol.*, 368, pp. 183-196.
- Ma C, Rowley PA, Macieszak A, Guga P & Jayaram M (2009b). Active site electrostatics protect genome integrity by blocking abortive hydrolysis during dna recombination. *EMBO J.*, 28, pp. 1745-1756.
- Maegley KA, Admiraal SJ & Herschlag D (1996). Ras-catalyzed hydrolysis of gtp: a new perspective from model studies. *Proc. Natl. Acad. Sci. U.S.A.*, 93, pp. 8160-8166.
- Martin SS, Pulido E, Chu VC, Lechner TS & Baldwin EP (2002). The order of strand exchanges in cre-loxp recombination and its basis suggested by the crystal structure of a cre-loxp holliday junction complex. *J. Mol. Biol.*, 319, pp. 107-127.
- Murshudov GN, Vagin AA & Dodson EJ (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, 53, pp. 240-255.
- Nagarajan R, Kwon K, Nawrot B, Stec WJ & Stivers JT (2005). Catalytic phosphoryl interactions of topoisomerase *ib*. *Biochemistry*, 44, pp. 11476-11485.
- Nash HA (1996). Site-specific recombination: integration, excision, resolution, and inversion of defined dna segments. In Neidhardt FC (Ed.), *Escherichia coli and salmonella: cellular and molecular biology*. : ASM Press, Washington, D.C.. pp. 2363-2376.
- Nunes-Duby SE, Kwon HJ, Tirumalai RS, Ellenberger T & Landy A (1998). Similarities and differences among 105 members of the int family of site-specific recombinases. *Nucleic Acids Res.*, 26, pp. 391-406.
- Nunes-Duby SE, Kwon HJ, Tirumalai RS, Ellenberger T & Landy A (1998). Similarities and differences among 105 members of the int family of site-specific recombinases. *Nucleic Acids Res.*, 26, pp. 391-406.
- Otwinowski Z & Minor W (1997). Processing of x-ray diffraction data collected in oscillation mode. *Meth. Enzymol.*, 276, pp. 307-326.

- Pan G, Luetke K & Sadowski PD (1993a). Mechanism of cleavage and ligation by flp recombinase: classification of mutations in flp protein by in vitro complementation analysis. *Mol. Cell. Biol.*, 13, pp. 3167-3175.
- Pan G, Luetke K, Juby CD, Brousseau R & Sadowski P (1993b). Ligation of synthetic activated dna substrates by site-specific recombinases and topoisomerase i. *J. Biol. Chem.*, 268, pp. 3683-3689.
- Parsons RL, Prasad PV, Harshey RM & Jayaram M (1988). Step-arrest mutants of flp recombinase: implications for the catalytic mechanism of dna recombination. *Mol. Cell. Biol.*, 8, pp. 3303-3310.
- Patel A, Shuman S & Mondragón A (2006). Crystal structure of a bacterial type ii dna topoisomerase reveals a preassembled active site in the absence of dna. *J. Biol. Chem.*, 281, pp. 6030-6037.
- Peakman TC, Harris RA & Gewert DR (1992). Highly efficient generation of recombinant baculoviruses by enzymatically mediated site-specific in vitro recombination. *Nucleic Acids Res.*, 20, pp. 495-500.
- Perry K, Hwang Y, Bushman FD & Van Duyne GD (2006). Structural basis for specificity in the poxvirus topoisomerase. *Mol. Cell*, 23, pp. 343-354.
- Perry K, Hwang Y, Bushman FD & Van Duyne GD (2010). Insights from the structure of a smallpox virus topoisomerase-dna transition state mimic. *Structure*, 18, pp. 127-137.
- Petersen BO & Shuman S (1997a). Dna strand transfer reactions catalyzed by vaccinia topoisomerase: hydrolysis and glycerololysis of the covalent protein-dna intermediate. *Nucleic Acids Res.*, 25, pp. 2091-2097.
- Petersen BO & Shuman S (1997b). Histidine 265 is important for covalent catalysis by vaccinia topoisomerase and is conserved in all eukaryotic type i enzymes. *J. Biol. Chem.*, 272, pp. 3891-3896.
- Phillips GJ (1999). New cloning vectors with temperature-sensitive replication. *Plasmid*, 41, pp. 78-81.
- Pruitt KD, Tatusova T & Maglott DR (2007). Ncbi reference sequences (refseq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, 35, p. D61-5.
- Redinbo MR, Stewart L, Kuhn P, Champoux JJ & Hol WG (1998). Crystal structures of human topoisomerase i in covalent and noncovalent complexes with dna. *Science*, 279, pp. 1504-1513.

- Ringrose L, Lounnas V, Ehrlich L, Buchholz F, Wade R & Stewart AF (1998). Comparative kinetic analysis of flp and cre recombinases: mathematical models for dna binding and recombination. *J. Mol. Biol.*, 284, pp. 363-384.
- Sadowski PD (1995). The flp recombinase of the 2-microns plasmid of *saccharomyces cerevisiae*. *Prog. Nucleic Acid Res. Mol. Biol.*, 51, pp. 53-91.
- Sarkar I, Hauber I, Hauber J & Buchholz F (2007). Hiv-1 proviral dna excision using an evolved recombinase. *Science*, 316, pp. 1912-1915.
- Sauer B (1987). Functional expression of the cre-lox site-specific recombination system in the yeast *saccharomyces cerevisiae*. *Mol. Cell. Biol.*, 7, pp. 2087-2096.
- Sauer B (2002). Chromosome manipulation by cre-lox recombination. In Craig N (Ed.), *Mobile dna ii*. : ASM Press, Washington, D.C. pp. 38-58.
- Serre MC & Jayaram M (1992). Half-site strand transfer by step-arrest mutants of yeast site-specific recombinase flp. *J. Mol. Biol.*, 225, pp. 643-649.
- Sherratt DJ & Wigley DB (1998). Conserved themes but novel activities in recombinases and topoisomerases. *Cell*, 93, pp. 149-152.
- Shuman S (1998). Vaccinia virus dna topoisomerase: a model eukaryotic type Ib enzyme. *Biochim. Biophys. Acta*, 1400, pp. 321-337.
- Spiers AJ & Sherratt DJ (1997). Relating primary structure to function in the *escherichia coli* xerD site-specific recombinase. *Mol. Microbiol.*, 24, pp. 1071-1082.
- Spiers AJ & Sherratt DJ (1999). C-terminal interactions between the xerc and xerD site-specific recombinases. *Mol. Microbiol.*, 32, pp. 1031-1042.
- Sternberg N, Hamilton D, Austin S, Yarmolinsky M & Hoess R (1981). Site-specific recombination and its role in the life cycle of bacteriophage p1. *Cold Spring Harb. Symp. Quant. Biol.*, 45 Pt 1, pp. 297-309.
- Sternberg N, Sauer B, Hoess R & Abremski K (1986). Bacteriophage p1 cre gene and its regulatory region. evidence for multiple promoters and for regulation by dna methylation. *J. Mol. Biol.*, 187, pp. 197-212.
- Stewart L, Redinbo MR, Qiu X, Hol WG & Champoux JJ (1998). A model for the mechanism of human topoisomerase I. *Science*, 279, pp. 1534-1541.

Stivers JT, Jagadeesh GJ, Nawrot B, Stec WJ & Shuman S (2000). Stereochemical outcome and kinetic effects of rp- and sp-phosphorothioate substitutions at the cleavage site of vaccinia type I DNA topoisomerase. *Biochemistry*, 39, pp. 5561-5572.

Stivers JT, Shuman S & Mildvan AS (1994). Vaccinia DNA topoisomerase I: kinetic evidence for general acid-base catalysis and a conformational step. *Biochemistry*, 33, pp. 15449-15458.

Subramanya HS, Arciszewska LK, Baker RA, Bird LE, Sherratt DJ & Wigley DB (1997). Crystal structure of the site-specific recombinase, XerD. *EMBO J.*, 16, pp. 5178-5187.

Tekle M, Warren DJ, Biswas T, Ellenberger T, Landy A & Nunes-Düby SE (2002). Attenuating functions of the C terminus of lambda integrase. *J. Mol. Biol.*, 324, pp. 649-665.

Tian L, Claeboe CD, Hecht SM & Shuman S (2003). Guarding the genome: electrostatic repulsion of water by DNA suppresses a potent nuclease activity of topoisomerase I $\beta$ . *Mol. Cell*, 12, pp. 199-208.

Van Duyne G (2008). Site-specific recombinases. In Rice P & Correll C (Eds.), *Protein-nucleic acid interactions*. : Royal Society of Chemistry. pp. 303-332.

Van Duyne GD (2001). A structural view of Cre-loxP site-specific recombination. *Annu Rev Biophys Biomol Struct*, 30, pp. 87-104.

Van Duyne GD (2002). A structural view of tyrosine recombinase site-specific recombination. In Craig N (Ed.), *Mobile DNA II*. : ASM Press, Washington, D.C. pp. 93-117.

Vanhooff V, Normand C, Galloy C, Segall AM & Hallet B (2009). Control of directionality in the DNA strand-exchange reaction catalysed by the tyrosine recombinase TnpI. *Nucleic Acids Res.*, , .

Warren D, Laxmikanthan G & Landy A (2008). A chimeric Cre recombinase with regulated directionality. *Proc. Natl. Acad. Sci. U.S.A.*, 105, pp. 18278-18283.

Whipple FW (1998). Genetic analysis of prokaryotic and eukaryotic DNA-binding proteins in *Escherichia coli*. *Nucleic Acids Res.*, 26, pp. 3700-3706.

Whiteson KL, Chen Y, Chopra N, Raymond AC & Rice PA (2007). Identification of a potential general acid/base in the reversible phosphoryl transfer reactions catalyzed by tyrosine recombinases: Flp H305. *Chem. Biol.*, 14, pp. 121-129.

Wierzbicki A, Kendall M, Abremski K & Hoess R (1987). A mutational analysis of the bacteriophage P1 recombinase Cre. *J. Mol. Biol.*, 195, pp. 785-794.

Wittschieben J & Shuman S (1997). Mechanism of dna transesterification by vaccinia topoisomerase: catalytic contributions of essential residues arg-130, gly-132, tyr-136 and lys-167. *Nucleic Acids Res.*, 25, pp. 3001-3008.

Woodfield G, Cheng C, Shuman S & Burgin AB (2000). Vaccinia topoisomerase and cre recombinase catalyze direct ligation of activated dna substrates containing a 3'-para-nitrophenyl phosphate ester. *Nucleic Acids Res.*, 28, pp. 3323-3331.

Xu CJ, Grainge I, Lee J, Harshey RM & Jayaram M (1998). Unveiling two distinct ribonuclease activities and a topoisomerase activity in a site-specific dna recombinase. *Mol. Cell*, 1, pp. 729-739.

Yakovleva L, Chen S, Hecht SM & Shuman S (2008). Chemical and traditional mutagenesis of vaccinia dna topoisomerase provides insights to cleavage site recognition and transesterification chemistry. *J. Biol. Chem.*, 283, pp. 16093-16103.

Yates J, Zhekov I, Baker R, Eklund B, Sherratt DJ & Arciszewska LK (2006). Dissection of a functional interaction between the dna translocase, ftsk, and the xerD recombinase. *Mol. Microbiol.*, 59, pp. 1754-1766.

Yuan P (2008). . Unpublished doctoral dissertation, ,