This doctoral thesis has been examined by a committee of the Department of Chemistry as follows:

Gerald N. Wogan, Professor of Chemistry and Biological Engineering, Chairman

John M. Essigmann
William and Betsy Leitch Professor of Chemistry and Toxicology, Thesis Supervisor

Barbara Imperiali, Class of 1922 Professor of Chemistry and Professor of Biology
Biochemical Characterization of the *E. coli* Very Short Patch Repair Pathway and its Coordination with Methyltransferase Repair of O\(^6\)-Methylguanine

by

Peter Thomas Rye

Submitted to the Department of Chemistry on April 28, 2006 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biological Chemistry

The *E. coli* Very Short Patch Repair (VSPR) system corrects T:G mismatches that arise through Dcm-mediated methylation and subsequent deamination of the underlined cytosine residue in the palindromic sequence 5'-CC\_WGG-3' (W is an adenine or thymine). Vsr initiates VSPR by producing a single stranded nick on the 5' side of the mismatched T. The MutS and MutL mismatch recognition proteins stimulate this activity, as cells lacking either of these proteins display diminished VSPR. Genetic studies also indicate that Pol I is responsible for removing and replacing a short tract of nucleotides downstream of the incision site and that DNA Ligase seals the nick to complete the repair event. However, until now, biochemical investigation of the repair steps downstream of Vsr incision have been lacking.

Herein, we describe two novel in vitro assays used to probe the biochemical events of VSPR. The first was used to verify the reconstitution of VSPR using purified *E. coli* Vsr, Pol I, and DNA Ligase enzymes, while the second was used to measure the distribution of VSPR patch sizes in whole cell extracts. By monitoring the loss of radiosignal from a series of substrates that contained the label at prescribed distances downstream of the T:G mismatch, we were able to determine that VSPR patches are distributed around 2 to 4 deoxynucleotides in length. Interestingly, under certain reaction conditions, the addition of DNA Ligase improved the efficiency of repair initiation by Vsr, suggesting that VSPR may be optimal in the context of a multi-protein complex.

Lastly, we investigated the effect of VSPR proteins on methyltransferase (MTase) repair of O\(^6\)-methylguanine (O\(^6\)mG). MTase repair of O\(^6\)mG opposite T results in a G:T mismatch that must be further processed to yield the native G:C base pairing. The G:T mismatch is therefore an intersection of the two pathways and led us to hypothesize that MTase and VSPR proteins might interact. Indeed, cells lacking the functions of MutS, MutL, or Vsr proteins displayed decreased MTase repair in vivo, revealing a previously unknown interaction. The cooperation between proteins of these two repair systems may shed light on the biological significance of the VSPR system.

Thesis Supervisor: John M. Essigmann
Title: William and Betsy Leitch Professor of Chemistry and Toxicology
ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. John Essigmann, for his guidance and support throughout the years. John was consistently receptive to my ideas and very helpful in thinking them through. His energy in his work provided a great model to follow and inspired me to achieve the standards of excellence he set for himself. He is a brilliant scientist and has an amazing talent to communicate ideas. I thank him for his time and persistence in teaching me those skills, and providing, all the while, an enjoyable and productive environment in which to work. John is an extremely kind and selfless individual, and he has taught me more than he will ever be know.

I would also like to thank my thesis committee members, Drs. Gerald Wogan and Barbara Imperiali, for taking the time and energy to assess my work and provide valuable feedback. I am especially indebted to Dr. Wogan for continually inspiring me to seek a greater understanding of my work, the implications it may have, and how to take it to the next level.

Everyone in the Essigmann Laboratory has positively impacted me in some way. I need to thank Sarah, Kyle, Jim, Lauren, and Sreeja for being such great friends and reading my many thesis drafts and suggesting how to make them better. Your many comments were appreciated and helped enormously in defining the quality of my work. Many thanks go to Kim for her attention and kindness in dealing with administrative matters. I am convinced no other could take her place. I would also like to thank many individuals (past and present) for their friendship and scientific discussions. Aida, Alphio, Annie, Bob, Bogdan, Charles, Denise, Eunsuk, Francis, Jeannette, Jen, Jim, John, Kaushik, Maryann, Michelle, Neena, Nicole, Paul, Pei-sze, Shawn, Uday, Will, Yuri, and Zoran were all excellent people with whom to work. I hope I stay in touch with you all and I look forward to seeing where your careers lead you. I consider all of you my friends, and I hope that I had impacted your life in a positive way like you have impacted mine.

Special thanks go to the numerous undergraduates I have had the privilege to mentor. Lucien, Meghan, Ai-ris, Jenny, Dana, and Jelly were all amazing students to teach and also to learn from. The saying that “there’s no greater way to learn than by teaching” holds very true. You all kept me on my toes by demanding an unparalleled understanding of our work. I owe a great amount of what I understand to our discussions. I also owe a great amount of what we accomplished in the lab to your efforts. I truly appreciate the devotion you all displayed in your work and the passion you had towards accumulating meaningful data. I hope that you feel great value, as I do, in the UROP experience we shared.

Many thanks go to fellow chemistry graduate students. I remember the great friendships formed in my first semester at MIT. I thank Aaron, Abi, Fran, Jen, Lyn, Mary, Matt, Shawn, Woody, and Zarixia for helping to make MIT the right choice. Our first year was one of the best. I look forward to seeing where you all end up. To Hector,
thanks for everything. You are an amazing person and friend. I hope we never lose touch.

I would like to thank the “Merrimack Crew” for keeping me grounded, cheering me on, and always being around. I will forever be grateful for the friendship of Andy, Cory, Derek, Erin, Greg, Holly, Jamie, Jen, Jennie, Jesse, Jessica, Mandy, Mark, Matt, Meredith, Mike, Mike, Salstrom, Sarah, and Steve. I feel blessed every day that I have such wonderful people in my life who I can count on for anything. Thanks to you, and your significant others, for everything in the past, present, and (most exciting) the future. One word, LIFE. I love you all.

I am also thankful for Syracuse folks including Aleta, Ben, Bry, Cindy, Edi, Green, Hosmer, Jill, Jordan, Kyoko, Larissa, Maria, Meth, Mooney, Morgan, Moran, Rachael, Rob, Steve, and Zander. You all had a positive impact on my life and helped me to remember the things that are truly important. May we continue to keep the good times rolling.

I would also like to acknowledge friends that I have made along the way. I feel fortunate to have had the support from people such as Cara, Carolyn, Kristen, Molly, Paul, Sylvia, and TK. You all have made the last years of my life more enjoyable. Special thanks goes to Sarah for her irreplaceable friendship. She has always helped me to enjoy life and appreciate its many graces. I will forever be touched by her unyielding optimism and the friendship we share.

Lastly, I would like to extend the largest thank you possible to my wonderful family. I have received so much love and support from my two sisters, Elise and Heide, their husbands, Paul and Randy, and their families. I am deeply touched by how much they selflessly helped me through the tough times and celebrated with me through the good ones. It was also amazing to have received copious amounts of letters, drawings, pictures, phone calls, and hugs from my wonderful nieces and nephews Ben, Billy, Claire, Hope, Jane, Katie, Kay, Morgan, and Sara. You all are amazing people and I thank you for the love and support that you gave me.

To my parents, the two most generous, kind, wise, fun, and loving people in my life, thank you for everything. You guys are always such an amazing source of direction, strength, and inspiration. I cannot thank you enough for the endless support you give me. Your attitude and advice keep me going through the rocky times and your excitement during the highs is incomparable. I thank God that I have such wonderful parents to push, pull, and stand by when needed. I also thank you for your talents in knowing just how to help me see things through. I know, from experience, that I can always look to your teachings for clarity and your example for guidance. By far and away, you contribute to my life’s work the most. Therefore, I dedicate this thesis to you. Remember, this is only the beginning.

PETER T. RYE
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Committee Page</td>
<td>2</td>
</tr>
<tr>
<td>Abstract</td>
<td>3</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>4</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>6</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>9</td>
</tr>
<tr>
<td>List of Figures</td>
<td>12</td>
</tr>
<tr>
<td>List of Tables</td>
<td>15</td>
</tr>
<tr>
<td>List of Schemes</td>
<td>16</td>
</tr>
<tr>
<td>Chapter 1. Mismatch Repair in <em>Escherichia coli</em>.</td>
<td>17</td>
</tr>
<tr>
<td>1.1. Mutation Avoidance</td>
<td>18</td>
</tr>
<tr>
<td>1.2. The Very Short Patch Repair System</td>
<td>19</td>
</tr>
<tr>
<td>1.2.1. Dcm – <em>in Vivo</em> and <em>in Vitro</em> Properties.</td>
<td>22</td>
</tr>
<tr>
<td>1.2.2. Vsr – <em>in Vivo</em> and <em>in Vitro</em> Properties</td>
<td>25</td>
</tr>
<tr>
<td>1.2.3. Structural Data for Vsr</td>
<td>28</td>
</tr>
<tr>
<td>1.2.4. Catalytic Mechanism of Vsr</td>
<td>34</td>
</tr>
<tr>
<td>1.3. Methyl-Directed Mismatch Repair – Lessons From <em>Streptococcus Pneumoniae</em></td>
<td>36</td>
</tr>
<tr>
<td>1.4. Methyl-Directed Mismatch Repair in <em>E. coli</em>.</td>
<td>37</td>
</tr>
<tr>
<td>1.4.1. Dam – <em>in Vivo</em> and <em>in Vitro</em> Properties</td>
<td>38</td>
</tr>
<tr>
<td>1.4.2. MutS – <em>in Vivo</em> and <em>in Vitro</em> Properties</td>
<td>39</td>
</tr>
<tr>
<td>1.4.3. MutL – <em>in Vivo</em> and <em>in Vitro</em> Properties</td>
<td>41</td>
</tr>
<tr>
<td>1.4.4. MutH – <em>in Vivo</em> and <em>in Vitro</em> Properties</td>
<td>42</td>
</tr>
<tr>
<td>1.4.5. Mechanism of Methyl-Directed Mismatch Repair</td>
<td>44</td>
</tr>
<tr>
<td>1.5. Mismatch Repair in Eukaryotes</td>
<td>45</td>
</tr>
<tr>
<td>1.6. Interaction Between the VSPR and MMR Systems</td>
<td>46</td>
</tr>
<tr>
<td>1.7. Mechanism of VSPR</td>
<td>51</td>
</tr>
<tr>
<td>1.8. Repair of T:G Mismatches in Eukaryotes</td>
<td>52</td>
</tr>
<tr>
<td>1.9. References</td>
<td>69</td>
</tr>
<tr>
<td>Chapter 2. Reconstitution of the <em>E. coli</em> Very Short Patch Repair System with</td>
<td>6</td>
</tr>
</tbody>
</table>
Abstract
Introduction
Experimental Procedures
Results
Discussion
References

CHAPTER 3. The Size of Repair Tracts Created by E. coli Very Short Patch Repair System

Abstract
Introduction
Experimental Procedures
Results
Discussion
References

CHAPTER 4. \(\beta\)-Methylguanine – Formation, Consequences, and Repair

Abstract
Introduction
Experimental Procedures
Results
Discussion
References

CHAPTER 5. Effect of Mismatch Repair Proteins on Methyltransferase Repair of \(\beta\)-Methylguanine in Vivo

Abstract
Introduction
Experimental Procedures
Results
References
ABBREVIATIONS

1mA 1-methyladenine
2D-NMR two dimensional nuclear magnetic resonance
3mA 3-methyladenine
3mC 3-methylcytosine
3mG 3-methylguanine
5mC 5-methylcytosine
7mA 7-methyladenine
7mG 7-methylguanine
Aa amino acid
A adenine
Amp Ampicillin
AP apurinic
AR androgen receptor
ATP adenosine triphosphate
BCNU 1,3-bis(2-chloroethyl)-1-nitrosourea
BER base excision repair
βME beta-mercaptoethanol
bp base pair
C cytosine
Cam Chloroamphenicol
CCNU N-(2-chloroethyl)-N'-cyclohexyl-N-nitrosourea
Da Dalton
DBD DNA binding domain
DMS dimethylsulphate
DMSO dimethylsulfoxide
DNA deoxyribonucleic acid
DTT 1,4-dithio-DL-threitol
E. coli Escherichia coli
EDTA ethylenediaminetetraacetic acid
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMS</td>
<td>ethyl methanesulphonate</td>
</tr>
<tr>
<td>EMSA</td>
<td>electrophoretic mobility shift assay</td>
</tr>
<tr>
<td>ENU</td>
<td>N-ethyl nitrosourea</td>
</tr>
<tr>
<td>ER</td>
<td>estrogen receptor</td>
</tr>
<tr>
<td>FAD</td>
<td>flavin adenine dinucleotide</td>
</tr>
<tr>
<td>g</td>
<td>gravity</td>
</tr>
<tr>
<td>G</td>
<td>guanine</td>
</tr>
<tr>
<td>GR</td>
<td>glucocorticoid receptor</td>
</tr>
<tr>
<td>h</td>
<td>hour</td>
</tr>
<tr>
<td>HEPES</td>
<td>4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid</td>
</tr>
<tr>
<td>HTH</td>
<td>helix-turn-helix</td>
</tr>
<tr>
<td>IDL</td>
<td>insertion-deletion loop</td>
</tr>
<tr>
<td>Kb</td>
<td>kilo-base</td>
</tr>
<tr>
<td>kDa</td>
<td>kilo Daltons</td>
</tr>
<tr>
<td>Kan</td>
<td>Kanamycin</td>
</tr>
<tr>
<td>KOH</td>
<td>potassium hydroxide</td>
</tr>
<tr>
<td>LB</td>
<td>Luria Broth</td>
</tr>
<tr>
<td>LBD</td>
<td>ligand binding domain</td>
</tr>
<tr>
<td>M</td>
<td>molar</td>
</tr>
<tr>
<td>MCS</td>
<td>multiple cloning sites</td>
</tr>
<tr>
<td>MGMT</td>
<td>O(^6)-methylguanine DNA methyltransferase</td>
</tr>
<tr>
<td>MeI</td>
<td>methyl iodide</td>
</tr>
<tr>
<td>MMR</td>
<td>methyl-directed mismatch repair</td>
</tr>
<tr>
<td>MMS</td>
<td>methyl methanesulphonate</td>
</tr>
<tr>
<td>MNNG</td>
<td>N-methyl-N(^\prime)-nitro-N-nitrosoguanidine</td>
</tr>
<tr>
<td>MNU</td>
<td>N-methylnitrosourea</td>
</tr>
<tr>
<td>MTase</td>
<td>methyltransferase</td>
</tr>
<tr>
<td>NaCl</td>
<td>sodium chloride</td>
</tr>
<tr>
<td>NEB</td>
<td>New England Biolabs (Ipswich, MA)</td>
</tr>
<tr>
<td>NEN</td>
<td>New England Nuclear</td>
</tr>
<tr>
<td>NER</td>
<td>nucleotide excision repair</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>NHR</td>
<td>nuclear hormone receptor</td>
</tr>
<tr>
<td>NTD</td>
<td>N terminal domain</td>
</tr>
<tr>
<td>O(^2)mC</td>
<td>O(^2)-methylcytosine</td>
</tr>
<tr>
<td>O(^2)mG</td>
<td>O(^2)-methylguanine</td>
</tr>
<tr>
<td>O(^2)mT</td>
<td>O(^2)-methylthymine</td>
</tr>
<tr>
<td>O(^4)mT</td>
<td>O(^4)-methylthymine</td>
</tr>
<tr>
<td>ODN</td>
<td>oligodeoxynucleotide</td>
</tr>
<tr>
<td>PAGE</td>
<td>poly-acrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCNA</td>
<td>proliferating cell nuclear antigen</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PNK</td>
<td>polynucleotide kinase</td>
</tr>
<tr>
<td>PR</td>
<td>progesterone receptor</td>
</tr>
<tr>
<td>RM</td>
<td>restriction modification</td>
</tr>
<tr>
<td>RT</td>
<td>room temperature</td>
</tr>
<tr>
<td>Rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>SAM</td>
<td>S-adenosylmethionine</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium Dodecyl Sulfate</td>
</tr>
<tr>
<td>T</td>
<td>thymine</td>
</tr>
<tr>
<td>T4PNK</td>
<td>T4 polynucleotide kinase</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris-Acetate-EDTA</td>
</tr>
<tr>
<td>TBE</td>
<td>Tris-Borate-EDTA</td>
</tr>
<tr>
<td>TE</td>
<td>Tris-EDTA</td>
</tr>
<tr>
<td>TLC</td>
<td>thin layer chromatography</td>
</tr>
<tr>
<td>Tris</td>
<td>2-amino-2-(hydroxymethyl)-1,3-propanediol</td>
</tr>
<tr>
<td>U</td>
<td>uracil</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>VSPR</td>
<td>Very Short Patch Repair</td>
</tr>
<tr>
<td>W</td>
<td>adenine or thymine</td>
</tr>
<tr>
<td>Xgal</td>
<td>5-bromo-4-chloro-3-indolyl-beta-D-galactoside</td>
</tr>
</tbody>
</table>
### LIST OF FIGURES

| Figure 1.1. | Organization of the *dcm* and *vsr* genes | 53 |
| Figure 1.2. | Repair of mismatches that arise through methylation and subsequent deamination of cytosine | 54 |
| Figure 1.3. | Unmethylated and hemimethylated *Vsr* substrates | 55 |
| Figure 1.4. | Proposed mechanisms for the hydrolytic deamination of cytidine | 56 |
| Figure 1.5. | Chemistry behind C:G to T:A transition mutations | 57 |
| Figure 1.6. | Model for how VSPR could be shaping the *E. coli* genome | 58 |
| Figure 1.7. | Structure of *Vsr* bound to its product – minor groove side | 59 |
| Figure 1.8. | Structure of *Vsr* bound to its product – major groove side | 60 |
| Figure 1.9. | Geometric characteristics of C:G and T:G base pairs | 61 |
| Figure 1.10. | Methyl direction of mismatch repair | 62 |
| Figure 1.11. | Methyl-directed mismatch repair is bidirectional | 63 |
| Figure 1.12. | ATP-hydrolysis dependent translocation model for MMR | 64 |
| Figure 1.13. | Molecular switch model for MMR, part 1 | 65 |
| Figure 1.14. | Molecular switch model for MMR, part 2 | 66 |
| Figure 1.15. | Molecular switch model for MMR, part 3 | 67 |
| Figure 2.1. | Unmethylated and hemimethylated *Vsr* substrates | 105 |
| Figure 2.2. | PhosphorImager results from *Vsr* nicking assays on unmethylated, hemimethylated, and control substrates | 106 |
| Figure 2.3. | Graphical representation of unmethylated and hemimethylated substrate nicking by *Vsr* | 107 |
| Figure 2.4. | PhosphorImagery results from incubating T:G containing duplexes with VSPR proteins | 108 |
| Figure 2.5. | Graphical representation of *Vsr* nicking results from reconstitution reaction conditions | 109 |

Representative PhosphorImagery results from TLC separation of mononucleotides at interrogation site.

Percentage of VSPR events initiated that went to completion.

Method for detecting VSPR by monitoring the incorporation of radioactivity into substrate DNA.

Extracts from cells over-expressing recombinant Vsr incorporate $[^{32}P]-dCTP$ into short duplex substrates via VSPR.

Graphical representation of results from measuring VSPR via radioactivity incorporation.

Method for detecting VSPR by measuring the depletion of radioactivity signal from substrate DNA.

Distribution of VSPR patch lengths created in vitro.

Addition of surplus Ligase to VSPR reactions.

Addition of Ligase to VSPR reactions containing excess substrate.

Sites of methylation on the DNA bases.

Conversion of MNNG to the active methanediazonium ion by cysteine.

The R and S stereoisomer of methylphosphotriesters.

Base pairing models for O$^6$mG:T and O$^6$mG:C.

Genetic structure and regulation of the adaptive response in E. coli.

Direct reversal of base damage by Ada.

Sequence of ogt and location of PCR primers.

Sequence of ada/alkB and location of PCR primers.

Sequence of mutH and location of PCR primers.

Sequence of mutL and location of PCR primers.
Figure 5.5. Sequence of \textit{mutS} and location of PCR primers \hspace{1cm} 210
Figure 5.6. Sequence of \textit{vsr} and location of PCR primers \hspace{1cm} 211
Figure 5.7. Relative efficiencies of MTase repair in MR defective strains \hspace{1cm} 212
Figure 5.8. PCR mediated gene replacement \hspace{1cm} 213
Figure 5.9. \textit{In vivo} assay for MTase repair \hspace{1cm} 214
Figure 5.10. Hypothetical mechanism for stimulation of MTase repair by VSPR proteins \hspace{1cm} 215
Figure 5.11. Bacterial two-hybrid system utilizing domains of adenylate cyclase \hspace{1cm} 216
Figure 6.1. Model for mechanism of bifunctional toxicants \hspace{1cm} 236
Figure 6.2. Restriction map of pSK278 \hspace{1cm} 237
Figure 6.3. Restriction map of pSK278-hER-LBD \hspace{1cm} 238
Figure 6.4. Restriction map of pSK278-hAR-LBD \hspace{1cm} 239
Figure 6.5. Restriction map of pSK278-hPR-LBD \hspace{1cm} 240
Figure 6.6. Restriction map of pSK278-hGR-LBD \hspace{1cm} 241
Figure 7.1. (A) Piperidine treatment of 5'-d(AATATTGGCCAATATT) treated with compound 1. (B) Retarded mobility of oligonucleotide-1 in the presence of ER-LBD illustrated by EMSA \hspace{1cm} 251
Figure 7.2. Survival of MCF-7 (ER+) and MD-MB231 (ER-) cells treated with 1 or chlorambucil \hspace{1cm} 252
Figure 8.1. Structure and molecular features of lead compound \hspace{1cm} 266
Figure 8.2. Survival of MCF-7 (ER+) and MDA-MB231 (ER-) breast cancer cells after 2 h exposure to estradiol-linked toxicants \hspace{1cm} 269
**List of Tables**

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td><em>E. coli</em> and <em>S. cerevisiae</em> Proteins Required for MMR</td>
<td>68</td>
</tr>
<tr>
<td>2.1</td>
<td>Oligodeoxynucleotides used for cloning Vsr into pET28a and construction of nicking assay and reconstitution assay substrates</td>
<td>113</td>
</tr>
<tr>
<td>2.2</td>
<td>Quantitative results from reconstitution assays containing Vsr, Pol I, and DNA Ligase protein</td>
<td>114</td>
</tr>
<tr>
<td>3.1</td>
<td>Oligodeoxynucleotides used to construct duplexes containing internal radiolabels at prescribed distances away from the mismatched T</td>
<td>142</td>
</tr>
<tr>
<td>3.2</td>
<td>DNA strands containing radiolabels at prescribed distances away from the T to be mismatched</td>
<td>143</td>
</tr>
<tr>
<td>4.1</td>
<td>Relative amounts of DNA alkylated products following reaction of double stranded DNA with MMS and MNU alkylating agents</td>
<td>171</td>
</tr>
<tr>
<td>5.1</td>
<td>PCR primers used for construction of recombination substrates</td>
<td>217</td>
</tr>
<tr>
<td>5.2</td>
<td>PCR primers used for analyzing strain genotypes</td>
<td>218</td>
</tr>
<tr>
<td>5.3</td>
<td>Sizes of genes and PCR amplification products</td>
<td>219</td>
</tr>
<tr>
<td>5.4</td>
<td>Initial plating results from MTase assay on C215 derivatives</td>
<td>220</td>
</tr>
<tr>
<td>5.5</td>
<td>Initial plating results from MTase assay on C216 derivatives</td>
<td>221</td>
</tr>
<tr>
<td>5.6</td>
<td>Initial plating results from MTase assay on C217 derivatives</td>
<td>222</td>
</tr>
<tr>
<td>5.7</td>
<td>Initial plating results from MTase assay on C218 derivatives</td>
<td>223</td>
</tr>
<tr>
<td>8.1</td>
<td>RBA = relative binding affinity for the rabbit uterine ER as compared to estradiol (RBA=100)</td>
<td>267</td>
</tr>
</tbody>
</table>
LIST OF SCHEMES

**Scheme 7.1.** Synthesis of compound 1 .............................................. 250

**Scheme 8.1.** Synthesis of estradiol-linked aniline mustards ................. 268
CHAPTER 1
Mismatch Repair in Escherichia coli
1.1. MUTATION AVOIDANCE

The formation of a base pair other than the canonical Watson-Crick G:C or A:T by a polymerase error, or by a chemical or hydrolytic reaction of an existing base pair, will create a mismatch and lead to a mutation if not repaired. Systems for mutation avoidance are present in both prokaryotic and eukaryotic cells (1-15). These mismatch repair (MR) systems are highly conserved from bacteria to humans and play an essential role in maintaining the integrity of the genetic code and health of the host. Cells deficient in MR exhibit a mutator phenotype, in which the spontaneous mutation rate is increased, and the cells are hyper-recombinogenic. In mammals, loss of MR is related to an increased susceptibility to cancer (6;16). Understanding mismatch avoidance is therefore paramount to human health.

Mismatched base pairs can arise in DNA through multiple processes (17). Errors in replication can cause an incorrect deoxynucleotide to be incorporated into the newly synthesized strand. In fact, in *Escherichia coli* (*E. coli*), MR contributes almost 1000-fold to the fidelity of DNA replication (4;18). Though, under normal conditions of replication, DNA polymerase III is responsible for proper base selection and insertion (19) and has associated proofreading activity (20), mistakes are made with a frequency of $10^{-9}$ to $10^{-10}$ (21). Additionally, slippage during replication can result in the formation of insertion-deletion loops (IDLs), which are composed of bases that are not matched at all. IDLs of less than four nucleotides have been shown to be good substrates for MR (22-24). IDLs of up to 4 or 5 bases can also be repaired by MR, albeit at lower efficiencies (25). In contrast, heteroduplexes containing large regions of nonhomology are not MR substrates (26). Recombination between two DNA strands that do not share complete sequence homology can also result in the formation of IDLs and mismatches that are subject to MR (27;28). Base pairs containing a chemically modified nucleobases can also substrates for MR (29). In particular, one mismatched base pair that can arise in DNA naturally is through the deamination of 5-methylcytosine (5mC). Small amounts of 5mC are present in both prokaryotic (30) and eukaryotic (31) DNA, and the increased propensity of 5mC to hydrolytically deaminate elevates the level of T:G mismatches.
The next sections will describe two mismatch repair pathways in *E. coli*, giving particular attention to the very short patch repair (VSPR) pathway and the cellular components it shares with the methyl-directed mismatch repair (MMR) pathway.

1.2. THE VERY SHORT PATCH REPAIR SYSTEM

In *E. coli*, the underlined cytosine of the pentanucleotide sequence 5'- CCWGG -3' (where W is an A or T) is methylated at the 5-position by a DNA cytosine methyltransferase (32;33) which is the protein product of the *dcm* gene (34). The 5'-methylcytosines (5mC) present in Dcm methylation recognition sites are prone to spontaneous deamination, which results in a T:G mismatch (35;36). If left unrepaired, a C:G → T:A transition mutation will arise following the next round of DNA replication. To avoid the formation of these mutations, the very short patch repair system removes and replaces the mismatched T with the proper base (37).

The first indication that such a system existed was observed in 1981 by Margaret Lieb when, in multifactor crosses, the recombination frequencies of certain amber mutations in the repressor (*cI*) gene of bacteriophage *λ* were higher than predicted by the physical distance separating them (38). Further study of the *am6* mutation revealed that these events arose from C → T transition mutations in a glutamine codon (5'- CAG -3') located within the Dcm recognition sequence 5'- CCAGG -3' (39), and recombinants had the C:G genotype (37;40). Additionally, Lieb demonstrated that a high frequency of these unexpected recombinants included co-repair of markers within 10 nucleotides of the mismatch but not in areas further than 20 nucleotides (39). Taken together, these findings led Lieb to define a very short repair (VSPR) system that acts on the sequence 5'- CTWGG -3' (where T is mismatched with G) to preferentially restore it to 5'- CCWGG -3'. The significance of the 5'- CCWGG -3' sequence, being that for Dcm recognition and methylation, was not overlooked.
Additional evidence for the existence of VSPR and a proposed role in maintaining Dcm methylation sites came from experiments with heteroduplex DNA (41-43). Substrate containing a T:G mismatch in the sequence 5'- CTAGG -3', which could result from the deamination of the Dcm product, was indeed subject to repair with a high (if not complete) bias in favor of replacing the mismatched T. Increasingly, it was accepted that VSPR functioned to reduce the mutation level associated with the spontaneous deamination of 5mC created by Dcm.

This hypothesis was corroborated by Lieb in 1991 when she observed that mutation at a particular hot spot was dependent upon Dcm and VSPR activity (44). In lysogens where only Dcm was present, the mutation frequency at the hot spot was 10-fold higher than in wt bacteria. This hot spot disappeared when both Dcm and Vsr were removed, while addition of Vsr to wild type (wt) bacteria caused a 4-fold decrease in mutation frequency. It is now accepted that a primary role of VSPR is to reduce the number of mutations that result from cytosine methylation and deamination (45).

In the years following the discovery of the VSPR system, a number of genetic and biochemical studies were conducted to implicate the proteins involved in the repair process. In 1987, Lieb reported (46) that the protein products of mismatch repair genes mutS and mutL were involved in VSPR on the basis that the frequency of in vivo crosses in cells lacking either one of these genes was greatly diminished. Strains lacking mutH or uvrD exhibited no difference from wt. Interestingly, the recombination frequency was diminished in dcm6 cells lacking the Dcm methylase, leading Lieb to conclude that the protein product of dcm had a repair function in addition to its methylase function. At the time, however, there was no explanation for this mixed phenotype.

In a report by Dzidic and Radman in 1989 (47) the genetic requirements for hyper-recombination by VSPR were extended to include polA (the gene for Pol I). In fact, the recombination frequency was diminished in cells lacking the polA gene sequences for either its 5' → 3' exonuclease or polymerase domains. This observation
led the authors to conclude that both activities of Pol I were paramount for effective VSPR. In contrast, *xth*, *nth* and *nfo* strains showed wt recombination frequencies.

In 1990, Sohail reported (48) that the two phenotypes of the *dcm* mutants, lacking both methylation and repair activities, had been genetically pinpointed by mapping. Additionally, they cloned genes that complemented the chromosomal mutation for methylation, but not repair, and *vice versa*. The gene essential for methylation, *dcm*, was predicted to code for a 473-amino-acid protein and was not required for VSPR, while the second gene, *vsv*, was predicted to code for a 156-amino-acid protein which was required for VSPR but not for methylation. The two genes were transcribed from a common promoter even though they were found in different translational registers. The 5' end of *vsv* overlaps the 3' end of *dcm* by 7 codons in a +1 reading frame (Figure 1.1). It was hypothesized that this genetic arrangement assures Vsr is always produced along with Dcm to minimize the mutagenic propensity of cytosine methylation. Sohail also noted (48) that although the two genes are transcribed from a common promoter, two separate proteins are produced. Thus, the genetic arrangement of *dcm* and *vsv* was the source of confusion regarding initial experiments conducted on the *dcm6* strain, as they lack *vsv* in addition to *dcm*. This misunderstanding led to the incorrect inference (46) that the *dcm* product is required for VSPR.

One year later, Hennecke showed that the *vsv* gene product was an 18 kDa mismatch endonuclease that produced a single strand nick 5' to the underlined T within duplex DNA sequences 5'- CTWGN -3' or 5'- NTWGG -3' (where N is A, T, C, or G) when the T was opposite a G (49). The incision was mismatch-dependent and strand-specific. These results therefore illustrated that Vsr was the protein that initiates VSPR. Consistent with these observations, strains lacking *vsv* are completely defective in VSPR and have a very high frequency of C → T transitions at 5mC sites (38;50).

Altogether, these experiments provided a model for the mechanism of VSPR (Figure 1.2). Evidence suggested that Vsr initiates repair by producing a nick 5’ to the mismatched T. The previous implication of MutS and MutL in mismatch recognition
suggested these proteins stimulated VSPR by assisting Vsr in finding or incising substrate mismatches. The exact mechanism of Vsr stimulation, however, is still unknown (see below). Subsequent to incision, Pol I removes the T residue with its 5' → 3' exonuclease activity and commences repair synthesis in the usual template-directed fashion, with the short synthesis tracts typical for the enzyme (51). DNA Ligase is reasonably assumed to seal the nicked strand on which repair has been performed.

1.2.1. DCM – IN VIVO AND IN VITRO PROPERTIES

Dcm is the 53 kDa product of dcm, which is located at 43 minutes on the E. coli chromosome (34). The protein methylates the underlined cytosine residue in the sequence 5' - CCWGG - 3' (where W is A or T) at the C5 position of the pyrimidine ring (32). The recognition sequence is palindromic and thus also results in methylation at the corresponding position of the opposite strand, as well (32). The deamination of a 5mC in fully dcm methylated DNA gives rise to a T:G mismatch in a hemimethylated duplex. If the G containing strand contains an unmethylated C, the mismatched duplex is said to be unmethylated (Figure 1.3). In E. coli, approximately 1% of all cytosine residues are methylated by Dcm (52). This translates roughly into approximately 10,000 Dcm recognition sites in the E. coli genome, or on average one recognition site every 400-500 base pairs.

Not surprisingly, the role of Dcm has been a topic of much discussion. However, no obvious biologically interesting phenotype of dcm mutants has been reported (52), and the reason for the existence of the dcm-vsr operon remains unclear (53). The closest characteristic to a phenotype for dcm mutants is the ability of their isolated genomic DNA to be cut by EcoRII endonuclease, which is part of a restriction modification (RM) system and recognizes the same sequence as Dcm (54;55). The EcoRII methyltransferase and Dcm share a high degree of sequence similarity (48), are functionally very similar, and are likely derived from a common ancestor.
An EcoRII RM system was found in drug resistance transfer plasmids (56), and therefore the function of Dcm could be to protect the host’s DNA from invasion (53). In support of this theory, cells bearing Dcm were shown to be resistant to the virulence of a parasitic EcoRII gene complex (53). It is also interesting to note that 5’-CCWGG-3’ is one of the most commonly utilized recognition sequences for RM systems (53). As of March 20th, 2006, the REBASE database (57) identified 169 of the total 3773 RM systems as recognizing the 5’-CCWGG-3’ sequence.

It has also been suggested that methylation by Dcm could be involved in gene regulation (52). A Dcm methylation site has been identified in the promoter region for lexA (58), however no relationship between the Dcm methylation status and expression has been established. Interestingly, some Dcm sites were found to remain unmethylated in dcm+ cells (59), suggesting that some sites are protected by DNA binding molecules or by DNA topology, and could play a role in a yet to be discovered biological process.

The resultant 5mC residues are subject to “spontaneous” hydrolytic deamination (60;61), analogous to the process that converts cytosine to uracil (62). Two chemical mechanisms have been proposed for the hydrolytic deamination of cytosine (62;63). One mechanism involves direct attack at the 4-position of the pyrimidine ring by a hydroxyl ion, and the other involves an addition-elimination reaction with the formation of dihydrocytosine as an intermediate (Figure 1.4). While the mechanism involving direct attack has the advantage of simplicity, relatively little direct evidence has been reported to support its occurrence in reactions involving nucleophilic displacement of the amino group of cytosine. On the other hand, the route involving addition to the C=C double bond is similar to that established for the reaction between hydroxylamine and cytosine derivatives (64). Additionally, it has been directly demonstrated that nucleophiles can add reversibly to the 5,6 double bond of a uracil ring (65). The deamination of dihydrocytosine derivatives (resembling intermediate 2 in Figure 1.4) was also shown to be a rapid process (66). Therefore, evidence suggests that an addition-elimination reaction is a probable mechanism for deamination of cytosine and 5mC in DNA.
Deamination of either cytosine or 5mC gives rise to products that yield C:G → T:A transition mutations upon the next round of replication (Figure 1.5). It is noteworthy, however, that the extent of cytosine and 5mC deamination is not equivalent. The higher rate of deamination for 5mC as compared to C (60:61), coupled to the finding that T:G mismatches are repaired less efficiently than U:G mispairs (67) results in cytosine methylation sequences being hot spots for C → T mutations. Interestingly, the accumulation of these hotspots in E. coli, depends on the phase of growth.

In exponentially growing cells (67), VSPR reduced the 5mC → T mutations by a factor of four. However, cells deficient in Dcm displayed more than 10-fold decrease in the 5mC → T mutation level. Therefore, the inefficiency of VSPR during exponential growth allows Dcm sequences to be hotspots for C → T mutations. In contrast, cells in stationary phase (68) are completely protected by VSPR from Dcm sequences being hotspots. It was observed that the 5mC → T mutation rate in cells that were stored for long periods of time was equivalent to that for cells lacking cytosine methylation. It was also observed that the 5mC → T mutation rate for cells lacking VSPR was consistent with the rate of 5mC deamination at 37 °C. Subsequently, it was shown that the concentration of Vsr in the cell is low during exponential phase and increases when cells start to enter stationary phase (69). To summarize, the ability of VSPR to protect against 5mC → T mutations is linked to the amount of Vsr that is available to initiate repair. Possible explanations for how the amount of Vsr is controlled in the cell (70;71), and why VSPR was designed to be less efficient in exponential phase have been proposed (72), and are discussed below.

In eukaryotes, methylation of cytosine occurs within CpG dinucleotides (73), and has been implicated, among other functions, in the regulation of gene expression (74). A large number of genetic diseases involve C to T changes at CpG sites (75), including β-thalassemia, hemophilia, and cancer (76-79), therefore establishing the maintenance of the sites as critical.
1.2.2. VSR – *in vivo* and *in vitro* properties

Vsr is the 18 kDa product of *vsr*, located at 43 minutes on the *E. coli* chromosome (48). This 156 amino acid protein functions as a strand-specific endonuclease to initiate repair of T:G mismatches in the sequence 5’- CTWGG -3’ (where T is mismatched with a G) (49). These T:G mismatches arise through the methylation of the underlined C in the sequence 5’- CCWGG -3’ by Dcm, followed by deamination of that methylated residue. Vsr, therefore, helps to reduce the mutation level associated with the spontaneous deamination of 5mC created by Dcm.

In 1995, Lieb and Rehmat (80) reported recombination experiments conducted on derivatives of Dcm sequences. Their findings showed that VSPR was most efficient in the pentanucleotide sequence 5’- CTAGG -3’, however there was also significant correction when either the 5’ most C or the 3’ most G was replaced by another base. In fact, some VSPR activity was observed in all of the sequences tested.

The same year, Glasner confirmed the broad sequence specificity of Vsr biochemically (81). Using fluorescence-labeled oligodeoxynucleotide (ODN) substrates, 14 substrates were found to be processed by the enzyme, with varying efficiencies. The substrates differed at one or two positions from the canonical pentanucleotide substrate sequence 5’- CTWGG -3’. Not surprisingly, the sequences 5’- CTAGG -3’ and 5’- CTTGG -3’ were the most reactive substrates with relative rates of 100 and 68% respectively. Sequence derivatives where the first or last base was varied (5’- NTWGG -3’ or 5’- CTWGN -3’) were all repaired with relative rates of 22% or higher. Therefore, Vsr will act efficiently on T:G mismatches in many sequence contexts.

This same study (81) investigated potential relationship between the efficiency of VSPR on individual sequences and their underrepresentation in the genome. It was previously hypothesized (82;83) that if a T:G mismatch resulted from misincorporation of a G opposite a T in the template DNA strand, it would be a potential substrate for both the methyl directed mismatch repair MMR and the VSPR systems. Further, if such a

25
mismatch evaded correct repair by MMR, VSPR might incorrectly carry out “repair” by removing and replacing the T. Therefore, over time, VSPR could operate on mismatches not intended for the system and would provide a driving force in depleting the genome of the T-containing sequences (for example 5’- CTAG -3’) and enriching it for the corresponding C-containing sequences (for example 5’- CCAG -3’). Ultimately, VSPR could have a hand in modifying the *E. coli* genome (Figure 1.6).

Interestingly, Glasner’s data (81) reflected a high quality of positive correlation between the relative second-order rate constants on VSPR removal of mismatched T and the underrepresentation rank for the 14 substrates studied. It was hypothesized that this process would lead to corresponding sequence polymorphisms in the total *E. coli* K-12 population. In a case where two genomes that differ in one or more such polymorphic sites combine, the T:G mismatches that are substrates for Vsr would be repaired in favor of G and thereby transfer the polymorphism to the other genome involved. Additionally, because VSPR replaces a short stretch of DNA downstream of the T:G mismatch, any mismatches within the reach of the DNA synthesis tract would also be repaired in favor of the G containing strand. In conclusion, Glasner postulated that VSPR might assist in the patchwise unidirectional transfer of genetic information from one DNA molecule to another, and contribute to the overall fitness of *E. coli* K-12.

Another function for VSPR appears to be to repair U:G mismatches in Dcm sequences that arise by deamination of cytosine (84). It was found that deamination of cytosine can be promoted by the Dcm methylase. In the absence of S-adenosylmethionine, DNA cytosine methylases promote the formation of a 5-dihydro intermediate that is much more susceptible to deamination than cytosine (85-88). Normally, uracils that arise by the deamination of C are recognized by uracil DNA glycosylase and are repaired by the base excision repair pathway. However, genetic reversion assays (84) revealed that although VSPR is not as efficient as the glycosylase system at correcting U:G mismatches to C:G, the VSPR system does contribute to the correction of U:G mismatches in Dcm sites.
The ability of Vsr to nick 5' to U:G mismatches in Dcm sequences was also studied biochemically (89). Results indicated that the single turnover rate for the U:G mismatch was about half that for the T:G mismatch. These biochemical parameters, however, were determined using protein that was not fully active. Turner observed that much of the Vsr purified from over-expressing E. coli strains is catalytically inactive, and less than 1% of the protein bound DNA (89). In fact, it was not until direct titration experiments were conducted (89;90), in which increasing amounts of Vsr were added to a constant amount of substrate DNA, that the activity of an individual protein preparation could be determined. Ultimately, the authors concluded that the U:G mismatch had a specificity constant ($k_{st}/K_D$) 10-fold lower than that for a T:G mismatch, due to an increase in equilibrium dissociation constant ($K_D$) and a reduction in the rate constant for single turnover ($k_{st}$). Similarly, the specificity constant for a T:G mismatch in an unmethylated duplex (Figure 1.3) was 10-fold less than that for a T:G mismatch in a hemimethylated duplex (Figure 1.3). Crystallographic data containing Vsr bound to duplex DNA (91;92) later suggested that the methyl group present in the hemimethylated duplex may pack against an aliphatic region of the protein (Phe-77, Lys-78, Val-79), and assist binding at the protein DNA interface. These findings support the notion that the biological role of Vsr is to initiate repair of T:G mismatches that arise from deamination of the Dcm product. Even so, the ability of Vsr to incise unmethylated substrates has also been confirmed in vivo (46).

To obtain the highest activity from Vsr protein, it has been reported (93) that current preparations should include freshly transforming E. coli with the plasmid that directs Vsr over-expression, purifying Vsr from the over-expressing strains, and using the purified protein within two weeks of preparation. Furthermore, loss of Vsr activity due to aggregation of protein during storage has been reported (94). Therefore, for best results, Vsr protein must be prepared fresh each time. While many changes have been made to the current methods for preparing Vsr, the reasons for the discrepancy in activity are not understood. Perhaps, as discussed below, the low activity of some Vsr preparations can be attributed to the loss of 14 residues from N-terminus of Vsr upon storage. Whatever
the case, adhering to these precautions has allowed fully active Vsr to be isolated and used to measure the binding and kinetic constants for its interaction with substrate DNA.

The binding affinity of Vsr for its hemimethylated substrate (Figure 1.3) was measured using competition gel-retardation assays (93). Complexes of Vsr and a $^{32}$P-labeled duplex were formed (in the presence of Ca$^{2+}$) and increasing amounts of unlabelled competitor DNA were added. Utilizing these methods, a $K_D$ of 10.9 nM was determined. Single turnover reaction conditions were utilized to saturate the substrate and reflect rates of turnover after substrate binding yet before product release. Under these conditions, the $k_{st}$ was $7.5 \text{ min}^{-1}$. For the unmethylated substrate (Figure 1.3), a $k_{st}$ of $2.8 \text{ min}^{-1}$ for Vsr was independently reported (94). A comparison of the kinetic measurements between the hemimethylated and unmethylated substrates suggests that the activity of Vsr is roughly 2.5-fold more efficient on the hemimethylated substrate. The kinetics of Vsr on these two substrates, however, have never been tested in parallel with fully active protein.

Determination of both $K_D$ and $k_{st}$ values (93) allowed the computation of the single turnover specificity constant ($k_{st}/K_D$), which was $0.7 \text{ min}^{-1} \text{nM}^{-1}$ for Vsr. This constant is generally agreed to be the most suitable metric for comparing the ability of different proteins to bind DNA substrates (95-97). In reactions where the steady state rate of hydrolysis is measured, the specificity constants oftentimes reflect the product release step as this step is commonly rate limiting. In such cases, the steady state constants are not particularly informative for comparisons. In light of the observation that Vsr produces a very stable enzyme product complex (89), the single turnover constants are indeed more useful than the steady state values.

1.2.3. STRUCTURAL DATA FOR VSR

To date, three structures of Vsr have been solved by X-ray diffraction and are available on the RCSB Protein Data Bank. The first structure, 1VSR, (98) was solved to
1.80 Å and contains residues 21 to 156. The second structure, 1CWO, (91) was solved to 2.30 Å and contains residues 2 to 156 of the Vsr protein bound to product duplex DNA. The third structure, 1ODG, (92) was solved to 2.80 Å and contains residues 22-155 of the Vsr protein bound to its reaction product site in duplex DNA. Both of the Vsr-DNA complexes contain hydrolyzed substrate. The following discussion is derived from the data presented in these three publications, and will go into the finer details of the structures and what can be learned upon their inspection.

Vsr has a three-layered α/β/α fold that is stabilized by a structural zinc site (98). The Zn^{2+} is coordinated by four protein residues (Cys-66, His-71, Cys-73, Cys-117) in a tetrahedral arrangement, three of which are on one loop. This coordination is distinct from the majority of zinc sites found in proteins. Typically, the four coordinating residues are divided equally on two strands (99). While Vsr shares no sequence homology with other endonucleases (49), the overall topology of Vsr resembles the type II restriction endonucleases, which have been extensively reviewed elsewhere (100). Members of this family include a growing list of restriction enzymes, as well as the DNA repair nucleases MutH (101), λ exonuclease (102), and archael Holliday junction resolvase (103). However, more specific comparison reveals that Vsr is not a standard type II endonuclease, as previously expected.

The central β sheet of Vsr is composed of two short and four long β strands, and braced on either side by α-helices. The active site of this type II restriction endonuclease family is located in a niche formed in the β sheet and is composed of the catalytic motif PDX_{6-30}(D/E)XK (104;105). For this family of enzymes, these conserved residues are superimposable from enzyme to enzyme and are essential for catalytic activity. Typically, the first two residues are acidic and bind an essential divalent cation. The last residue, typically Lys, is the most variant amino acid between enzymes, and has been proposed to stabilize the transition state (106;107) or orient the attacking water (108;109). Superimposition of the β sheet of Vsr onto members of the type II restriction enzyme family shows that Asp-51 (D) of Vsr (a catalytically essential residue) is the first conserved Asp of the catalytic motif. In keeping with the motif, this
residue does bind an essential divalent cation. The rest of the catalytic motif for type II restriction enzymes, however, is not conserved in Vsr.

Phe-62 in Vsr superimposes onto the second catalytic residue of the motif, however this residue can not engage in metal coordination. The third residue in the motif correlates to a partially conserved His or Asp 64 in Vsr, and while this residue is important for activity, it is not essential. Additionally, His 69, which is close in vicinity to His 64 but positionally distinct from the catalytic residues of type II restriction enzymes, was found to be absolutely required for Vsr activity. These findings have disqualified Vsr from strict conservation with the type II restriction enzyme catalytic motif, and imply that the mechanism for Vsr differs from that of typical type II restriction enzymes. These structural data helped to clarify the discrepancies between how the motif is used for typical type II restriction enzymes and Vsr.

In contrast to typical type II restriction enzymes which utilize highly coordinated metal ions, Vsr coordinates two Mg\(^{2+}\) ions through extensive metal water clusters and has very little direct coordination with the Mg\(^{2+}\) ions. Besides Asp-51, the only other direct metal coordination includes the main chain carbonyl of Thr-63, one residue away from the predicted Phe-62. His-64 and Glu-25 coordinate water molecules in the water metal clusters. Alanine scanning has shown each of these residues to be important for Vsr activity. The crystal structure also revealed that the product DNA duplex, containing a 5’ phosphate on the mismatched thymine and a 3’ hydroxyl on the neighboring cytosine, was involved in Mg\(^{2+}\) coordination. The oxygen on the C3’ of the cytosine coordinates one metal ion. Additionally, the 5’ phosphate of the Vsr reaction product was coordinated with both metal ions. Interestingly, His-69 was coordinating one of the terminal phosphate oxygens. This coordination is atypical since Mg\(^{2+}\) generally prefers the harder oxygen atoms of carboxyl residues as ligands.

In addition to the structural involvement of zinc, Vsr requires Ca\(^{2+}\) or Mg\(^{2+}\) for substrate binding (90). This observation is in contrast to a number of restriction enzymes including EcoRI and BamHI, which require metal ions for catalysis but not for binding
This distinction implies that the metal ions in Vsr play a structural role in addition to a catalytic role. Consistent with this idea, it was observed that Vsr can bind substrate, but not cleave it, in the presence of Ca$^{2+}$ (89). This property allows the binding and cleavage activities of Vsr to be separated, and has been useful for a number of biochemical studies (89;90;93).

In conclusion, the first structure of the Vsr protein alone revealed many important features of the molecule including its topology in relation to type II restriction enzymes, certain conserved elements in its catalytic site, its novel metal binding coordination, and why certain residues are important or critical to catalysis. The second structure of Vsr contained the protein bound to duplex DNA product, and provided valuable information on how the two molecules interact.

The N-terminus of Vsr, which was partially removed in the truncated structure, does not form a single alpha helix as was predicted by secondary structure analysis of the primary sequence (91). Rather, the N-terminus is composed of two short alpha helices separated by a flexible linker (Figure 1.7). These results are in good agreement with limited proteolysis studies which suggested that Vsr had exposed turn-arounds at amino acids 14 and 20 (91;98). The main core of Vsr binds the DNA duplex from the major groove while the most N-terminal helix (Lys-7 to Arg 15) pins down the DNA from the minor groove side. A large opening has been formed in the center of the DNA duplex via intercalation of multiple side chains between the DNA strands. The protein surface shows a strong electrostatic and steric complementarity to the DNA duplex. The basic regions follow along the phosphate backbone of the DNA and a strong acidic patch is found close to the metal binding sites in the catalytic center. Met-14 and Ile-17 project from the minor groove side and leave only enough space for the phosphate backbone to pass on either side.

Remarkably, three aromatic residues (Phe-67, Trp-68, and Trp 86) insert into the DNA duplex from the major groove side (Figure 1.8). Phe-67 stacks with the A:T base pair in the recognition sequence, Trp-68 stacks with the mispaired T, and Trp-86 stacks with two sugars on the strand opposite the mismatched T. The major groove is opened
up, the DNA is significantly unwound, and the duplex is kinked by approximately 44 degrees. It is surmised then that these residues recognize and enhance the structural distortion in DNA resulting from the T:G mismatch (99). Not surprisingly, these three residues are invariantly conserved in the Vsr family (91).

Two of the intercalating residues (Phe-67 and Trp-68), are adjacent to His-69, which is absolutely required for activity and proposed to activate the attacking water. Additionally, all these residues are part of the loop that contains three (Cys-66, His-71, and Cys-73) of the four residues that coordinate the structural Zn$^{2+}$ ion. Therefore, these residues are constrained conformationally and precisely aligned for mismatch recognition and DNA strand incision.

Vsr recognizes its mismatch within the DNA helix, primarily by the structural distortion in DNA caused by wobble pairing between T and G (Figure 1.9). In a T:G wobble pair, the $\mathcal{O}^6$ and N1 of guanine hydrogen bond to the N3 and the $\mathcal{O}^2$ of thymine, respectively. This pairing is preferred by Vsr for a number of reasons. Firstly, the geometry of a T:G mispair is such that it can enter the active site of Vsr. A normal G:C base pair would encounter steric clash between the $\mathcal{O}^2$ of cytosine and Thr-19. However, because the $\mathcal{O}^2$ of T in a T:G mispair is pushed into the major groove, these mismatches can fit into the Vsr active site. Secondly, the wobble pairing achieves a shape such that the scissile phosphate of the mismatched thymine is in proximity to the catalytic residues in the active site. In other words, normal base pairs would be situated such that their phosphate group would not be proximal to the Vsr catalytic residues. Thirdly, the wobble pairing disrupts the DNA stacking between the mismatched G and the base on the 5' side such that it causes (1) an opening of the DNA by unwinding, (2) increased distance between G and its 5' neighbor, and (3) the ability of Vsr to recognize and bind the G:T mismatch by intercalating within the recognition site. In addition to the ability of Vsr to recognize T:G mismatches via the drastically different geometry, there are a few hydrogen bonding contacts between the mismatched base pair and Vsr residues. The invariant Asn-93 binds to the $\mathcal{O}^6$ of thymine and presumably plays a role in cytosine discrimination. Guanine is bound at its $\mathcal{O}^6$ position by Lys-89 and at its N$^2$ by the main
chain of Met-14. Biochemical analysis of mutants in these residues will be necessary to
determine the extent to which these interactions are required for recognition.

The five base pair recognition sequence occurs through an interlaced network of
direct and water mediated hydrogen bonding interactions. Two notable features are that
(1) Arg-120 has the only direct contact with the major groove, and it binds to the G:C
base pair that is required for recognition, and (2) the third base pair in the middle of the
recognition sequence (which can be A:T or T:A) has the most interactions. The sequence
specificity of Vsr may therefore lie primarily in the ability of duplex DNA to tolerate
intercalation.

The third structure of Vsr was obtained using a DNA duplex that generated a site
 corresponding to the hemimethylated cleavage product of Vsr (92). This second
protein/DNA structure is quite similar to the co-crystal structure previously reported (91),
but some differences are worth mentioning. The structure shows a 60° angle along the
minor groove as compared 44° in the previous structure solved by Tsutakawa.
Additionally, the middle A:T base pair (in the sequence 5’- CTAGG -3’) was in a
Hoogsteen conformation, with the A being in the syn conformation. This pairing is
believed to exist in bulk DNA at low levels (in equilibrium with the Watson-Crick form)
but has not, until recently, been observed in a protein-DNA complex (110). To explain
why this Hoogsteen pairing was first observed in this structure, it is relevant to note that,
in this structure, Vsr is bound to a hemimethylated DNA substrate. The methyl group on
the C5 position of C would potentially clash with the C5 methyl group of T if the
thymine residue was not positioned closer to the minor groove through a Hoogsteen pair.
Therefore, it was hypothesized that, in this structure, the Hoogsteen A:T base pair is more
energetically favorable than a normal A:T base pair that would introduce methyl-methyl
clash.
1.2.4. CATALYTIC MECHANISM OF VSR

Taken together, these structural data \((91;98)\) lead Tsutakawa to propose a catalytic mechanism for Vsr \((99)\). When Vsr binds the mismatch-containing DNA duplex, the scissile phosphate is coordinated and stabilized by both metal water clusters. His-69 abstracts a proton from one of the waters in the metal water clusters and the resulting activated water attacks the phosphate group between the mismatched thymine and its 5’ cytosine neighbor. The orientation of the His-69, terminal phosphate, and deoxyribose oxygen groups suggests an inline attack. The lack of significant side chain conservation in the area surrounding the deoxyribose oxygen suggests that its protonation is from a water molecule and not the protein.

Additional support for His-69 acting as a catalytic base comes from the observation that replacement of this residue with alanine results in a variant with dramatically reduced activity \((91;98)\). Additionally, a hydrogen bond between His-69 and Asp-97 would increase the pKa of His-69, facilitating its role as a base \((99)\). Several endonucleases including DNaseI \((111)\), exonuclease III \((112)\), and HAP1 \((113)\) use a similar Asp-His pair to provide basic catalysis.

The two magnesium water clusters are also probably required for stabilizing the additional negative change that develops on the scissile phosphate in the transition state. It has also been hypothesized that one of the magnesium ions may have a role in correctly aligning the activated water molecule for attack, while the other magnesium ion may provide Lewis acid catalysis by coordinating to the leaving group and neutralize the developing negative charge on the scissile phosphate \((93)\).

Furthermore, the possibility of His-69 involvement with direct attack of the scissile phosphate to form an enzyme-DNA covalent intermediate, subsequently hydrolyzed by water, was excluded by Elliot and coworkers on the basis of stereochemical analysis of the phosphate group \((93;114;115)\). They deduced that the formation and subsequent hydrolysis of an enzyme-DNA intermediate would proceed
with two displacements at the phosphorus atom and therefore two inversions of configuration. Ultimately, the resulting product would have overall retention of stereochemistry (115;116). Conversely, if His-69 deprotonates a water molecule which then attacks the scissile phosphate directly, only one displacement occurs at the phosphate and the resulting product would have overall inversion of stereochemistry. Experiments conducted using H$_2^{18}$O and Vsr substrates that contained the Rp phosphorothioate isomer at the scissile phosphate site showed that the Vsr reaction proceeds with inversion of configuration at the scissile phosphate, and, therefore, His-69 functions as a base to deprotonate the attacking water molecule.

The final catalytic requirement for nucleases is either a Lewis or a proton acid to interact with, and thereby stabilize, the developing negative charge on the leaving group (114;117). In the case of Vsr, the leaving group is the 3'-bridging oxygen atom at the scissile phosphate. To probe if stabilization of the leaving group is achieved through coordination with a metal center, this group was modified to contain a 3'-phosphorothioate (which coordinates Mn$^{2+}$ better than Mg$^{2+}$) and binding and hydrolysis rates were measured (93). While binding of Vsr to this modified substrate was unchanged, hydrolysis in the presence of Mg$^{2+}$ was strongly reduced. However, cleavage was fully restored to wild type levels when Mg$^{2+}$ was replaced with Mn$^{2+}$, strongly suggesting that one Mg$^{2+}$ serves as a Lewis acid and stabilizes the negative charge on the leaving group as it develops.

In summary, when Vsr binds to a T:G mismatch in duplex DNA, the scissile phosphate is coordinated and stabilized by two Mg$^{2+}$ water clusters. His-69, acts as a general base and abstracts a proton from a water in the metal water cluster. The resulting activated water conducts a direct inline attack on the phosphate group between the mismatched thymine and its 5' cytosine neighbor. The increased negative charge in the transition state is localized on the two non-bridging oxygen atoms and is stabilized by the two Mg$^{2+}$ water clusters. One of the Mg$^{2+}$ ions also acts as a Lewis acid to labilize the leaving group by neutralizing the negative charge that develops as the reaction proceeds.
Evidence of the existence of mismatch repair in prokaryotes and the idea of strand discrimination originated from early work involving transformation studies in *Streptococcus pneumoniae* (*S. pneumoniae*) (118). In a *S. pneumoniae* wt strain, the efficiencies to which different transformants integrated was varied. Mismatch repair was proposed to be responsible for this phenomenon because heteroduplexes formed during integration would have contained mismatches, since the regions involved in recombination were not completely homologous. It was also observed that the transformation efficiency was dependent not only upon the system’s preference for undertaking repair on the donor strand, underscoring strand discrimination, but also on its preference for the type of mismatch on which repair was carried out. In other words, different mismatches were repaired with different efficiencies.

Treatment of *S. pneumoniae* wt cells with mutagens led to the isolation of mutants that did not discriminate between high and low efficiency markers (119). The mutations responsible for this phenotype were designated *hex* (originally for “high efficiency unknown [x]”, now for heteroduplex repair deficiency) (118). Interestingly, *hex* mutants showed an elevated level of mutation accumulation (120), suggesting that the *hex-* dependent MMR system could play a role in mutation avoidance.

It was hypothesized that strand discrimination was established by the ability of the system to recognize that donor strands possessed strand breaks at their ends (121). In support of this hypothesis, low efficiency markers were sensitive to donor DNA that had been UV irradiated (and thus contained a high number of strand breaks) in a fashion that required the *hex* genes (118;122). Therefore, the mutator phenotype of the *hex* mutants was likely due to their inability to recognize strand breaks and discriminate between strands (123-125). Since the lagging strand of DNA is continuously forming breaks at the ends of its Okazaki fragments during replication (124;126) it is believed that *hex*
dependent MMR is programmed to carry out repair on errors that arise during replication to the newly synthesized strand.

Methyl-directed mismatch repair in *E. coli* was believed to be similar to that for *S. pneumoniae*. Corroboration for that hypothesis was two-fold. It was observed that HexA from *S. pneumoniae* is homologous to the MutS mismatch recognition protein from *Salmonella typhimurium*, and could complement an *E. coli* strain deficient in MutS (127). Additionally, the long standing hypotheses that repair was directed to the newly synthesized strand *via* the methylation state of d(GATC) sites (128;129) was confirmed (130).

1.4. METHYL-DIRECTED MISMATCH REPAIR IN *E. COLI*

Methyl-directed mismatch repair (MMR) in *E. coli* (131) is initiated when the mismatch recognition protein MutS finds a substrate in DNA. Upon mismatch recognition, MutS interacts with its canonical partner MutL to activate the latent endonuclease activity of MutH (132). MutH is a type II restriction endonuclease that cleaves the newly synthesized strand, containing the incorrect base, adjacent to the A at the nearest unmethylated d(GATC) site (132). The ability of MutH to incise the correct strand in the palindromic d(GATC) sequence is achieved by the methylation status of the two strands (Figure 1.10). For a short time after replication, the d(GATC) sequence on the nascent DNA strand is not methylated (133;134). MutH can only incise unmethylated d(GATC) sequences (132). Thus, repair is directed by the methylation status of the DNA, and, therefore, this characteristic has been reflected in the naming of the repair system.

The incision site, which can be either 5' or 3' to the error, is the entry point for removing the strand which contains the mismatch. MutL has been shown to play a pivotal role in loading DNA helicase II and recruiting single-stranded DNA binding protein (135;136). Additionally, contacts between MutL (as part of a MutL-MutS
complex) and DNA helicase II have been shown to activate the helicase (135;136). Therefore, MutL plays an important role in communicating the direction of the error. Models for how this communication is accomplished are discussed below.

Working in concert, the helicase and single-stranded binding protein expose single stranded DNA that is subsequently digested by either 3' or 5' exonucleases, depending on the orientation of the error (17;137). Originally, Exo I was implicated in the degradation of DNA in the 3' direction (138). It was later shown that Exo X (139) and Exo VII (140) could provide this function as well (141;142). On the other hand, degradation of DNA in the 5' direction has been shown to involve RecJ or ExoVII (143). Subsequent to excision of the error containing strand, Pol III and DNA Ligase to resynthesize and ligate the DNA to restore the native structure. Remarkably, the resynthesis of a significant tract of DNA can be mandated by an incision site located up to a couple of thousand bases away from the replication error (128;144). Though all repair events are not necessarily long (145), MMR is sometimes referred to as long-patch repair.

1.4.1. DAM - IN VIVO AND IN VITRO PROPERTIES

Methylation of the exocyclic $N^\delta$ position of adenine in d(GATC) sequences is entrusted to the DNA adenine methylase, Dam (146), which is a 278 amino acid protein that has an approximate molecular mass of 32 kDa (147). The importance of methylation was confirmed by the observation that dam- strains of E. coli have increased rates of spontaneous mutation (148;149). The responsibility of dam+ dependent methylation in providing the strand discrimination signal was later established in experiments with heteroduplexes made from two strands of different methylation status (129;146;150-152). In hemimethylated sequences, repair was highly based towards the unmethylated strand, with the methylated strand serving as the template during repair. Experiments on heteroduplexes showed that MMR will not act unless an unmethylated d(GATC) site is present (153). Similarly, reduced expression of Dam results in reduced repair and hyper-
mutation (33;149). Over-expression of Dam also results in an increase in the spontaneous mutation rate (154;155) because the window in which MMR has to act is shortened (156) and, as a consequence, more mutations evade repair.

1.4.2. MutS - IN VIVO AND IN VITRO PROPERTIES

The mutS gene codes for an 853 amino acid protein (157) with a monomer molecular mass of approximately 97 kDa (158). The protein belongs to the Walker A/B sequence motif of the NTPase families (159). Sedimentation experiments revealed that a MutS homodimer acts as the recognition component of MMR by distinguishing mismatches (160) and IDLs (23). Binding of MutS to mismatches contained in stretches of up to 20 nucleotides was first demonstrated by DnaseI footprinting analysis (160). The MutS footprint was also found to expand in the presence of MutL and ATP (161). These conditions also activated a MutH dependent endonuclease activity at a hemimethylated d(GATC) site (162).

MutS itself demonstrates a weak ATPase activity (163), and proper functioning of this enzyme has been attributed to the essential need for ATP hydrolysis (132). Indeed, introduction of mutations in the phosphate binding loop (P-loop) motif of the MutS ATP-binding site results in a dominant negative mutator phenotype (164). However, the role of hydrolysis and the mode of signal transduction between mismatched nucleotides and the excision machinery is under debate (165).

MutS has varying affinities for the eight possible mispairs, ranging from 10 to more than 1,500-fold higher for mismatched DNA as compared with perfectly matched DNA (166;167). The relative affinities of MutS for mismatches ranked in decreasing order are T:G > C:A > G:G, A:A > T:T, C:T, G:A > C:C (150;158;160). The C:C mismatch is recognized by MutS (158) but has not been reported as subject to MMR. Although mispairs leading to transition mutations (purine to purine or pyrimidine to pyrimidine) are, for the most part, better repaired than those leading to transversion
mutations (purine to pyrimidine or visa versa), increased efficiency of repair may be dependent upon sequence context. It has been reported that an increase in GC content in the area surrounding the mismatch may increase repair efficiency (168).

Structures of MutS from *Thermus aquaticus* (169) and *E. coli* (170;170-175) have been solved. The overall structure of a carboxyl-truncated *E. coli* MutS bound to a T:G mismatch shows an asymmetric MutS dimer, where only one monomer is bound to ADP and is recognizing the T:G mismatch. The ATPase domain of the other monomer is unoccupied and reinforces the notion that the two subunits do not share equal roles in mismatch recognition. The monomer units make contacts at the top and bottom to form a dimer that has two large holes in the middle. The DNA is bound at one of these holes and the other is unoccupied. It has been proposed that this hole could serve as a second DNA binding site to allow for MutS to simultaneously bind DNA at the mismatch site and at the MutH cleavage site by looping the DNA through this second hole (169).

Interestingly, the ATPase sites are located at the end of the complex opposite that which binds DNA, invoking questions as to how ATP hydrolysis is coupled to mismatch recognition. While it has been observed that the binding of DNA by MutS stimulates ATP hydrolysis (176), which results in a decrease in DNA binding affinity (177), the physical basis for how DNA binding and ATPase activity is coupled is not understood.

The duplex DNA is in the B-form and has a strong kink of approximately 60° at the mismatch site. The bases at the mismatch site are unstacked and have atypical roll, tilt, and twist angles. The kink in the DNA is manifested in sugar pucker rearrangements from C2'-endo (typical for B-form) to C3'-endo (typical for A-form) for a stretch of six nucleotides.

The N-terminal mismatch binding domain of one subunit inserts into the minor groove and makes direct contacts with the mismatch. At the center of the DNA binding interface, a conserved Phe-36 residue is stacked against the unpaired thymine, and the protein forms hydrogen bonds with a number of the unpaired bases. Additionally, Glu-38
is engaged in hydrogen bonding with the mismatched thymine residue. In normal base pairs, these groups contacted by the protein would not be exposed. Therefore, it appears as though MutS probes for the mismatch by looking for contacts that would not exist in perfectly matched DNA.

Although MutS readily forms tetramers (160;176;177) the carboxyl-truncated fragments of MutS do not crystallize as such (169;171) and the orientation of the two dimers is therefore unknown. Perhaps such a structure would reveal more about how signaling between MutS and other mismatch proteins is achieved. Regardless, the fact that bacterial MutS dimers form structural heterodimers is interesting because of the use of true heterodimers by eukaryotes (178). This asymmetry could be involved in strand specific signaling of mismatches or evidence that the MutS monomer units operate through conformational switching between the two units (170).

There is recent evidence that MutS, Pol I, and DNA Ligase all interact with the β-clamp accessory protein (179). This protein is typically associated with Pol III and is responsible for processive DNA replication. The observation that the β-clamp associates with DNA repair proteins suggests that it is involved in a variety of reactions in addition to chromosomal replication, and that certain instances of repair could be highly orchestrated involving multi-component repairosomes.

1.4.3. MutL - in Vivo and in Vitro Properties

The mutL gene has been cloned from E. coli (180) and encodes a 70 kDa protein monomer that can be over-expressed in E. coli and purified (161). MutL exists as a homodimer through interactions in its C-terminus (161;181). The MutL dimer is recruited to the ATP bound MutS complex (161), in a manner that does not require binding or hydrolysis of ATP by MutL (182).
MutL, like MutS, has an ATPase domain that is critical to MMR (183). Binding of ATP by MutL occurs at the N-terminus, causing conformational changes that induce additional dimer contacts at its N-terminus (184). N-terminal dimerization, however, does not require ATP hydrolysis, as nonhydrolyzable ATP analogs also produce N-terminal contacts (184). Mutations that inactivate hydrolysis, but not binding of ATP by MutL, are strongly dominant negative to MMR in vivo (185) and in vitro (186). Hydrolysis of ATP is therefore a critical for a step of MMR after the interaction between MutS and MutL. Possibly, hydrolysis of ATP by MutL serves to activate MutH (101) and assists the actions of DNA helicase II (186). In any case, the changes that take place upon hydrolysis of ATP by MutL help to coordinate and regulate interactions critical for MMR (184). In this regard, MutL has been deemed a molecular matchmaker (187).

Structural data for MutL (181;183;184;188) reveal that the MutL ring-like dimer contains many residues for DNA binding on the inner surface of the ring. These data suggest that MutL, like MutS, encircles DNA and possibly assists strand specific excision by looping out the DNA to bring the mismatch site and MutH incision site within close proximity (181).

1.4.4. MutH - IN VIVO AND IN VITRO PROPERTIES

The mutH gene codes for a 229 amino acid protein with a molecular mass of approximately 25 kDa (162). MutH is structurally similar to a variety of type II restriction enzymes (101). The protein consists of two globular domains connected by a flexible junction, and a comparison of multiple MutH structures reveals that this junction allows for multiple protein conformations (189). This flexibility could explain, in part, why association with MutL is required for activity. MutL might assist in stabilizing MutH in an active conformation. The active site of MutH is located in a shallow cleft between the two globular domains and is coated with a number of basic residues that likely attract the negatively charged phosphate backbone of the DNA substrate.
MutH has an extremely weak Mg$^{2+}$ dependent endonuclease activity that nicks hemimethylated DNA, 5' to the G in the sequence d(GATC) (162). Initial measurements of the endonuclease activity for the purified enzyme demonstrated less than one cleavage event per hour per MutH monomer, leading to the hypothesis that the activity of MutH was somehow stimulated by other components of MMR (131;162). This notion was strengthened by the observation that the activity of the same protein preparation was approximately 20 to 70-fold higher when added to a reconstituted methyl-directed mismatch repair system. It was later demonstrated that a MutS-MutL mismatch ternary complex, and ATP hydrolysis, allowed for full activation of MutH (132). Interestingly, high concentrations of MutL have been shown to be sufficient for activating MutH in an ATP-dependent manner (101). The need for the MutH and/or the d(GATC) incision site can be bypassed altogether if a nick is present in the undermethylated strand (131;153).

An important observation is that MutH can cleave both DNA strands if the d(GATC) sites are completely unmethylated (132). Decades earlier, it was observed that dam mutants, which contain fully unmethylated d(GATC) sites, were killed by growth in the presence of base analogs that form substrates for MMR in DNA, such as 2-aminopurine (190). This observation lead to the suggestion that, in the absence of methylation, MMR would initiate repair on either strand, leading to the formation of double strand breaks (190). This hypothesis received criticism on the grounds that mismatches did not occur frequently enough to produce excision tracts that would overlap (191). Therefore, the finding that MutH can create a double strand break from one mismatch, assists in explaining cell death in dam$^-$ strains.

Nicking of the undermethylated strand by MutH can take place either 3' or 5' to the mismatch (Figure 1.11) (132). The choice of which single stranded exonuclease will carry out the excision and degradation of the DNA is contingent upon DNA helicase II unwinding the DNA at the nick site in the appropriate direction towards the mismatch (143). Though the appropriate direction for DNA unwinding is coordinated by MutL (186), the precise method of signaling is not understood. The next section describes two prominent models for how MMR is orchestrated.
1.4.5. MECHANISM OF METHYL-DIRECTED MISMATCH REPAIR

One model (18;145) which is commonly referred to as the Hydrolysis Dependent Translocation Model (Figure 1.12) postulates that the assembly of a MutS/MutL complex at the mismatch site uses ATP hydrolysis to motor bidirectionally and to mediate the formation of an α-shaped loop structure containing the substrate mismatch (177). MutL enhanced the rate of MutS-mediated DNA loop formation and both MutS and MutL appeared to be bound at the base of the α-loop. Loop formation required a mismatch and ATP. Hydrolysis of ATP (and not just binding) was necessary for translocation, as nonhydrolyzable ATP analogs were not sufficient to promote loop growth. The original model was modified (192) to accommodate the observation that MutS appears to move a large distance on the DNA per ATP hydrolysis event. Such a DNA tracking process was envisioned to link MutS mismatch recognition with MutH endonuclease activity at a nearby d(GATC) site as well as to provide directionality for subsequent loading of the UvrD helicase and the appropriate ssDNA exonuclease.

An explanation as to how MutS can translocate bidirectionally (193-195), is based on the idea that the nucleotide binding sites of MutS can act as “latch (L) sites” that can be in the open or closed conformation depending on their occupancy. If ADP is bound to one L site, it will be in the closed conformation and not allow translocation at that site. Simultaneously, however, the other L site will be ATP bound and open, therefore allowing translocation in that direction. Hydrolysis of ATP in the second site, accompanied by replacement of ADP with ATP in the first site, allows translocation to occur in the other direction. This process repeats, so long as ADP and ATP are around, to achieve bidirectional translocation.

Another model (165) commonly referred to as the Molecular Switch Model (Figure 1.13, 1.14, and 1.15) is based on the observations (182) that ADP-bound MutS binds to mismatched DNA and then exchanges ADP for ATP. Nucleotide exchange is said to result in conformational changes that destabilize mismatch binding and give rise to an ATP-bound MutS sliding clamp that diffuses along the helix, independent of ATP-
Interestingly, ATP binding by the human MutS homolog, hMSH, has been shown to result in the formation of a DNA sliding clamp capable of hydrolysis-independent diffusion for several thousand nucleotides (198). It is also noteworthy to mention that only the nucleotide-free and ADP bound structures of MutS have been solved because infusion of ATP or ATPγS disintegrates the MutS crystal (172). It appears that ATP binding by MutS invokes conformational transition(s). Iterative loading of multiple clamps was responsible for marking the mismatch and creating a gradient of clamps, providing directionality, along the duplex surrounding the mismatch site (198;199). MutL is said to interact with ATP-bound MutS clamps to produce complexes that can interact with, and activate, MutH endonuclease. Observations indicate that interaction between the MutL complex and MutH increases the ability of MutL to bind ATP, and in turn the ATP-MutL complex activates the endonuclease activity of MutH (182). MutL, in this sense, is a protein activated molecular switch. Subsequent to incision, UvrD helicase begins unwinding the DNA with an activity that is enhanced by MutS and MutL. This property was hypothesized to direct the helicase towards MutS-MutL complexes and therefore towards the mismatch.

1.5. MISMATCH REPAIR IN EUKARYOTES

Eukaryotic mismatch repair (6;8;16) is functionally homologous to that of *E. coli*, plays an important role in mutation avoidance, and is critical to health. Defects in human mismatch repair have been linked with a predisposition to hereditary nonpolyposis colon cancer and ovarian cancer. Eukaryotic MMR proteins also have been associated with the triggering of programmed cell death.

Many eukaryotic DNA repair enzyme homologs have been identified and their functions characterized (Table 1.1) (18;200;201). However, to date, no MutH homologs have been assigned. Additionally, the mechanism by which strand discrimination is achieved is not thought to include methylation, as in *E. coli*. Rather, strand discrimination is possibly directed by nicks which exist at the ends of Okazaki fragments.
in lagging strand synthesis and at the growing 3' end of the leading strand (202;203). Another possibility is that the proliferating cell nuclear antigen (PCNA) could assist in discriminating the parental strand from newly synthesized strand through its interactions with various MMR proteins (204;205).

Interestingly, the homologs to *E. coli* MutS do not form asymmetrical homodimers as described above. Rather, they form a variety of protein-protein heteroduplexes depending on the substrate. It is hypothesized that subsequent to recognition of the MMR substrate, the MutS complex will form a clamp-like structure and slide along the DNA. Formation of a reairosome proceeds by association of MutL-like heterodimers and PCNA to the MutS-like complex. This complex then identifies the nascent strand and initiates repair by some mechanism yet to be described. Repair is then carried out by proteins including DNA polymerase δ, replication protein A (RPA), replication factor C (RFC), and exonuclease EXO I/HEX 1.

### 1.6. Interaction Between the VSPR and MMR Systems

A comparison between the long patch (MMR) and short patch (VSPR) mismatch repair systems discussed above reveals that they vary in a number of ways. As highlighted by the names of the repair systems, the average length of the repair tract associated with each is quite different. While a VSPR repair tract rarely exceeds 10 nucleotides in length, that for MMR has been observed to reach thousands of nucleotides. MMR repairs a large number of substrates including a multitude of base-base mismatches, IDLs, and chemically modified bases. VSPR, on the other hand, repairs T:G and U:G mispairs to yield C:G in all cases. MMR has the requirement that the DNA must contain an unmethylated d(GATC) sites for MutH incision. In contrast, the methylation status of d(GATC) sites has no effect on VSPR. However, VSPR does have sequence context specificity. Therefore, many differences exist between these two repair pathways.
One thing in common between these two repair pathways is that they share the ability to act on T:G mismatches. Furthermore, VSPR has the opportunity to act incorrectly on T:G mismatches that arise from the misincorporation of G opposite T. These types of mismatches should be processed correctly by MMR before VSPR gets a chance to act, thus avoiding a fixation of the mutation. Therefore, given the intention of cellular repair systems to avoid DNA mutations, the VSPR and MMR systems should be coordinated with each other. This section discusses how the two systems may interact.

As highlighted above, both the VSPR and MMR systems use MutS and MutL for initial mismatch recognition. It was originally thought that VSPR could not function in cells lacking the *mutS* or *mutL* gene (41-43;46). However, it was subsequently shown that the level of VSPR in *mutS* and *mutL* mutants was only reduced, not eliminated (70). The reduction of VSPR was also observed upon over-expression of MutS from a multicopy plasmid (80;206), suggesting that the level of MutS maintained by the cell is optimal for its functions in VSPR. In contrast, introducing a MutS producing plasmid into *mutS* cells does not affect MMR (164). These observations led Lieb (206) to study a number of *mutS* alleles, that were dominant negative for MMR, for their effects on VSPR. Lieb found that many of the MutS mutants (Class II) did prevent VSPR. These alleles had previously been found to reject complementation by *mutS*, *mutL*, or *mutH* plasmids (164), indicating that the host mutant allele products bind to mismatches and prevent further access. Some mutants (Class I), however, were found to not disrupt VSPR in *mutS* cells and actually had the effect of stimulating VSPR in *mutS* cells. These results indicated that specific MutS alleles, lacking functions for MMR, can still participate in VSPR.

A study of these Class I alleles revealed that some were complemented by over-expression of MutL and/or MutH from plasmids (164), indicating that the nature of the host *mutS* mutations affected the way in which the alleles interacted with the MutL and/or MutH proteins. While VSPR is compromised in cells lacking MutL (41-43;46), *mutH* strains retain full repair activity (46). Therefore, these alleles were capable of interacting with MutL for VSPR activity but not for MMR. Not only did these observations...
highlight that the roles of these recognition proteins in VSPR and MMR are different, but also that MutL plays an integral role in each repair system.

Evidence that MutL actually balances the two repair systems came from experiments done in the Cupples Lab. Mutational studies in cells that were over-expressing Vsr revealed that not only did high levels of Vsr result in T → C mutations, but also a large number of other unexpected mutations, including frameshifts (207,208). It was unlikely that this mutator phenotype resulted from competition between Vsr and MutS for substrates. Rather, a more likely explanation was that Vsr was competing for a component of, and preventing, MMR.

*In vitro* studies revealed that binding of Vsr to its substrate is promoted by MutL catalytically (209). In contrast, stimulation of MutS binding to the Vsr substrate by MutL was increased by 6-fold stoicheometrically. In agreement with these findings, bacterial two-hybrid studies have confirmed *in vivo* interaction between Vsr and MutL (210). Additionally, interaction between Vsr and MutL inhibited the ability of MutL to dimerize, interact with MutS and/or MutH, and to mediate a previously unknown interaction between MutS and MutH. No interaction between Vsr and MutS and/or MutH was observed. All these data together strongly suggested that Vsr makes critical contacts with MutL and that these interactions can detract MutL from its role in MMR.

Further investigation (94) into the nature of the interaction between Vsr and MutL revealed that a N-terminal truncation mutant of Vsr (Vsr-A14), cannot be stimulated by MutL *in vivo*, and promotes diminished mutagenicity. In contrast, four derivatives of full-length Vsr that were enzymatically inactive (F67A, W68A, H69A, and H71A), were as mutagenic as the wt protein. Therefore, the ability of Vsr activity to be stimulated by MutL, and the ability of the Vsr-MutL interaction to be detrimental to MMR, resides in the N-terminal domain (NTD) of Vsr and not in the proteins incision activity.

Interaction between Vsr and MutL was shown to be independent of the first 19 residues of Vsr (210). Therefore, the observation that the Vsr-A14 mutant has diminished mutagenicity is not simply because it does not bind MutL. How then does Vsr inhibit
MutL from participating in MMR? To better understand how inhibition might occur, it will be useful to revisit the role of ATP hydrolysis by MutL in MMR.

Studies of MutL revealed that the protein must bind and hydrolyze ATP to be functional in MMR (183). Further investigation (described above) revealed that hydrolysis of ATP by MutL stimulates the endonucleolytic activity of MutH, and results in loss of N-terminal contacts in the MutL dimer (184). Additionally, re-dimerization of the N-termini is thought to be the rate limiting step in the ATPase cycle of MutL (184). Therefore, after stimulating the protein responsible for nicking DNA, MutL needs to bind another ATP and dimerize at its N-terminal before it can take part in another repair event. It is plausible that MutL goes through a similar cycle with Vsr. That is, interaction between Vsr and MutL, invokes MutL to hydrolyze ATP to convert Vsr into its active form. In the process, MutL is rendered incapable of assisting another repair event for a short time. Given the observation that the Vsr-Δ14 mutant has diminished mutagenicity, as compared with the wt enzyme, it is plausible that the NTD of Vsr is involved in MutL dependent stimulation of the endonuclease. Temporary incapacitation of MutL would explain how over-expression of the full length Vsr is mutagenic. It also explains how the Vsr-Δ14 is not mutagenic. If the NTD of Vsr is not present for stimulating the activity of Vsr, MutL will not give up its activated state by hydrolyzing its ATP. To understand how MutL interacts with the NTD of Vsr, to stimulate endonuclease activity, we must look at how the NTD of Vsr is ordered.

Early on, trypsin and chymotrypsin proteolysis studies revealed that the NTD of Vsr is unstructured (98), and it had to be removed before the free protein would crystallize (98). It was also observed that the tail was labile in vitro (see below) (81). However, crystallographic studies revealed that, at some point, the NTD of Vsr becomes ordered and forms an α-helical structure that lies across the minor groove of DNA (91). In purified protein, ordering of the NTD of Vsr appears to be very slow (211). It was therefore hypothesized that the interaction between MutL and full length Vsr in vivo helps the NTD of Vsr achieve an ordered conformation that stabilizes Vsr binding to DNA (94). This notion is consistent with data showing that MutL can stimulate Vsr directly (89).
It is interesting to note that a N-terminally truncated form of Vsr, missing 12 to 14 residues, was observed to form upon storage (81). This finding shed light on the lability of the NTD, and also complicated initial attempts to characterize Vsr biochemically (89). Early preparations of Vsr were reported to contain enzyme with only 1% activity (89). Interestingly, reliable biochemical studies have reported a $k_{cat}$ of 2.8 min$^{-1}$ and 0.09 min$^{-1}$ for wt Vsr and Vsr-Δ14 respectively (94). These data indicate an activity for the Vsr-Δ14 mutant of approximately 3%, as compared with the wt Vsr. Therefore, early kinetic studies on Vsr (89;90) were likely conducted on the truncated form of Vsr and might not accurately reflect behavior in vivo.

Given that the VSPR and MMR pathways are balanced by the availability of MutL, a mechanism by which they avoid competition with each other is imperative. As mentioned earlier, the level of Vsr in the cell is dependent on the phase of growth. Vsr was found to be present in the cell at low levels during exponential phase and at elevated levels as the cells entered stationary phase (69). This trend is consistent with the notion that VSPR utilizes the MutL protein after it has carried out its duties in MMR.

MMR is active just after replication, during the time the newly synthesized strand is unmethylated at its d(GATC) sites (133;134), strand discrimination can occur (152), and MutH is capable of initiating a repair event by incising the DNA (153). Once Dam has fully methylated the d(GATC) sites in DNA, MutH cannot initiate MMR (162), and therefore relieves MutS and MutL from their commitment to that repair pathway.

Accordingly, like Vsr, the level of MutH in the cell is dependent on the phase of cell growth, and has been observed to decrease by 3-fold when cells are in non-replicating conditions (212;213). Interestingly, in stationary cells, the level of MutS is decreased by approximately 10-fold whereas the level of MutL remains constant (212). It therefore appears that the ability of the cell to balance MMR and VSPR relies on regulating the levels of Vsr and MutH so that they do not compete for MutL. This system explains why, during stationary phase, MMR is relatively inefficient (214) and VSPR is most efficient (68). Given that replication errors produce substrates for MMR during the growth phase, and accumulation of deaminated 5mC residues occurs in stationary phase,
it is reasonable that *E. coli* has evolved to balance two distinct repair pathways and maximize the efficient use of the proteins they share in common.

A number of mechanisms for how the level of Vsr is controlled in the cell have been proposed. One posits that the expression of Vsr is regulated by overlap with Dcm. The 5' end of *vsr* overlaps the 3' end of *dcm* by 6 codons in a +1 reading frame (Figure 1.1) (48). The two genes are transcribed from a common promoter, and therefore the ribosome must frameshift to produce two proteins from the same transcription product. The shift in translational reading frames has been studied and found to be approximately 4-fold more efficient as cells enter stationary phase (215). The increase of Vsr in cells as enter stationary phase (69) might therefore be regulated by the cells efficiency at translational frameshifting.

Another theory for how the level of Vsr might be controlled in the cell came from *in vivo* VSPR assays in *dam* strains (70). Dcm-induced CCAGG to CTAGG mutations were found to be increased in *dam* strains more than in *mutS* strains. The mutation level was dependent on the concentration of Dcm methylase, and therefore resulted from a deficiency in VSPR, not MMR. Interestingly, western analysis revealed that the regulation of Vsr in *dam* strains was found to take place post-transcriptionally. While the level of *vsr* transcript was the same in wt and *dam* strains, the absolute amount of Vsr in the *dam* strains was reduced. Therefore, though the Dam methylase does not affect the growth phase-dependent level of Vsr, it has a significant effect in maintaining the absolute level of Vsr in the cell.

1.7. MECHANISM OF VSPR

The mechanism by which MutS and MutL stimulate Vsr is not fully understood. The observations (as discussed above) that a slow conformational change of the Vsr NTD occurs upon DNA binding (211), that ordering of the NTD is essential for forming a stable protein DNA complex (91), and that MutL can stimulate Vsr directly (89), suggest that MutL activates Vsr by ordering its NTD. It has also been hypothesized that MutS
and MutL assist MMR by presenting mismatches within a α-shaped loop (177). As a result, the location of the mismatch is more defined. Also, depending on the size of the loop, it could have the potential to increase negative supercoiling of the DNA and cause unwinding, and thus more exposure, to the mismatch site.

Altogether a likely explanation for how MutS and MutL stimulate Vsr is as follows. Initial recognition of the mismatch is carried out by MutS, which then recruits MutL to assist in presenting the mismatch in a looped DNA structure. Vsr is then recruited to the complex because fully methylated d(GATC) sites render MutH incapable of DNA incision. Contacts are made between MutL and Vsr which activate Vsr by ordering its NTD so that it forms a stable complex with its DNA substrate. Vsr then identifies its substrate DNA mismatch located within the loop and incises the T containing strand to initiate VSPR.

1.8. REPAIR OF T:G MISMATCHES IN EUKARYOTES

At least two proteins exist for the specific repair of T:G mismatches in eukaryotes. In 1993, Nedderman and Jiricny isolated a human enzyme that excises T from T:G mismatches (216). The repair protein was found to be a glycosylase (TDG) that also initiates BER on uracil sites in DNA (217). However, unlike previous uracil glycosylases (Family 1) (218), TDG only excises uracil when it was opposite a guanine. Uracils opposite adenine or in single-stranded DNA are not excised by this enzyme. A structural homolog to TDG does exist in bacteria; however this enzyme, Mug, only acts efficiently on uracil and does not share sequence homology with TDG (219). Multiple reports are consistent in the conclusion that Mug does not excise T opposite G (220-222). The second protein present in eukaryotes for the repair of T:G mismatches is the 5-methylcytosine-binding domain glycosylase 4 (MBD4) (223). This protein also initiates BER by recognizing and hydrolytically removing T opposite G in the context of a methylated CpG site.
Figure 1.1. Organization of the \textit{dcm} and \textit{vsr} genes. The 5’ end of \textit{vsr} overlaps the 3’ end of \textit{dcm} by 7 codons in a +1 reading frame. This genetic arrangement was hypothesized to assure that Vsr is always produced along with Dcm to minimize the mutagenic effects of cytosine methylation. Although the two genes are transcribed from a common promoter, two separate proteins are produced.
Figure 1.2. Repair of mismatches that arise through methylation and subsequent deamination of cytosine. Dcm methylates the C5 of the pyrimidine ring of the underlined C in the sequence 5'- CC(A/T)GG -3'. Subsequent deamination of that residue gives rise to a T:G mismatch, which is a substrate for very short patch repair. Vsr initiates repair by producing a single strand nick 5' to the mismatched T. This activity is stimulated by, but does not require, MutS and MutL. Pol I carries out nick translation by removing and replacing the mismatched T and a small number of nucleotides downstream of the mismatch. DNA Ligase seals the nick to complete the repair event.
Figure 1.3. Unmethylated and hemimethylated Vsr substrates

Unmethylated Vsr Substrate

Hemimethylated Vsr Substrate
Figure 1.4. Proposed mechanisms for the hydrolytic deamination of cytidine. One mechanism involves direct attack at the 4-position of the pyrimidine ring of cytidine (1) by a hydroxyl ion followed by loss of ammonia to yield uridine (4). The other mechanism involves an addition-elimination reaction where water is added to the 5,6 double bond of cytidine (1) to form a dihydrocytidine intermediate (2). This intermediate is then further attacked by water to release ammonia and yield uridine hydrate (5), which then dehydrates to uridine (4). In DNA, R symbolizes the deoxyribose-phosphate backbone.
Figure 1.5. Chemistry behind C:G to T:A transition mutations. Modification of cytosine (A) can occur via deamination and methylation. Deamination of cytosine gives rise to uracil (B), and methylation of cytosine produces 5-methylcytosine (C). Subsequent deamination of 5-methylcytosine produces thymine (D). Replication past either uracil or thymine will give rise to a C:G to T:A mutation.
Figure 1.6. Model for how VSPR could be shaping the *E. coli* genome. If a replication error occurred such that a guanine was misincorporated opposite thymine, the resulting mismatch could be a substrate for both MMR and VSPR. While MMR would repair the error correctly in favor of thymine, VSPR would repair the mismatch in favor of guanine. Over time, this process could lead to the underrepresentation of 5’-CTAGG-3’ and over-representation of 5’-CCAGG-3’ sequences.
Figure 1.7. Structure of Vsr bound to product DNA, minor groove side. The N-terminus of Vsr forms two α-helices that pin down the DNA. The main core of Vsr binds the DNA from the major groove side, where the chemistry takes place. The two magnesium ions are bound by residues from the β-sheet and assist DNA binding through multiple water mediated contacts.
Figure 1.8. Structure of Vsr bound to product DNA, major groove side. Vsr has a three layered $\alpha/\beta/\alpha$ fold that is stabilized by a structural zinc. Vsr also has two magnesium ions that participate in water clusters to bind DNA. The DNA is significantly unwound and bends off axis by $44^\circ$. Three residues insert themselves into the opening and help orient His-69 (shown in ball and stick) to activate a water for attack of the scissile phosphate.
Figure 1.9. Geometric characteristics of C:G and T:G base pairs. In contrast to C:G base pairs, T:G base pairs are structurally distorted. This geometry pushes the mismatched thymine into the major groove and disrupts base pair stacking within the DNA duplex. Adapted from reference (224).
Figure 1.10. Methyl direction of mismatch repair. Replication can cause mismatches to form. The mismatched base in the newly synthesized unmethylated strand is targeted for repair. MMR excises the mismatched base and carries out resynthesis. The newly synthesized strand is then methylated by Dam.
Figure 1.11. Methyl-directed mismatch repair is bidirectional. The point of incision by MutH can be upstream or downstream of the mismatch. Depending on the orientation of the mismatch with respect to the nick, either 3' exonuclease (Exo I) or 5' exonuclease (Exo VII or RecJ) activity is appropriate for removing the intervening nucleotides. Loading of the helicase in the correct orientation, for subsequent exonuclease action, is accomplished by the MutS-MutL complex. Pol III fills in the repair patch and DNA Ligase seals the nick to complete repair.
Figure 1.12. ATP-hydrolysis dependent translocation model for MMR. MutS binds to the mismatch and recruits MutL to the site. This complex uses ATP to motor bidirectionally and mediate the formation of a loop structure. MutH becomes activated by the MutS-MutL complex and makes an incision on the error-containing strand at the unmethylated d(GATC) site (not shown). DNA helicase II then loads onto the DNA at the incision site and unwinds around the DNA loop.
**Figure 1.13.** Molecular switch model for MMR, part 1. MutS bound to ADP acts as a mismatch sensor. Mismatch binding promotes MutS to exchange ADP with ATP and then diffuse along DNA, independent of hydrolysis. Diffusion of ATP-MutS re-exposes the mismatch, allowing multiple clamps to load and slide. MutL interacts with ATP-MutS clamps and slides with them as part of a ATP-MutS-MutL complex, independent of hydrolysis. Binding of MutL to MutS promotes unloading of the clamp and will therefore abort sliding in the absence of key MMR components.
Figure 1.14. Molecular switch model for MMR, part 2. MutL physically connects the ATP-MutS-MutL complex to MutH at the unmethylated d(GATC) site. MutH increases the ability of MutL to bind ATP, which in turn activates the endonucleolytic activity of MutH. In this regard, MutL is a protein-activated molecular switch.
Figure 1.15. Molecular switch model for MMR, part 3. DNA helicase II is loaded onto the DNA and initiates unwinding in a manner which requires ATP and is stimulated by MutS plus MutL. Theories posit that the helicase unwinds in the appropriate direction because MutL either assists in its loading or it orients the helicase correctly by interacting with it from the direction of the mismatch. Upon unwinding, the DNA is degraded by an exonuclease and resynthesized by Pol III.
### Table 1.1. *E. coli* and *S. cerevisiae* proteins required for MMR. Adapted from Schofield and Hsieh (200).

<table>
<thead>
<tr>
<th><strong>E. coli</strong></th>
<th><strong>S. cerevisiae</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>MutS - Binds to mismatches and small IDLs</td>
<td>Msh2/Msh6 - Binds to mismatches and 1-base IDLs</td>
</tr>
<tr>
<td>MutL - Molecular matchmaker; interacts with MutS in a mismatch-specific manner to activate MutH cleavage activity</td>
<td>Mlh1/Pms1 - Primary MutL homolog for postreplication repair</td>
</tr>
<tr>
<td>MutH - Nicks unmethylated strand at a hemimethylated d(GATC) site, initiating repair</td>
<td>Mlh1/Mlh3 - Involved in repair of some IDLs; also functions in meiosis</td>
</tr>
<tr>
<td>β-clamp - Processivity clamp for Pol III, interacts with MutS in vitro; may recruit MutS to the replication fork</td>
<td>Mlh1/Mlh2 - Minor role in suppression of frameshifts</td>
</tr>
<tr>
<td>Helicase II - (UvrD) delivered to nick by MutS and MutL; unwinds DNA prior to its excision</td>
<td>No known homolog</td>
</tr>
<tr>
<td>RecJ, Exo VII - Required for 3' to 5' Excision between nick and mismatch</td>
<td>PCNA - Processivity clamp for Pol δ and Pol ε; interacts with Msh3 and Msh6, increases mispair binding specificity and also involved in repair resynthesis</td>
</tr>
<tr>
<td>Exo I, Exo X - Required for 5' to 3' Excision between nick and mismatch</td>
<td>No known homolog</td>
</tr>
<tr>
<td>Pol III - DNA polymerase required For repair resynthesis</td>
<td>Exo I - 5' to 3' exonuclease, deletion of which results in mild mutator phenotype</td>
</tr>
<tr>
<td>SSB - Single-strand-DNA-binding protein; aids excision and resynthesis</td>
<td>Polδ - DNA polymerase required for repair resynthesis</td>
</tr>
<tr>
<td>DNA Ligase – Seals nicks</td>
<td>RPA – Single-strand-DNA-binding protein; aids resynthesis</td>
</tr>
</tbody>
</table>

DNA Ligase – Seals nicks
1.9. REFERENCES


72


73


Chapter I


- Chapter 1 -

dependence of formation and dynamics of hMutSalpha.hMutLalpha.heteroduplex

complex hMSH2-hMSH6 functions as a novel molecular switch. *Cell.* 91(7), 995-
1005.

recognition and ATPase activity by hMSH2-hMSH3. *J. Biol. Chem.* 274(31),
21659-21664.

(198) Gradia, S., Subramanian, D., Wilson, T., Acharya, S., Makhov, A., Griffith, J.,
and Fishel, R. (1999) hMSH2-hMSH6 forms a hydrolysis-independent sliding

(199) Heinen, C. D., Wilson, T., Mazurek, A., Berardini, M., Butz, C., and Fishel, R.
(2002) HNPCC mutations in hMSH2 result in reduced hMSH2-hMSH6 molecular


Arabidopsis thaliana gene that predicts a protein belonging to the MSH2

characterization of mismatch-containing circular DNA molecules competent for
30(3), E14.

(203) Wang, H., and Hays, J. B. (2004) Signaling from DNA mispairs to mismatch-
repair excision sites despite intervening blockades. *EMBO J.* 23(10), 2126-2133.

repair initiation factors and the replication processivity factor PCNA. *J. Mol. Biol.*
355(2), 175-184.


some dominant-negative mutS mutations that disable methyladenine-directed
mismatch repair are active in very-short-patch repair. *J. Bacteriol.* 183(21), 6487-
6490.


85


CHAPTER 2
Reconstitution of the *E. coli* Very Short Patch Repair System with Purified Proteins
2.1. ABSTRACT

The *E. coli* Very Short Patch Repair (VSPR) system repairs T:G mismatches that arise from the deamination of 5-methylcytosines (5mC) created by Dcm. For reasons yet to be understood, the Dcm protein methylates the second cytosine of 5'- CCAGG -3' sequences, and to the detriment of the cell these 5mC residues can spontaneously deaminate and give rise to T:G mismatches. To avoid the accumulation of mutations and to maintain the sequence integrity of Dcm sites, VSPR removes the mismatched T and restores the native DNA sequence. Repair is initiated by the Vsr endonuclease, which produces a single strand nick 5' to the mismatched T. Studies done on cells lacking the genes for either Pol I or DNA Ligase show diminished levels of VSPR and therefore these proteins have been implicated in the repair process. However, the steps downstream of incision have not been studied biochemically. Herein, we report a novel *in vitro* assay for monitoring the entire VSPR pathway. Using this assay, we demonstrate repair of T:G mismatches contained within short oligodeoxynucleotides by the combination of Vsr, Pol I, and DNA Ligase proteins. These studies represent the first reconstitution of the VSPR pathway and lay the foundation for studying the mechanism of repair more closely.
In *E. coli*, the second cytosine of each strand of the palindromic sequence 5'-'CCWGG' -3' (where W is A or T) is methylated at the C5 position of cytosine by the Dcm methylase to create a fully methylated duplex (Figure 2.1) (1;2). The precise function of methylation is not understood, but it has been hypothesized to play a role in defense against restriction modification systems of invading organisms (3). To the detriment of the cell, 5-methylcytosine (5mC) residues created by Dcm can spontaneously deaminate in a process analogous to that for cytosine (4-6). Deamination of 5mC in these sequences gives rise to T:G mismatches in hemimethylated Dcm sites that are known mutagenic hotspots (7;8). If removal of the mismatch does not precede replication, a C → T transition mutation, and loss of the Dcm methylation site, will result. Mutation avoidance and maintenance of these sites is the responsibility of the very short patch repair (VSPR) pathway (9).

VSPR of mismatched thymine in the sequence 5'-CTAGG -3' (where the T is mismatched with G) is initiated by Vsr (10). This protein is a strand and sequence specific endonuclease that creates a single strand nick on the 5' side of the mismatched T (11). Recombination assays that measure VSPR *in vivo* have implicated the 5' to 3' exonuclease activity of Pol I in removal of the mismatched T in addition to a few bases downstream of the mismatch, and the polymerase activity of Pol I to conduct resynthesis of the error containing strand (12). Therefore, the model for VSPR hypothetically involves nicking by Vsr, nick translation by Pol I, and ligation by the DNA Ligase. However, no studies to date have explicitly tested the ability of these three enzymes to carry out VSPR *in vitro*.

We describe here an assay to monitor the complete VSPR pathway *in vitro*. Moreover, we report that the combination of purified Vsr, Pol I, and DNA Ligase protein from *E. coli* are necessary and sufficient to repair T:G mismatches to C:G base pairs contained within a short oligodeoxynucleotide (ODN). Consistent with the literature, we observed that the nicking efficiency of Vsr is greater on hemimethylated duplexes as
compared to their unmethylated counterparts (Figure 2.1) \((13;14)\). Interestingly, this preference of Vsr affected both the efficiency of G:T to G:C repair and the extent to which the repair reactions went to completion. We also observed that the efficiency of Vsr was increased in the presence of Pol I and DNA Ligase proteins. Our results therefore support a model for VSPR involving multiple protein-protein interactions for maximal efficiency.
2.3. EXPERIMENTAL PROCEDURES

Cloning, Expression, and Purification of Vsr. Preparation of *E. coli* genomic DNA for *vsr* amplification was carried out essentially as described in Current Protocols (15). Briefly, 1 mL of cells from an overnight culture was spun down and the resulting cell pellet was resuspended in 570 μL TE 8.0 (10 mM Tris, 1 mM EDTA). After the addition of 30 μL of 10% SDS and 1.8 units of Proteinase K (Roche), the solution was incubated at 37 °C for 1 h. This solution was then supplemented with 100 μL of 5 M NaCl, 80 μL of a 10% CTAC/0.7 M NaCl solution, and incubated at 65 °C for 10 minutes. This solution was extracted twice from 750 μL of a 25:24:1 ratio of phenol to chloroform to isoamyl alcohol (Invitrogen). DNA was precipitated from the final aqueous layer with 420 μL of Isopropanol, and the DNA pellet was washed with 250 μL of 70% Ethanol. The resulting DNA was dried *in vacuo* for 5 minutes, resuspended in 100 μL TE, and quantified using UV spectrometry.

The gene sequence for Vsr was amplified from GM30 genomic DNA by PCR using the primers ODN 1 and ODN 2 (Table 2.1). PCR was carried out according to the Opti-prime PCR Optimization Kit (Stratagene). The combination of Taq polymerase (NEB), buffering conditions #6 or #7, and 2% DMSO as a reaction adjunct provided the best results. The thermocycling parameters were as follows: 94 °C for 120 seconds, then thirty rounds of 95 °C for 30 seconds, 54 °C for 45 seconds, and 72 °C for 60 seconds, and then a final incubation at 72 °C for ten minutes. The 471 bp product was identified by separating a fraction of the reaction products on a 1% agarose in TAE gel and purified using the QIAquick PCR Purification kit (QIAGEN). Using the *EcoRI* and *HindIII* restriction sites contained within the primers, the amplified segment was inserted into complementary sites of pET28a (Novagen) using T4 DNA ligase (NEB). Following transformation into competent *E. coli* DH5α, DNA from isolated clones were characterized by restriction analyses and sequencing. One successful recombinant expression vector was then transformed into BL21-Gold(DE3)pLysS cells (Stratagene). Expression of the expected 18 kDa protein containing Vsr (aa 1 to 156) fused to a N-
terminal 6 x His tag was confirmed by Western blot analysis using anti-His tag antibodies (Amersham).

To prepare cell pellets containing over-expressed Vsr, an overnight culture, grown in LB containing 25 µg/mL Chloroamphenicol (Cam) and 50 µg/mL Kanamycin (Kan), was diluted (100 µL to 10mL) in LB and grown (roller drum, 37 °C) to mid log phase (OD<sub>600</sub>~0.65). Expression of Vsr was induced by adding IPTG (Roche) to a final concentration of 0.75 mM and the cells were grown, as above, for an additional 3 h. Cells were then harvested by centrifugation (6000 xg, 15 min) and stored at -20 °C for up to one week.

When purified Vsr was desired, cell pellets were lysed using BugBuster® protein extraction reagent (Novagen). Briefly, 500 µL of lysis buffer (containing 400 µL of wash buffer (25mM HEPES, 300mM NaCl, 10mM MgCl₂, 2mM βME, pH 8.0), 50 µL of 10x BugBuster®, 50 µL of SIGMA P8340 protease inhibitors cocktail, and 15 units benzonase) was used to resuspend the pellet and transfer it to a 1.5 mL tube. Disruption of the cell membrane proceeded by nutating the mixture for 15 min at RT.

The cell lysate was isolated from insoluble material by centrifugation (13,000 xg for 15 min, at 4 °C) and transferred to a QIAGEN Ni-NTA Spin Column pre-equilibrated by loading the column with 600 µL wash buffer and centrifuging at 2,000 xg for 2 min at 4 °C. The Vsr protein mixture was passed through the column by centrifugation as above, washed five times with 600 µL wash buffer, and purified Vsr was eluted from the column by passing two fractions of elution buffer (wash buffer plus 250 mM imidazole, 100 µL each) through the column. The Vsr eluates (~0.15 mg/mL) were combined and kept on ice until subsequent reactions.

Nicking assays. To prepare substrates for nicking assays, 1 nmole of 38-mer ODN 3 was 5'-end labeled with 16.5 pmole [γ-³²P]ATP (6000 Ci/mmol)(NEN brand, Perken Elmer) using 10 units of T4 polynucleotide kinase (T4PNK)(NEB) in the supplier’s reaction buffer. The 20 µL reaction proceeded for 30 min at 37 °C and was then heated to 65 °C.
for 30 min to inactivate the reaction. Fractions (2 μL, 100 pmoles) were annealed in 50 mM NaCl with equimolar amounts of complementary strands ODN 4, ODN 5, and ODN 6 to construct duplexes containing unmethylated T:G substrate, hemimethylated T:G substrate, and unmethylated T:A control duplex respectively (Figure 2.1). These duplexes (1 μM) were stored at -20 °C for up to 2 weeks.

Nicking assays were performed by combining 250 fmoles substrate, 1.25 μL salmon sperm competitor DNA (Stratagene), 10 μL Vsr protein (~1.5 μg), and reaction buffer (25 mM HEPES, 100 mM NaCl, 10mM MgCl₂, 2mM βME, pH 7.5) to a final volume of 20 μL. Reactions were incubated at 37 °C (for 0, 15, 30, 60, 120, and 180 min) and stopped by adding an equal volume of formamide plus 10mM EDTA loading buffer and putting them at -20 °C. DNA fragments were separated via 20% denaturing PAGE, visualized by PhosphorImagery (Storm 840, Molecular Dynamics), and quantified using ImageQuant 5.2 software (Molecular Dynamics).

Reconstitution of VSPR with Purified E. coli Proteins. To prepare substrates for reconstitution reactions, 500 pmoles of 28-mer ODN 7 was 5’ phosphorylated using 6.6 pmoles [γ-32P]ATP (6000 Ci/mmol)(NEN brand, Perken Elmer) and 15 units of T4PNK for 10 min in a total reaction volume of 20 μL, buffered according to the supplier. After incubation at 37 °C for 30 min, the reaction was supplemented with excess (300 nmoles) non-radiolabeled ATP, incubated at 37 °C for an additional 2.5 h, and then heated to 65 °C for 30 min to inactivate the reaction. The completely 5’ phosphorylated 28-mer ODN 7 was then ligated to an equimolar amount of 17-mer ODN 8 (500 pmoles), using the scaffold ODN 9 (500 pmoles), and 1200 units of T4 DNA Ligase (NEB). The 60 μL reaction was buffered according to the enzyme supplier and proceeded at 16 °C overnight. A fraction (6 μL, 50 pmoles) of the resulting 45-mer ODN, containing an internal radiolabel, was then mixed with an equimolar amount of 45-mer ODN 10, and an equimolar amount of 45-mer ODN 11, in separate tubes. The two DNA mixtures (1000 μL in Vsr reaction buffer) were heated to 75 °C and cooled to room temperature slowly, allowing the unmethylated T:G duplex and hemimethylated T:G duplex to anneal.
Reconstitution reactions were performed by combining 50 fmoles of substrate, 1 µL salmon sperm competitor DNA (Stratagene), 15 µL Vsr (2.25 µg), 5 units *E. coli* DNA Polymerase I (NEB), and 5 units *E. coli* DNA Ligase (NEB) in reconstitution buffer (25 mM HEPES, 100 mM NaCl, 10mM MgCl₂, 2mM βME, pH 7.5). Each reaction was also supplemented to contain 2 mM ATP (Roche), 100 µM β–NAD (free acid, grade I, Roche), 100 µM each dNTP (NEB), 50 µg Creatine Kinase (Roche), and 50 mM Creatine Phosphate (Roche). Reactions (50 µL total) were incubated at 37 °C for 1 h and were stopped by adding an equal volume of formamide plus 10mM EDTA loading buffer. In addition to performing VSPR reaction in triplicate for each duplex, parallel control reactions were performed in the presence of all reactants except Pol I and DNA Ligase proteins to assess Vsr nicking.

**Analysis of Reconstitution.** After reconstitution reaction, the substrate product mix was separated from the other reaction components via PAGE (20% in TBE, 7 M Urea). The 45-mer substrate product mix was identified by its position on the gel and its location relative to the other bands. The appropriate cuts were made to excise the 45-mer and the gel slice was put into a 1.5 mL screw top tube, crushed thoroughly with 2 wood applicators, resuspended in 750 µL ddH₂O, vortexed for 20 seconds, and put on a rotator overnight at room temperature for the DNA to elute from the gel bits.

The next day, the DNA was isolated from the gel bits as follows. The top of the tube was removed and the center of the bottom was punctured using a 21 guage needle. This tube was then put on top of a 1.5 mL flip top tube which had its top cut off, its bottom punctured with a 26 guage needle, and contained 0.2 g of sand. This tube was, in turn, put on top of a collection tube consisting of a 1.5 ml flip top tube with its top cut off. This three tiered set up was put into a 15 mL centrifugation tube for stability and centrifuged at 8500 xg for 1.5 hours. The DNA eluate was transferred from the collection tube into a sterile 1.5 mL tube by pipet. The middle tube containing the sand and the trapped gel bits was disposed of appropriately as a radioactive material.
The resulting DNA solution was concentrated to dryness in vacuo and resuspended in 40 µL ddH₂O. The final volume of each sample was brought to 100 µL with ddH₂O. Desalting was accomplished via two G-25 Sephadex Quick Spin Columns (Roche). The substrate product mixes were then suitable for downstream enzymatic steps.

Digests were conducted on the 60 µL scale by adding 5 µL 10X NEB buffer #2, 1 µL BsmI (10 units), 1 µL BstYI (10 units), and 3 µL ddH₂O to each 50 µL of DNA. The reactions were allowed to proceed at 60°C for 2 h. Then, 2 µL of phosphatase (Roche, 1 units/µL) was added to the digests and incubated at 37°C for an additional 2 h. The digest/phosphatase reactions were terminated by heating to 80°C for 15 minutes.

The resulting DNA fragments in each mixture were radiolabeled as follows. Each 60 µL mixture was supplemented with 2 µL of a mix containing 0.5 µM [γ-32P]-ATP (NEN brand, Perkin Elmer), 0.75 µM ATP (Roche), 2.5 units/µL T4PNK (NEB), and 1 X T4PNK buffer (supplied by NEB). The kinase reactions were incubated at 37 °C for 45 minutes to ensure that every DNA fragment 5' end was phosphorylated.

The mixtures containing radiolabeled DNA fragments were separated via PAGE (20%, 7 M Urea). The target 10-mer ODN was located via PhosphorImagery, excised from the gel, and isolated from the gel slice as described above. Once completely dried, the DNA was resuspended in a total of 50 µL of 20 mM sodium acetate pH 5.3, and desalted using Micro Bio-Spin Chromatography Columns (BioRad) according to protocol. Before applying the sample to the columns, however, the resin was equilibrated with 500 µL of 20 mM sodium acetate pH 5.3 two times. The resulting DNA sample was therefore buffered in sodium acetate and ready for digestion into mononucleotides.

To each DNA sample, 2 µL of nuclease P1 (United States Biological, resuspended with ddH₂O to 0.3 units/µL) was added. Digestion was allowed to take place at 65°C for 2 h. Afterwards, the mixtures of mononucleotides were concentrated to ~10 µL in vacuo and ready for TLC separation.
TLC was conducted using cellulose 20 cm x 20 cm Baker-flex PEI-F sheets (J.T. Baker, Phillipsburg NJ), which were previously washed in ddH$_2$O and allowed to dry completely. Approximately 5 µL of mixture was spotted and separation of the mononucleotides proceeded for 12 h using saturated ammonium phosphate pH 5.8 as the mobile phase. The TLC plate then dried on the bench for ~1 h., was protected with Saran Wrap, and visualized by PhosphorImagery (Storm 840, Molecular Dynamics). The relative amounts of thymine and cytosine mononucleotides were calculated using ImageQuant 5.2 software (Molecular Dynamics).

To account for any contamination that survived the Vsr protein preparation and contributed to VSPR, the 45-mer DNA from control reconstitution reactions lacking Pol I and DNA Ligase was isolated and treated in parallel. Given that these reactions were not supplemented with Pol I and DNA Ligase, isolated 45-mer DNA was theoretically 100% substrate and representative of 100% T at the interrogation site. The %C found for these "control reactions" could therefore account for the total background in the assay and was subtracted from the sample values.
Analysis of nicking activity. The ability of Vsr to incise substrates that contained T:G mismatches in the canonical sequence 5'- CTAGG -3' (where the T is mismatched with G) was measured using a nicking assay. Incubation of purified Vsr with duplex substrates that contained a mismatch afforded incision 5' to the mismatched T and converted the radiolabeled 38-mer T containing strand into a radiolabeled 17-mer and an unlabeled 21-mer. Triplicates of identical reactions were stopped over a range of incubation times and the substrate to product ratio was determined by separating the reaction mixture by PAGE and visualizing the distribution of 38 and 17-mers by PhosphorImagery. A control duplex, in which the radiolabeled T was correctly paired with A in duplex DNA, was analyzed in parallel.

In total, three DNA duplexes were analyzed by the nicking assay: an unmethylated duplex, a hemimethylated duplex, and a control duplex (Figure 2.1). Gel analysis of the replicates (Figure 2.2) shows that there was a good degree of reproducibility of nicking extent for each substrate. The control duplex showed no fragmentation over time. Additionally, gel analysis of the control duplex reactions revealed that the oligo experienced essentially no degradation by exonucleases. The observation that the substrates and products are well defined bands reflects the propensity of phosphorothioate bonds in deterring exonuclease degradation, which manifests in a laddering effect on the gel. In the presence of normal phosphodiester bonds, oligonucleotide substrates were readily degraded, primarily from their 3' ends (data not shown).

Both the unmethylated substrate and the hemimethylated substrates showed significant fragmentation of the 38-mer substrate to the 17-mer product over time (Figure 2.2). Quantification of the amount of nicked substrate was carried out using ImageQuant 5.2 software and revealed that Vsr was more active on the hemimethylated substrate as compared to the unmethylated substrate. Depending on the time of incubation, the preference of Vsr for hemimethylated substrate was between 2- and 2.5-fold (Figure 2.3).
The ability of Vsr to cleave a hemimethylated substrate more efficiently than an unmethylated substrate was also observed when the duplexes were incubated with Vsr under reaction conditions that were supplemented for VSPR reconstitution (Figure 2.4). Again the replicates for each duplex showed good correlation with each other, and the preference of Vsr for hemimethylated substrate was approximately 3-fold. After the time allotted, 19 ± 5% of unmethylated substrate was nicked and 63 ± 2% of hemimethylated substrate was nicked (Figure 2.5).

There was no signal from nicked product in the VSPR reconstitution reactions that were supplemented with Vsr, Pol I, and DNA Ligase (Figure 2.4). These results are consistent with the notion that Pol I and DNA Ligase were capable of the repair events downstream of Vsr nicking. According to the model for the mechanism of VSPR, action by Pol I would lead to the removal and replacement of the radiolabeled T mononucleotide and a small number of additional nucleotides downstream. Upon ligation of the remaining nick, a 45-mer product would be formed that lacked a radiolabel and was the same size as the 45-mer substrate. The absence of nicked product and the loss of radioactivity signal from the 45-mer are both consistent with this model.

The signal loss in the reactions containing the hemimethylated substrate (72 ± 3%) was more than those for the unmethylated substrate (55 ± 4%) (Figure 2.5), and was consistent with the notion that nicking and subsequent repair was more efficient on the hemimethylated duplex. It is interesting to note that the amount of unmethylated and hemimethylated substrate converted to the nicked species was less in the control reconstitution reactions lacking Pol I and DNA Ligase (19 ± 5% and 63 ± 2%) as compared to the reactions that contained these enzymes (55 ± 4% and 72 ± 3%) (Figure 2.5).

After reconstitution of VSPR, 45-mer DNA was separated from the reaction mixture by denaturing PAGE, isolated from the gel, and purified (Figure 2.6). The 45-mer VSPR substrate/product mixture was digested into fragments that were then
radiolabeled. The 10-mer DNA fragment that contained the radiolabel at the interrogation site was separated from the reaction mixture by denaturing PAGE and isolated from the gel. The target oligonucleotides, containing a radiolabeled T or C at the interrogation site, were digested to mononucleotides, and separated by TLC. The resulting TLC plate was analyzed by PhosphorImagery (Figure 2.7) to quantify percentage cytosine. Some enzymatic degradation of the oligo mixture was evident as PhosphorImagery revealed DNA fragmentation. Therefore, alongside the target 10-mer oligo containing a T or C at the interrogation site, a small amount of undesired 10-mer was isolated.

The resulting values for %C at the interrogation site were 21, 27, and 23 for the unmethylated substrate. The values for the hemimethylated substrate were 54, 50, and 4. This third value was analyzed and rejected using the Q test. The $Q_{crit}$ value for three observations at 90% confidence is 0.9 and the $Q_{calc}$ for the data point was 0.92. Therefore, the average and standard deviation of three observations is used for the unmethylated substrate and two observations for the hemimethylated substrate. The % C at the interrogation site for the unmethylated and hemimethylated substrate was 24 ± 3 and 52 ± 3 respectively.

For the reconstitution reactions containing all the enzymes, the number of repair events initiated that went to completion can be calculated using the %C at the interrogation site and the radiosignal from the 45-mer DNA after reaction. In the case of unmethylated duplex, total 45-mer DNA (radiolabeled and non-radiolabeled) isolated from the gel contained $0.45 ± 0.04$ parts radiolabeled T from the unreacted substrate and some number of additional parts non-radiolabeled C from VSPR. The number of parts of repaired C was calculated to be $0.14 ± 0.02$, as this value over the total ($0.45 ± 0.04 + 0.14 ± 0.02$) is equal to the portion determined by TLC analysis ($0.24 ± 0.03$). Therefore, 14 ± 2% of total unmethylated substrate was repaired. This value corresponds to 25 ± 4% of the repair events that were initiated ($55 ± 4$%)(Figure 2.8). The value for the hemimethylated duplex repaired was derived, as above, to be 30 ± 4%. This value
corresponds to $42 \pm 6\%$ of the repair events that were initiated ($72 \pm 3\%$)(Figure 2.8). These data are organized in Table 2.2.
2.5. DISCUSSION

Repair of T:G mismatches that arise in the sequence 5'-CTAGG-3' from the methylation and deamination of cytosine are repaired in *E. coli* by the very short patch repair (VSPR) pathway (16). Initiation of repair is carried out by the Vsr endonuclease which nicks the T containing strand 5' to the mismatch. Studies have previously established that the activity of Vsr for hemimethylated substrates (14) is approximately 2.5-fold better than that for unmethylated substrates (13). Presence of the C5-methyl group on cytosine in the hemimethylated substrate has been reported to correlate both a decrease in the dissociation constant (K_D) and an increase in the single turnover rate constant (k_{at}) (17). Analysis of available crystallographic data for Vsr bound to duplex DNA (18;19) suggests that the C5-methyl group packs against an aliphatic surface of the protein (Phe-77, Lys-78, Val-79), and assists binding at the protein DNA interface (20).

Consistent with these observations, our purified Vsr protein showed a nicking efficiency of approximately 2- to 3-fold higher for the hemimethylated substrate as compared to the unmethylated one. This preference was observed in both the nicking assay conditions (Figure 2.3) and in the control reconstitution reactions lacking Pol I and DNA Ligase protein (Figure 2.5). Gel separation of the reaction products from the nicking assays also shows that, in line with previous literature, strand incision by Vsr is carried out on the T containing strand and is specific to a T:G mismatch (Figure 2.2).

Subsequent to nicking at the mismatch site, the mismatch must be removed and replaced to complete the repair event. The model for the steps subsequent to Vsr nicking was elucidated by data gathered from *in vivo* VSPR assays in cells of a variety of genetic backgrounds. Studies done by Dzidic and Radman established that the *polA* gene fragments associated with the 5' to 3' exonuclease activity of Pol I and the polymerase activity of Pol I were both necessary for efficient VSPR (12). These results were consistent with the notion that Pol I removes the mismatched T and conducts strand resynthesis. This same body of work investigated the level of VSPR in strains lacking *xth, nth,* and *nfo* genes and revealed that these nucleases did not affect VSPR. Lieb
established that strains which lack the functions of the MutS and MutL recognition proteins show reduced, but not eliminated, levels of VSPR (21). Therefore, these proteins are thought to assist Vsr in mismatch recognition but are not necessary for the endonuclease activity. In line with this observation, the decrease in VSPR in *mutS* or *mutL* strains can be counteracted by complementing the cell with additional Vsr from a plasmid (10).

Additional evidence that Vsr does not require the MutS and MutL proteins has come from biochemical experiments. While it has been hypothesized that MutL assists Vsr by orienting its N-terminus to a conformation that stabilizes DNA binding (22), standard investigation of the incision event by Vsr is conducted without any additional enzymes (23). We therefore felt comfortable in the ability of Vsr to initiate VSPR, and did not include the MutS and MutL proteins in the reconstitution assay mixtures.

Reconstitution reactions in the presence of Vsr, Pol I, and DNA Ligase created products that were consistent with reconstitution of the VSPR pathway (Figure 2.4). The presence of one radioactive band, as opposed to the two bands seen when Pol I and DNA Ligase were not included, suggested that the product of Vsr incision was processed further by the Pol I and DNA Ligase enzymes. Removal of the mismatched T by Pol I releases the mismatched mononucleotide and thereby removes the radiosignal from the nicked product. Subsequent to removing the mismatched T, Pol I carries out strand resynthesis and replaces a short tract of the DNA typical for the enzyme (24). The resulting non-radiolabeled product co-migrated with the substrate in the gel, facilitating the isolation of the substrate/product mixture to be analyzed for the distribution of T and C at the interrogation site.

The method by which the distribution of T and C at the interrogation site was analyzed is outlined in Figure 2.6. The assay involves cleavage of the duplexes into smaller fragments using restriction enzymes. This step also serves to expose the interrogation site for 5’ radiolabeling. The DNA fragments were all of different size allowing the target oligo containing the radiolabeled interrogation site to be separated and
identified by PAGE. After isolation, the target oligodeoxynucleotide mixture containing a radiolabeled T or C was digested to mononucleotides and separated by TLC. Phosphorimager analysis was then used to quantify the %C at the interrogation site. This parameter was used (as described in the experimental procedures section) in conjunction with the radiosignal from the 45-mer after the reaction to calculate the extent of repair.

Both substrates were indeed repaired by the combination of Vsr, Pol I, and DNA Ligase proteins, though with different efficiencies and extent of completion (Table 2.2). The hemimethylated substrate was repaired (30 ± 4%) about twice as well as the unmethylated substrate (14 ± 2%). In light that the hemimethylated substrate was also cleaved more efficiently, it appears as though the nicking step was at least partially rate limiting in the reaction conditions used. Interestingly, the percent completion was also higher for the hemimethylated duplex (42 ± 6%) as compared to the unmethylated duplex (25 ± 4%). These results may indicate that increased affinity of Vsr for the hemimethylated substrate results in protection of the incised DNA until Pol I and DNA Ligase can properly process the nick. As a result, a greater percentage of the events initiated by Vsr on the hemimethylated substrate are reserved for downstream events. Additionally, it could be the case that Pol I identifies nicked sites more efficiency by the presence of Vsr protein bound to the DNA.

Vsr is known to form a stable complex with its DNA product (25). Indeed, Vsr bound to its product has been crystallized (18;19) and observed by band-shift experiments (25). It is therefore very likely that Vsr stays bound to and protects its incised DNA product in vivo until Pol I arrives. In turn, Pol I may protect its product until DNA Ligase can seal the juxtaposing DNA ends. Accordingly, the efficiency of VSPR would depend on the availability of all proteins involved, and would benefit from a coordinated process. Figure 2.8 shows that only a fraction of VSPR events initiated went to completion. Therefore, it is a possibility that additional components are critical for maximal VSPR efficiency. Studies to investigate this question are underway.
It is also noteworthy that the amount of each duplex incised by Vsr was less in the control reconstitution reactions as compared to the reactions containing Pol I and DNA Ligase protein (Figure 2.5). Perhaps, these proteins play a role in functionally assisting the activity of Vsr and increase its incision efficiency. There is a growing body of evidence that supports the notion of multi-component repair complexes. Association of the MutS, Pol I, and DNA Ligase proteins with the β-clamp has recently been reported (26). These findings suggest that certain types of repair may have mismatch recognition and repair that is coordinated. Slight variations of the assay described here may be able to address these possibilities in the context of VSPR.
Figure 2.1. Unmethylated and hemimethylated Vsr substrates

Unmethylated Vsr Substrate

5' - CCAGG - 3'
3' - GGTCC - 5'

Hemimethylated Vsr Substrate

5' - CTAGG - 3'
3' - GGTCC - 5'

Spontaneous Deamination of 5mC

Dcm
Figure 2.2. PhosphorImager results from Vsr nicking assays on unmethylated, hemimethylated, and control substrates.
Figure 2.3. Graphical representation of unmethylated and hemimethylated substrate nicking by Vsr.

<table>
<thead>
<tr>
<th>% Substrate</th>
<th>0 min</th>
<th>15 min</th>
<th>30 min</th>
<th>60 min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmeth.</td>
<td>100 ± 12</td>
<td>85 ± 2</td>
<td>71 ± 12</td>
<td>57 ± 4</td>
</tr>
<tr>
<td>Hemimeth.</td>
<td>100 ± 15</td>
<td>62 ± 13</td>
<td>43 ± 11</td>
<td>19 ± 7</td>
</tr>
</tbody>
</table>
Figure 2.4. PhosphorImagery results from incubating T:G containing duplexes with VSPR proteins. Reaction products were separated by 20% denaturing PAGE and visualized by PhosphorImagery. The first three lanes of each gel are replicates of control reconstitution reactions containing Vsr but lacking Pol I and DNA Ligase. The second set of reactions are replicates of reconstitution reactions containing Vsr, Pol I, and DNA Ligase enzymes.
Figure 2.5. Graphical representation of Vsr nicking results from reconstitution reaction conditions. Signal loss is represented for control reconstitution reactions lacking Pol I and DNA Ligase (gray columns) and reconstitution reactions containing those proteins (black columns).
Figure 2.6. Methodology for assessing VSPR efficiency in vitro

Substrate/Product Mixture

5'-TACGGTAGTGGAATGCXAGGCTTCGAGATCCCATTTGCTTGAAACAT-3'
3'-ATGCCATCACCTTAGGCTCGAAGCTCTCTAGTAAACGAACTTGTA-5'

X = T, C

1. Ssml, BstYI digest
2. Phosphatase
3. Radiolabel (*) 5' ends

20% denaturing PAGE

1. 10 mers isolated and digested
2. Mononucleotides separated by TLC

Standards Rxn 1 Rxn 2

% C = 100 * C / (C + T)
**Figure 2.7.** Representative PhosphorImagery results from TLC separation of mononucleotides at interrogation site. Separation of cytosine (C) and thymine (T) was accomplished with TLC and visualized by PhosphorImagery. Shown are representative results from C and T standards and VSPR reactions on (1) unmethylated substrate, and (2) hemimethylated substrate.
Figure 2.8. Percentage of VSPR events initiated that went to completion

- Chart 2 -
Table 2.1. Oligodeoxynucleotides (ODNs) used for cloning Vsr into pET28a and construction of nicking assay and reconstitution assay substrates.

<table>
<thead>
<tr>
<th>ODN</th>
<th>SEQUENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5'-TTT GAA TTC ATG ATG GCC GAC GTT CAC GAT-3'</td>
</tr>
<tr>
<td>2</td>
<td>5'-GAC AAG CTT TCA AGC GAG TAA ATG AAT CCC-3'</td>
</tr>
<tr>
<td>3</td>
<td>5'-C<em>A</em>C<em>GGTAGATGGA</em>C<em>G</em>CCTAGGTTCATGCTT<em>G</em>A*T-3'</td>
</tr>
<tr>
<td>4</td>
<td>5'-A<em>T</em>C<em>A</em>A<em>GCATAGATGGA</em>C<em>CCTAGGTCATGCTT</em>G<em>A</em>T-3'</td>
</tr>
<tr>
<td>5</td>
<td>5'-A<em>T</em>C<em>A</em>A<em>GCATAGATGGA</em>C<em>CCTAGGTCATGCTT</em>G<em>A</em>T-3'</td>
</tr>
<tr>
<td>6</td>
<td>5'-A<em>T</em>C<em>A</em>A<em>GCATAGATGGA</em>C<em>CCTAGGTCATGCTT</em>G<em>A</em>T-3'</td>
</tr>
<tr>
<td>7</td>
<td>5'-TAGGCTTCGAGATGACGCTTG<em>A</em>A<em>C</em>A*T-3'</td>
</tr>
<tr>
<td>8</td>
<td>5'-T<em>A</em>C<em>G</em>G*TAGTGGAATGCC-3'</td>
</tr>
<tr>
<td>9</td>
<td>5'-GAAGCCTAGGCAATTCC-3'</td>
</tr>
<tr>
<td>10</td>
<td>5'-A<em>T</em>G<em>T</em>T<em>CAAGCAATGGATCTCAAGCCTGGCATTCCACTA</em>C<em>G</em>T*A-3'</td>
</tr>
<tr>
<td>11</td>
<td>5'-A<em>T</em>G<em>T</em>T<em>CAAGCAATGGATCTCAAGCSTGGCATTCCACTA</em>C<em>G</em>T*A-3'</td>
</tr>
</tbody>
</table>

5mC - 5-methylcytosine

* - phosphorothioate linkage
Table 2.2. Quantitative results from reconstitution assays containing Vsr, Pol I, and DNA Ligase protein

<table>
<thead>
<tr>
<th></th>
<th>Unmethylated Substrate</th>
<th>Hemimethylated Substrate</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Substrate Remaining (T at interrogation site)</td>
<td>45 ± 4</td>
<td>28 ± 3</td>
</tr>
<tr>
<td>% Substrate Loss (Repair Initiated)</td>
<td>55 ± 4</td>
<td>72 ± 3</td>
</tr>
<tr>
<td>%C at interrogation site</td>
<td>24 ± 3</td>
<td>52 ± 3</td>
</tr>
<tr>
<td>% of starting material repaired</td>
<td>14 ± 2</td>
<td>30 ± 4</td>
</tr>
<tr>
<td>% of repair initiated that went to completion</td>
<td>25 ± 4</td>
<td>42 ± 6</td>
</tr>
</tbody>
</table>
2.6. REFERENCES


CHAPTER 3
The Size of Repair Tracts Created by *E. coli* Very Short Patch Repair System
3.1 ABSTRACT

The *E. coli* Very Short Patch Repair (VSPR) pathway is responsible for repairing T:G mismatches that arise through the Dcm-mediated methylation and subsequent deamination of the underlined cytosine residue in the sequence 5’-CCAGG -3’. While the purpose of Dcm-mediated methylation is not well understood, the importance of Dcm site maintenance is evident, based on the overlap between the *dcm* and *vsr* genes. The Vsr protein initiates repair of T:G mismatches that arise in Dcm sites by producing a single strand nick 5’ to the mismatched T. Pol I then removes and replaces a short tract of nucleotides downstream of the incision and DNA Ligase seals the nick to complete the repair event. The small size of repair tracts excised is reflected in the repair system’s name. Bacteriophage crosses revealed that a significant number of VSPR events do not include co-repair of markers as close as five bases away from the T:G mismatch. However, closer examination of the repair patch size and distribution was convoluted by monitoring repair of T:G in close proximity to another marker mismatch. To circumvent this issue, we have devised a novel *in vitro* assay to measure the distribution of VSPR patch sizes in whole cell extracts. The assay involves monitoring the loss of radiosignal from a series of substrates that contain the signal at prescribed distances downstream of the T:G mismatch. We found that the repair patch is distributed about 2 to 4 deoxynucleotides in length. Interestingly, under certain reaction conditions, the addition of DNA Ligase actually improved the efficiency of repair initiation, suggesting that VSPR may operate best in the context of a multi-protein complex.
3.2 Introduction

In *E. coli*, the second cytosine of 5'-CCWGG-3' sequences (where W is A or T) is subject to methylation at the C5 position of the pyrimidine ring by the Dcm methylase (1;2). These 5-methylcytosine (5mC) residues are known to deaminate spontaneously, giving rise to T:G mismatches (3;4). If correction of the mismatch does not precede replication, a C→T mutation will result and the Dcm site will be lost. Therefore, a system must exist to avoid mutations and maintain these sites. Maintenance of the 5mC sites is the responsibility of the Very Short Patch Repair (VSPR) system (5).

The VSPR system in *E. coli* is so named because of the short repair tracts it creates (6). Studies done in the early 1980s by Lieb estimated that the repair tract size rarely exceeds 10 deoxynucleotides *in vivo* (7). VSPR of the mismatched thymine is initiated by Vsr. This protein is a strand and sequence specific endonuclease that creates a single strand nick 5' to the mismatched T (8). The activity of Vsr is increased by association with the mismatch recognition proteins MutS and MutL (9-11). Recombination assays used to measure VSPR *in vivo* have implicated the 5' to 3' exonuclease activity of Pol I in removal of the mismatched T, and the polymerase activity of Pol I in resynthesis of the error containing strand, because cells that are deficient in either fragment of the *polA* gene are deficient in VSPR (12). DNA Ligase is reasonably assumed to seal the nick and complete the repair event. While the Vsr protein has been studied biochemically (9;13-16), *in vitro* investigation of the VSPR events downstream of incision is lacking (17).

An advantage to studying systems biochemically is that the concentrations of substrate(s) and protein(s) can be varied incrementally. The ability to easily modulate individual components facilitates observing their relative contributions to the overall event. Additionally, these types of studies can yield insight in to how the proteins cooperate. As a result, biochemical studies can provide a more complete understanding of the system.
Herein, we describe studies investigating the VSPR system in whole cell extracts bearing over-expressed Vsr. We have developed a novel assay for monitoring the repair patch size and distribution created by VSPR, as well as the effect of DNA Ligase on the repair event. The assay involves a series of oligodeoxynucleotides (ODNs) containing radioactive signals in the DNA backbone at prescribed distances downstream of the T:G mismatch. VSPR is monitored through the detection of radiosignal loss from substrate DNA as the $^{32}$P-group is removed as part of the repair patch. Our results indicate that VSPR removes and replaces 2 to 4 deoxynucleotides and may act as a multi-protein complex for optimal efficiency.
3.3 EXPERIMENTAL PROCEDURES

Cloning and Expression of Vsr. Preparation of *E. coli* genomic DNA for *vsr* amplification was carried out essentially as described in Current Protocols (18). Briefly, 1 mL of GM30 cells from an overnight culture was spun down and the resulting cell pellet was resuspended in 570 μL TE 8.0 (10 mM Tris, 1 mM EDTA). After the addition of 30 μL of 10% SDS and 1.8 units of Proteinase K (Roche), the solution was incubated at 37 °C for 1 h. This solution was then supplemented with 100 μL of 5 M NaCl, 80 μL of a 10% CTAC/0.7 M NaCl solution, and incubated at 65 °C for 10 minutes. This solution was extracted twice from 750 μL of a 25:24:1 ratio of phenol:chloroform:isoamyl alcohol (Invitrogen). DNA was precipitated from the final aqueous layer with 420 μL of isopropanol, and the DNA pellet was washed with 250 μL of 70% Ethanol. The resulting DNA was dried *in vacuo* for 5 minutes, resuspended in 100 μL TE, and quantified using UV spectrometry.

The gene sequence for Vsr was amplified from GM30 genomic DNA by PCR using the primers 5’- TTT GAA TTC ATG ATG GCC GAC GTT CAC GAT -3’ and 5’- GAC AAG CTT TCA AGC GAG TAA ATG AAT CCC -3’. PCR was carried out according to the Opti-prime PCR Optimization Kit (Stratagene). The combination of Taq polymerase (NEB), buffering conditions #6 or #7, and 2% DMSO as a reaction adjunct provided the best results. The thermocycling parameters were as follows: 94 °C for 120 seconds, then thirty rounds of 95 °C for 30 seconds, 54 °C for 45 seconds, and 72 °C for 60 seconds, and then a final incubation at 72 °C for ten minutes. The ~500 bp product was identified by separating a fraction of the reaction products on a 1% agarose in TAE gel and purified using the QIAquick PCR Purification kit (QIAGEN). Using the *EcoR* I and *Hind* III restriction sites contained within the primers, the amplified segment was inserted into complementary sites of pET28a (Novagen) using T4 DNA ligase (NEB). Following transformation into competent *E. coli* DH5α, DNA from isolated clones was characterized by restriction analysis and sequencing. One successful recombinant expression vector was then transformed into BL21-Gold(DE3)pLysS cells (Stratagene). Expression of the expected 18 kDa protein containing Vsr (aa 1 to 156) fused to a N-
terminal 6 x His tag was confirmed by Western blot analysis using an anti His tag antibody (Amersham).

To prepare cell pellets containing over-expressed Vsr, an overnight culture grown in LB containing 25 µg/mL Cam and 50 µg/mL Kan, was diluted (100 µL to 10mL) in LB and grown on roller drum at 37 °C to mid log phase (OD\textsubscript{600} ~0.65). Expression of Vsr was induced by adding IPTG (Roche) to a final concentration of 0.75 mM following which the culture was grown, as above, for an additional 3 h. Cells were then harvested by centrifugation (6000 xg, 15 min, 4 °C) and stored at -20 °C for up to one week.

When protein extract bearing over-expressed Vsr was desired, cell pellets were lysed using BugBuster\textsuperscript{®} protein extraction reagent (Novagen). Briefly, 300-500 µL of lysis buffer (containing, by volume, 8/10ths reaction buffer (25mM HEPES, 100mM NaCl, 10mM MgCl\textsubscript{2}, 2mM βME, pH 8.0), 1/10th 10x BugBuster\textsuperscript{®}, and 1/10th SIGMA P8340 protease inhibitors cocktail) were used to resuspend the cell pellet and transfer it to a 1.5 mL tube. Disruption of the cell membrane proceeded by nutating the mixture for 15 min at RT. If the solution remained extremely viscous, 0.5 units of benzonase (Novagen) was added and the tube was nutated for an additional 10-20 minutes. Protein extract was isolated from insoluble material by centrifugation (13,000 xg, 15 min, 4 °C), transferred to a sterile tube, and kept on ice until use. Depending on the volume of lysis buffer used, the total protein concentration ranged from 3.3 to 5.6 mg/mL.

Radioactivity Incorporation Assay To prepare substrate for the radioactivity incorporation assay, the ODN 5' - C*A*C *G*G*T AGA TGG ACG CCT AGG TCT TCA GTA TGC T*T*G *A*A*ddC - 3' (where * denotes a phosphorothioate linkage and ddC represents dideoxy-cytosine) was annealed to 5' - T*T*C *A*A*G CAT ACT GAA GAC CTG GGC GTC CAT C*C*TG*ddC - 3' in a mixture (250 µL) containing 5 nmoles of each oligo, 10 mM Tris, and 20 mM NaCl. The resulting duplex (20 µM) was stored at -20 °C.
Radioactivity incorporation assays were performed by combining 20 pmoles duplex substrate and 70 μg protein extract in reconstitution buffer (25 mM HEPES, 100 mM NaCl, 30mM MgCl₂, 2mM βME, pH 7.5). Each reaction (Vᵢ = 50 μL) was also supplemented to contain 2 mM ATP (Roche), 100 μM β-NAD (free acid, grade I, Roche), 50 μg Creatine Kinase (Roche), 50 mM Creatine Phosphate (Roche), 1 nmole of each dNTP, and 3.3 pmoles of [α²⁻³⁻]dCTP (6000 Ci/mmol, NEN brand, Perkin Elmer). Reactions were incubated at 37 °C (for 0, 15, 30, 60, and 120 min), stopped by adding an equal volume of formamide plus 10 mM EDTA loading buffer, and stored at -20 °C until electrophoretic separation. Reaction mixtures were separated via PAGE (20% in TBE, 7 M Urea), and the radioactive signals were visualized with PhosphorImagery (Storm 840, Molecular Dynamics) and quantified using ImageQuant 5.2 software (Molecular Dynamics).

**Radioactivity Depletion Assays** To prepare substrates for the radioactivity depletion assays, 500 pmoles of ODNs 2, 4, 6, 8, 10 and 12 (Table 3.1) were each incubated at 37 °C with kinase (T4 PNK, 10 units, NEB) and 16.5 pmoles of [γ⁻³²⁻⁻]ATP (6000 Ci/mmol, NEN brand, Perkin Elmer) for 30 min under conditions optimal for the enzyme in a volume of 20 μL. After supplementation with 200 nmoles of cold ATP (Roche), the reactions proceeded for an additional 60 min. The resulting mixtures, containing fully 5' phosphorylated strands, were then heat inactivated by incubation at 65 °C for 25 min and added to mixtures containing 500 pmoles of ODNs 1, 3, 5, 7, 9 and 11 respectively, 500 pmoles of scaffold 5'- GAA GAC CTA GGC GTC C -3', and 3 μL of 10X Ligase buffer (NEB). Scaffolding ODN 5'- GAA GAC CTA GGC GTC C -3' was utilized for ODNs 1, 3, 5, 7, and 9, while ODN 5'- ATA CTG AAG ACC TAG G -3' was utilized for ODN 10. The resulting solutions (50 μL) were heated to 80 °C and cooled to RT slowly to allow the DNA fragments to anneal. A 5 μL aliquot of each sample (50 pmoles) was removed to analyze accurately the radioactivity signal from a known amount of DNA. To ligate the strands held together by the scaffold, each tube was supplemented with 1200 units of T4 DNA Ligase (NEB) and incubated overnight at 16 °C. PAGE analysis (20% in TBE, 7 M Urea) of a fraction of the pre- and post-ligation mixtures was
conducted to verify the presence of DNA fragments of the correct size. The radiolabeled 38-mers (1nt, 2nt, 3nt, 4nt, 5nt, and 10nt Table 3.2) were purified from all reactants by sequentially centrifuging the ligation mixture through two ProbeQuant G-50 Micro Columns (Amersham) according to the supplied protocol. A fraction of each purified 38-mer was separated by PAGE (20% in TBE, 7 M Urea) next to a fraction of the pre-ligation control. The radioactivity signals were visualized with PhosphorImagery (Storm 840, Molecular Dynamics) and quantified using ImageQuant 5.2 software (Molecular Dynamics). The stock concentration of each 38-mer was calculated by comparing the signal intensity of gel-separated product to that for the control aliquot of known concentration. Substrate mixtures were made by annealing, as described above, 50 pmoles of radiolabeled DNA to 50 pmoles of the ODN 5'- T*T*C*A*A*G CAT ACT GAA GAC CTG GGC GTC CAT CTA *C*C*G*T*G -3' in a total of 500 μL 50 mM NaCl. Additionally, a control duplex was prepared by annealing 50 pmoles of 1nt with 50 pmoles of 5'- T*T*C*A*A*G CAT ACT GAA GAC CTG GGC GTC CAT CTA *C*C*G*T*G - 3', as above, to produce a radiolabeled duplex containing a perfectly matched T:A pair at the mismatch site.

Radioactivity depletion assays were typically performed by combining 100 fmols duplex substrate and ~55 μg protein extract in reconstitution buffer (25 mM HEPES, 100 mM NaCl, 30mM MgCl2, 2mM βME, pH 7.5). Each reaction (Vt = 50 μL) was also supplemented to contain 2 mM ATP (Roche), 100 μM β-NAD (free acid, grade I, Roche), 50 μg Creatine Kinase (Roche), 50 mM Creatine Phosphate (Roche), and 500 pmoles of each dNTP. Reactions were incubated at 37 °C for 120 min, stopped by adding an equal volume of formamide plus 10 mM EDTA loading buffer, and stored at -20 °C until electrophoretic separation. Zero time points were analyzed for each substrate in each reaction condition to verify integrity of substrate and normalize for error. Reaction mixtures were separated via PAGE (20% in TBE, 7 M Urea), and the radioactive signals were visualized with PhosphorImagery (Storm 840, Molecular Dynamics) and quantified using ImageQuant 5.2 software (Molecular Dynamics).
In the reactions where the effect of DNA Ligase was studied, serial dilutions of *E. coli* DNA Ligase (10 units/μL, NEB) were made using 1X buffer, and 5-10 μL of the total 50 μL reaction volume were added to the reactions just before incubation. Under normal conditions, each reaction contained 100 fmoles of substrate and 35 μg protein extract. For assay conditions with excess substrate, 1000 fmoles of duplex and 45 μg protein extract was used in each reaction. All other buffering conditions, supplements, and handling procedures were the same as above.

Following quantification of radioactive signals, the intensity for each substrate at \( t = 0 \) was designated as 100% and used as a reference for the percent loss of signal from that substrate after reaction. Additionally, as a redundant control, the relative intensities of the zero time points were compared to those for the control aliquots to account for minute deviations from equal amounts of substrates in every reaction. Contribution of exonuclease degradation from the DNA ends was defined by signal loss from the control duplex containing no mismatch and subtracted from the respective data values. These values were then normalized so that the relative signal loss from position 1 was 100.
3.4 Results

The ability of whole cell extracts bearing over-expressed Vsr to carry out VSPR was first monitored using a simple assay involving incorporation of radiolabeled $[^{32}\text{P}]-d\text{CTP}$ into substrate DNA (Figure 3.1). It was hypothesized that if VSPR could execute repair of T:G mismatches in short duplexes, the repair event would incorporate dCTP opposite G and provide a radiosignal by which to measure the extent to which VSPR occurred over time. Gel separation of the reaction products confirmed this hypothesis and also fortuitously revealed the presence of a small amount of VSPR reaction intermediate (Figure 3.2). The incorporation of $[^{32}\text{P}]-d\text{CTP}$ increased over time in two species that detail the abundance of resynthesis product pre- (repair intermediate) and post- (repair product) ligation (Figure 3.3). The level of the signal in the repair intermediate remains relatively constant while the signal for the repair product increases steadily for the first 30 minutes and then levels off after approximately 1 h. The incorporation of $[^{32}\text{P}]-d\text{CTP}$ into substrate DNA was dependent on Vsr activity and too low to be detected in wild type strains (data not shown).

Upon the observation that whole cell extracts bearing over-expressed Vsr can carry out VSPR on T:G mismatches, we modified our assay to address the relative frequencies of VSPR patch sizes created in whole cell extracts. In this second assay, we started with mismatch containing duplexes that possessed a radiolabel in the phosphate backbone at a prescribed distance downstream from the mismatch. In these reactions, the ability to monitor VSPR depended on detecting radiosignal loss from substrate as the $^{32}\text{P}$-group was removed as part of the repair patch (Figure 3.4). Signal loss from the duplexes containing the signal adjacent to the mismatched T was representative of the total of repair on all substrates as replacement of mismatched T is a qualifier for repair. The relative signal loss for this position was therefore normalized to 100. For the substrates containing a radio signal downstream of the mismatch, the degree of signal loss from those substrates related the frequency that the resynthesis patch continued through the position of that signal. In quantifying the signal losses from a series of substrates...
containing radiosignals at prescribed distances downstream from the mismatch, we could therefore address the frequency of specific repair patch sizes.

The relative signal losses for position one, two, and three are very high and roughly equivalent (Figure 3.5). In contrast, the signal loss from position 4, 5, and 10 are significantly lower than that for the first three, indicating that a smaller fraction of VSPR events monitored included repair tracts of these sizes. For eight individual reaction sets analyzed (data not shown), the signal loss from position 10 was 98 ± 7% of that for position 5.

To investigate the effect of surplus DNA Ligase on the relative signal losses and address the step at which exonucleases access the incised DNA, we conducted reactions containing zero, 0.005 units, and 0.5 units of DNA Ligase. Signal losses from all the reactions showed a trend similar to what we observed previously (Figure 3.6). The relative losses for positions 1, 2, and 3 were similarly high and the signal losses for position 4 and 5 were significantly less. The relative signal loss values for positions 4 and 5 ranged from low fifties to low sixties, and showed no general trend that depended on the concentration of DNA Ligase. These data suggested that the observed noise did not result from a discontinuity in the VSPR pathway in between nick translation by Pol I and Ligation. Rather, the noise in these values was likely due to exonuclease removal of the radiosignal before nick translation could occur.

To further investigate the effect of DNA Ligase, we conducted assays in which an overwhelming amount of substrate was used. We hypothesized that, in the absence of sufficient DNA Ligase, we would observe equal loss of signal from all positions because DNA ends would persist and exonuclease activity would remove the radiosignal independent of how far downstream it was located. Our data are consistent with this hypothesis. The relative signal loss in the reactions containing no supplemental DNA Ligase are similar and do not follow a trend of decreasing values (Figure 3.7, white columns). The absence of decreased signal losses at positions 4 and 5 are consistent with
the notion that exonuclease activity is degrading the DNA from the nick site and removing the radiolabel.

Additional roles for DNA Ligase are suggested by the data from the reactions containing excess substrate and supplemented protein (Figure 3.7). Not only do the reactions supplemented with additional enzyme show the characteristic decrease in relative signal loss around position 3, but they reveal much greater signal losses at positions 1, 2, and 3 as compared to the reactions containing no DNA Ligase. This observation was surprising because the protein responsible for terminating the repair event had a positive influence on the number of events initiated.

The relative signal losses for the reaction containing excess DNA Ligase showed a gradual decrease as the distance from the mismatch to the radiolabel increased (Figure 3.7, black columns). These results showed a slightly broader distribution of relative signal losses than previously observed, and indicate that repair tract sizes in these reaction conditions were approximately 2 to 4 deoxynucleotides in length. The relative signal loss from position 5 was similar to the relative signal losses from this position in previous reactions. This result is consistent with the notion that the majority of noise in this assay is due to exonuclease degradation from the Vsr incision site before Pol I and DNA Ligase can restore the DNA to its native form.
3.5 Discussion

The Very Short Patch Repair (VSPR) pathway was discovered upon the observation that the recombination frequencies of certain amber mutations in the \((cl)\) gene of bacteriophage \(\lambda\) were higher than predicted by the physical distance separating them (19). Further study of the \(am\theta\) amber mutation revealed that these events arose from \(C \rightarrow T\) transition mutations in a 5' - CAG -3' glutamine codon located within the Dcm recognition sequence 5' - CCAGG -3' (6). The recombinants observed included co-repair of markers within 10 nucleotides of the mismatch but not in areas further away (6). Moreover, recombinants were found to contain the C:G genotype and were not affected by the adenine methylation activity of Dam (5). These observations distinguished VSPR from methyl-directed mismatch repair, and spawned many studies to understand its mechanism.

Lieb and coworkers launched a thorough investigation into the length of the repair tracts created by VSPR (7). By crossing all available repair-prone amber mutations in gene \(cl\) with available \(cl\) markers up to 68 bp away, their data afforded a comprehensive picture of VSPR tract lengths. They observed that the frequency of co-repair for markers 10 base pairs and further away was the same. Therefore, VSPR repair tracts rarely exceed 10 bases \(in vivo\). Closer examination revealed that co-repair was only observed for markers downstream of the amber mutation, and markers as close as 5 base pairs away were not always co-repaired. The very short nature of VSPR repair tracts was becoming clear. Using this assay to investigate the retention of markers within a few base pairs of the amber mutation, however, was complicated.

The authors noted that one drawback of measuring the repair tract length using their recombination assay was that VSPR needed to act on T:G mismatches which were located close to another mismatch, the marker. It was hypothesized that the base pairing at one mispair could hinder the base pairing at another, and thereby affect repair of the T:G mismatch. Thus, interpreting the frequency to which repair tracts went short
distances was limited. One primary goal of ours was to design an assay free from such limitations.

Our *in vitro* assay utilizes a series of duplex substrates that all have the same sequence and bear a centrally located T:G mismatch contained in the sequence 5'-CTAGG-3'. The substrates differ by the location of a radioactive phosphate in their backbone that is adjacent to the mismatched thymine (position 1) or another base downstream of the mismatch (position 2, 3, 4, 5, or 10). Repair of the mismatch proceeds by the removal of the mismatched T and some number of nucleotides downstream. Therefore, the number of nucleotides that are removed as part of the repair patch may be probed by monitoring the signal loss from each of the substrates as VSPR occurs (Figure 3.3).

Before conducting these assays, however, we first verified that extracts bearing over-expressed Vsr would carry out complete VSPR on short substrate duplexes. We constructed a reaction using a non-radioactive substrate and extracts supplemented with a small amount of \( [\alpha^{32}\text{P}]-\text{dCTP} \) (Figure 3.1). In doing so, we were able to monitor the repair event by restoration of a correct C:G base pairing in the DNA. Results in Figure 3.2 show that we were able to resolve the Pol I resynthesis product pre- (repair intermediate) and post- (repair product) ligation.

The relative abundance (Figure 3.3) of the two species revealed that ligation of the resynthesis product was partially rate limiting. The ability to see the repair intermediate before ligation revealed that the polymerase was providing substrate for the DNA Ligase faster than the DNA Ligase could act on that substrate. Therefore, the repair intermediate was kinetically competent and was observed as part of the reaction mixture. The relatively small increase in the signal for the repair intermediate, as compared to that for the repair product, demonstrated that while the ligation step was partially rate limiting, the rate was not dramatically slower than those of the polymerase activities. Therefore, our reaction conditions were competent for VSPR.
We next conducted our assay on duplexes containing the radiosignal at positions adjacent to and downstream of the T:G mismatch. Our *in vitro* results (Figure 3.5) are in good agreement with what has previously been estimated *in vivo* (7), and provide additional detail as to the distribution of VSPR patch lengths. The frequencies of repair tracts 1, 2, and 3 bases in length were similarly high. The frequencies of repair tracts 4, 5, and 10 bases in length were also similar to each other but lower than the first three positions. Therefore, the distribution of VSPR tracts we observed was quite narrow and consistent with a length of 3 deoxynucleotides.

Our duplexes experienced a significant amount of nonspecific radiosignal loss. VSPR repair tracts rarely exceed 10 nucleotides in length (7). Therefore, the radiosignal loss we observed from this position was a result of interfering activity. We believe the signal losses for positions 4, 5, and 10 are predominantly from nonspecific degradation of nicked DNA by nuclease. Each strand of our duplexes contained phosphorothioate bonds to deter exonucleolytic degradation from the ends. Additionally, loss of radiosignal from exonuclease degradation through the substrate ends was accounted for using a control duplex that did not contain a mismatch. We therefore surmised that the loss of radiosignal was due to 5' → 3' exonuclease degradation from the DNA ends produced by Vsr incision.

We rationalized that the degradation could occur (1) after Vsr incision but before Pol I nick translation, or (2) after Pol I nick translation but before Ligation of the juxtaposed DNA ends. If the latter case was the primary contribution to noise in the assay, it could be overcome by supplementing the reactions with purified *E. coli* DNA Ligase. Addition of surplus DNA Ligase had no significant effect of reducing the relative signal loss at positions 4 and 5 (Figure 3.6). Therefore, exonuclease degradation of duplex DNA takes place just after incision by Vsr. Given this observation, the substrate duplexes we used may not have been optimal.

Vsr binds and cleaves DNA duplexes composed of 5'- CTAGG -3' and 5'-C5mCTGG -3' (where 5mC is 5-methylcytosine) strands more efficiently than those
composed of 5'-CTAGG-3' and 5'-CCTGG-3' strands (15). This preference has also been rationalized using structural data (20;21), and is justified as the hemimethylated substrate is the form encountered in vivo (1). Our duplex substrates were not hemimethylated and therefore could have experienced relatively destabilized binding to Vsr. Perhaps, the use of hemimethylated substrates in our assay would increase Vsr binding and decrease exposure of DNA ends susceptible to degradation.

Conducting a repair reaction under standard conditions plus supplemental DNA Ligase also allowed us to study the effects of DNA Ligase on VSPR. We rationalized that if the size of the repair tract created by Pol I was affected by the concentration of DNA Ligase present, we would be able to observe that affect with our assay. We first constructed assay conditions that allowed us to observe the contribution of surplus DNA Ligase to reactions (Figure 3.6). The size and distribution of the repair patch was not affected by the addition of DNA Ligase, providing good evidence that surplus DNA Ligase does not influence Pol I to creating short repair tracts.

To test the contribution of DNA Ligase to reactions that did not have sufficient DNA Ligase activity to keep up with substrate repair, we used reaction conditions which contained excess substrate and a range of DNA Ligase supplementation. In these reactions (Figure 3.7), the added protein did benefit the repair and assist in VSPR completion. Without the addition of DNA Ligase the relative signal loss from the substrates was not typical for a 3 nucleotide repair tract. Instead, the signal losses were relatively equivalent and did not decrease significantly between positions 1 and 5. In contrast to reactions containing no supplement, those containing additional DNA Ligase did show signal losses characteristic for repair tracts of 2 to 4 nucleotides in length. These results suggested that DNA Ligase does not have a significant effect on the length of VSPR tract lengths.

One of the most interesting observations was that the relative level of VSPR was greater in reactions containing sufficient amounts of DNA Ligase (Figure 3.7, black columns). We did not expect that the concentration of DNA Ligase, responsible for
terminating the repair event, would have an influence on the efficiency of repair initiation. This finding suggests that VSPR might be coordinated so that the proteins are more efficient at starting repair events that can be successfully completed because sufficient downstream repair pathway proteins are present. VSPR is possibly most efficient when accomplished by multiprotein repair complexes.

It has been observed that the MutS, DNA Ligase, and Pol I proteins all interact with the β sliding clamp protein (22). This protein was originally studied for its role in tethering DNA polymerase III holoenzyme to DNA by forming a ring around the DNA and thereby assuring replication processivity. More recently, however, its ability to coordinate multiple proteins on DNA has received attention in the context of DNA repair. One hypothesis (22) posits that the clamps loaded on the lagging strand during DNA synthesis can assist MutS damage recognition by localizing MutS on newly synthesized DNA. If the presence of DNA Ligase was somehow beneficial for this interaction, the result may be more efficient damage recognition and repair initiation. Though the substrates used in this study are likely too small for β-clamp sliding to affect the efficiency of repair initiation, the possibility remains that the β-clamp can organize a complete assembly of machinery for optimal repair.
**Figure 3.1.** Method for detecting VSPR by monitoring the incorporation of radioactivity into substrate DNA

5' - CACGGTAGATGGACGCCTAGGTCTTCAGTATGCTTGAA - 3'
3' - GTGCCATCTACCTGCAGGGTCCAGAAGTCATACGAACTT - 5'

MutS recognizes and binds to the mismatch

Vsr creates an incision 5' to the mismatched T

ATP
ADP + Pi

MutS/MutL complex presents the mismatch in a loop

DNA polymerase I removes and replaces a short patch of DNA

DNA ligase seals the nick

5' - CACGGTAGATGGACGCCTAGGTCTTCAGTATGCTTGAA - 3'
3' - GTGCCATCTACCTGCAGGGTCCAGAAGTCATACGAACTT - 5'
Figure 3.2. Extracts from cells over-expressing recombinant Vsr incorporate $[\alpha^{32}\text{P}]$-dCTP into short duplex substrates via VSPR.
Figure 3.3. Graphical representation of results from measuring VSPR via radioactivity incorporation

<table>
<thead>
<tr>
<th></th>
<th>0 min</th>
<th>15 min</th>
<th>30 min</th>
<th>60 min</th>
<th>120 min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intermediate</td>
<td>1</td>
<td>9</td>
<td>10</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>Product</td>
<td>0</td>
<td>51</td>
<td>74</td>
<td>97</td>
<td>100</td>
</tr>
</tbody>
</table>
**Figure 3.4.** Method for detecting VSPR by measuring the depletion of radioactivity signal from substrate DNA

5′ - CACGGTAGATGGACGCCCAGGTCTTCAGTATGCTTGAA - 3′
3′ - GTGCCATCTACCTGCGGGTCCAGAAGTCATACGAACCTT - 5′

5′ - CACGGTAGATGGACGCCCAGGTCTTCCTTCTCATATGCTTGAA - 3′
3′ - GTGCCATCTACCTGCGGGTCCAGAAGTCATACGAACCTT - 5′

MutS recognizes and binds to the mismatch

Vsr creates an incision 5′ to the mismatched T

MutS/ MutL complex presents the mismatch in a loop

ADP + P

ATP

DNA polymerase I removes and replaces a short patch of DNA

DNA ligase seals the nick

5′ - CACGGTAGATGGACGCCCAGGTCTTCCTTCTCATATGCTTGAA - 3′
3′ - GTGCCATCTACCTGCGGGTCCAGAAGTCATACGAACCTT - 5′
Figure 3.5. Distribution of VSPR patch lengths created \textit{in vitro}
Figure 3.6. Addition of surplus DNA Ligase to VSPR reactions
Figure 3.7. Addition of DNA Ligase to VSPR reactions containing excess substrate

<table>
<thead>
<tr>
<th>Position</th>
<th>No Ligase added</th>
<th>0.2 units Ligase</th>
<th>20 units Ligase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position 1</td>
<td>100</td>
<td>108</td>
<td>132</td>
</tr>
<tr>
<td>Position 2</td>
<td>111</td>
<td>111</td>
<td>130</td>
</tr>
<tr>
<td>Position 3</td>
<td>89</td>
<td>100</td>
<td>111</td>
</tr>
<tr>
<td>Position 4</td>
<td>86</td>
<td>89</td>
<td>81</td>
</tr>
<tr>
<td>Position 5</td>
<td>78</td>
<td>73</td>
<td>46</td>
</tr>
</tbody>
</table>
Table 3.1. Oligodeoxynucleotides (ODNs) used to construct duplexes containing internal radiolabels at prescribed distances away from the mismatched T

<table>
<thead>
<tr>
<th></th>
<th>Oligonucleotide Sequence (5' → 3')</th>
<th>Complement Sequence (5' → 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5'- C<em>A</em>C<em>G</em>G*TAGATGGACGCC - 3'</td>
<td>3' - TAGGTCTTCAGTATGC<em>T</em>T<em>G</em>A*A</td>
</tr>
<tr>
<td>2</td>
<td>5'- TAGGTCTTCAGTATGC<em>T</em>T<em>G</em>A*A - 3'</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5'- C<em>A</em>C<em>G</em>G*TAGATGGACGCCT - 3'</td>
<td>5' - AGGTCTTCAGTATGC<em>T</em>T<em>G</em>A*A - 3'</td>
</tr>
<tr>
<td>4</td>
<td>5'- AGGTCTTCAGTATGC<em>T</em>T<em>G</em>A*A - 3'</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5'- C<em>A</em>C<em>G</em>G*TAGATGGACGCCTA - 3'</td>
<td>5' - GGTCTTCAGTATGC<em>T</em>T<em>G</em>A*A - 3'</td>
</tr>
<tr>
<td>6</td>
<td>5'- GGTCTTCAGTATGC<em>T</em>T<em>G</em>A*A - 3'</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>5'- C<em>A</em>C<em>G</em>G*TAGATGGACGCCTAG - 3'</td>
<td>5' - GTCTTCAGTATGC<em>T</em>T<em>G</em>A*A - 3'</td>
</tr>
<tr>
<td>8</td>
<td>5'- GTCTTCAGTATGC<em>T</em>T<em>G</em>A*A - 3'</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>5'- C<em>A</em>C<em>G</em>G*TAGATGGACGCCTAGG - 3'</td>
<td>5' - TCTTCAGTATGC<em>T</em>T<em>G</em>A*A - 3'</td>
</tr>
<tr>
<td>10</td>
<td>5' - TCTTCAGTATGC<em>T</em>T<em>G</em>A*A - 3'</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>5'- C<em>A</em>C<em>G</em>G*TAGATGGACGCCTAGGTCTTC - 3'</td>
<td>5' - AGTATGC<em>T</em>T<em>G</em>A*A - 3'</td>
</tr>
<tr>
<td>12</td>
<td>5' - AGTATGC<em>T</em>T<em>G</em>A*A - 3'</td>
<td></td>
</tr>
</tbody>
</table>

Complement: 5' - T*T*C*A*A*GCATACTGAAGACCTGGGCGTCCATCTA**C*G*T*G - 3'  
* = phosphorothioate linkage
Table 3.2. DNA strands containing radiolabels at prescribed distances away from the T

<table>
<thead>
<tr>
<th>Distance</th>
<th>Strand Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1nt</td>
<td>5'- C<em>A</em>C<em>G</em>G<em>TAGATGGACGCTGTTCAGAGTTGATA</em>G* - 3'</td>
</tr>
<tr>
<td>2nt</td>
<td>5'- C<em>A</em>C<em>G</em>G<em>TAGATGGACGCTGTTCAGAGTTGATA</em>G* - 3'</td>
</tr>
<tr>
<td>3nt</td>
<td>5'- C<em>A</em>C<em>G</em>G<em>TAGATGGACGCTGTTCAGAGTTGATA</em>G* - 3'</td>
</tr>
<tr>
<td>4nt</td>
<td>5'- C<em>A</em>C<em>G</em>G<em>TAGATGGACGCTGTTCAGAGTTGATA</em>G* - 3'</td>
</tr>
<tr>
<td>5nt</td>
<td>5'- C<em>A</em>C<em>G</em>G<em>TAGATGGACGCTGTTCAGAGTTGATA</em>G* - 3'</td>
</tr>
<tr>
<td>10nt</td>
<td>5'- C<em>A</em>C<em>G</em>G<em>TAGATGGACGCTGTTCAGAGTTGATA</em>G* - 3'</td>
</tr>
</tbody>
</table>

● denotes the position of the radiolabel
* = phosphorothioate linkage
3.6. REFERENCES


CHAPTER 4

$\alpha$-Methylguanine – Formation, Consequences, and Repair
4.1. O\textsuperscript{6}-METHYL GUANINE – A BIOLOGICALLY IMPORTANT ADDUCT

In the early 1960s, alkylating agents were thought to cause mutations primarily through modification of the N7 position of guanine, resulting in atypical base pairing during replication (1;2). Acid hydrolysates of DNA that had been treated with a variety of agents demonstrated the major product to be 7-alkyl-guanine (3;4). However, the observation that the ability of an agent to produce N7-guanine adducts did not correlate with its mutagenicity suggested that some less abundant DNA modification could be responsible for the mutagenic effects of alkylating agents (3;5).

In 1963, Friedman reported that treating deoxyguanosine with diazomethane produced an alkylation product, O\textsuperscript{6}-methylguanine (O\textsuperscript{6}mG), that was converted to guanine upon perchloric acid hydrolysis (6;7). This finding revealed that certain alkylation products did not survive typical DNA acid hydrolysis workups and could be the hidden source of the mutagenic properties of certain agents. The identification of such products generated within DNA would require modifications to the typical hydrolysis procedure.

In the late 1960s, Loveless isolated the O\textsuperscript{6}-alkylG product by allowing deoxyguanosine to react with N-methylnitrosourea (MNU), N-ethyl nitrosourea (ENU), or ethyl methanesulphonate (EMS) (5). This lesion, could therefore be produced by classical alkylating agents and contribute to their mutagenic effects in vivo. The hydrogen bonding properties of the O\textsuperscript{6} and N1 atoms in G are altered in O\textsuperscript{6}mG and change its base pairing properties as a consequence. In response to these findings, Loveless commented,

"The relationship between nucleic acid alkylation and the carcinogenicity of nitrosamines and nitroamides compared with "classical" alkylating agents is also unsatisfactory. The extent of nucleic acid alkylation, with which a correlation with carcinogenicity has been vainly sought, has been estimated after acid hydrolysis. It is particularly interesting, in this context, that Kruger et al
found no ethylation of tissue RNAs following NEU administration. This substance is a powerful carcinogen and of all the reagents discussed here affords the largest yield of O-alkyl-deoxyguanosine. It seems therefore that it may yet be possible to relate carcinogenicity with nucleic acid reaction but only when the extent of O-alkylation can be measured.”

These observations paved the way for a better understanding of the mechanisms underlying the mutagenicity and cytotoxicity of alkylating agents. It is now generally accepted that although O\textsuperscript{6}mG is one of the less abundant lesions created by alkylating agents, it is a primary adduct contributing to mutagenesis and carcinogenesis (5;8;9). The following sections detail what has been learned about the lesion and how it is repaired in cells.

4.2. O\textsuperscript{6}-Methylguanine is Mutagenic

DNA contains many sites susceptible towards alkylation (Figure 4.1), and the relative proportions of the different alkylation products is dependent on the reaction mechanism of the agent and the secondary structure of the DNA target. Alkylating agents can be categorized into two groups that differ with respect to the transition state of reaction with the DNA bases. Agents such as methyl methanesulphonate (MMS), dimethylsulphate (DMS) and methyl iodide (MeI) react through a S\textsubscript{N}2 (bimolecular) transition state because the nitrogen groups on DNA are sufficiently nucleophilic for reaction before the leaving group has left. On the other hand, the activated forms of agents such as N-methyl-N’-nitro-N-nitrosoguanidine (MNNG)(Figure 4.2) and MNU are less sensitive to the groups on DNA with which they react and through a S\textsubscript{N}1 (unimolecular) transition state alkylate the oxygen atoms in addition to the nitrogen groups in DNA.

The difference in reactivity between S\textsubscript{N}1 and S\textsubscript{N}2 agents is roughly correlated with the Swain-Scott “selectivity” constant (S) that ranges from 0 to 1 (10). Agents with
low S values tend to react more extensively with the less nucleophilic oxygen centers on DNA, while agents with high S values react with the more nucleophilic nitrogen centers (11). For example, the $S_{1}$ agent MNU has a S value of 0.42 (12) and produces about 20-fold more $\text{O}^{6}\text{mG}$ than the $S_{2}$ alkyator MMS (13). A summary of the alkylation product spectra for a typical $S_{2}$ agent, MMS, and a typical $S_{1}$ agent, MNU, is presented in Table 4.1.

In duplex DNA, a number of the reactive sites are involved in base pairing and therefore shielded from alkylation. Accordingly, the most abundant adducts created in double strand DNA are 7-methylguanine (7mG), 3-methyladenine (3mA), and $\text{O}^{6}\text{mG}$. In single stranded DNA, the 1-methyladenine (1mA) and 3-methylcytosine (3mC) lesions are efficiently produced, as well (14). In addition to the DNA bases, alkylation of the phosphate backbone can occur to create a mixture of R and S diastereoisomers (Figure 4.3) (15).

In addition to the exogenous agents discussed above, there are a number of endogenous sources that can damage DNA and cause mutations. A number of biomethylation processes utilize S-adenosylmethionine (SAM) as a methyl source, and this molecule has been observed to weakly methylate DNA in vitro (16). However, varying the cellular levels of SAM over a 100-fold range had no effect on the spontaneous mutagenesis in E. coli (17). Bacterially catalyzed nitrosation has been shown to result in DNA alkylation, likely through the activation of amino acids and peptides (18;19). Moreover, mutants deficient in nitrite reductase activity was less susceptible to spontaneous mutagenesis suggesting that nitrosation may be a biologically relevant source of mutations (18;20;21). In humans, potent mutagens such as nitrosamines and related N-nitroso compounds are produced as a result of metabolism (22).

There are many factors that influence whether a particular DNA alkylation product will be harmful to the cell. These factors include if the lesion is mutagenic, a replication block, and/or efficiently repaired. The 7mG lesion is not mutagenic (13) and
does not interfere with transcription (23) or replication (24). However, methylation at this position increases the rate at which this base depurinates (25). The resulting apurinic (AP) site can result in a mutation during DNA replication (26). Alternatively, 7mG residues have been shown to undergo an imidazole ring opening that produces a formamidopyrimidine residue which is a block to DNA synthesis in vitro (27). Therefore, one cannot base the biological importance of lesions merely upon their abundance.

By the 1970s, it was well understood that in addition to reacting with DNA in vitro the S\textsubscript{N}1 alkylating agent MNNG was mutagenic in vivo (28-30). Using a forward mutational assay, Coulondre and Miller studied the base changes necessary to generate nonsense codons and determined that the predominant class of mutations caused by MNNG was G:C \rightarrow A:T transition mutations (31;32). Years earlier, Loveless had hypothesized T to pair with O\textsubscript{6}mG during replication (5), and these results were consistent with that notion. Subsequently, this hypothesis was directly demonstrated in vivo using adducts that were site specifically incorporated into DNA (8;33;34). The preference of O\textsubscript{6}mG to base pair with T during replication was also confirmed in vitro using a variety of purified polymerases (35).

Initially, the mutagenicity of O\textsubscript{6}mG was thought to arise from the stability of the resulting O\textsubscript{6}mG:T base pair, however this idea proved too simplistic as more data on O\textsubscript{6}mG pairings became available (36). The stability of duplexes containing O\textsubscript{6}mG opposite the four possible bases demonstrated that the O\textsubscript{6}mG:C base pair was actually the most stable (37). Moreover, 2D-NMR studies (38;39) revealed a O\textsubscript{6}mG:T pair in solution with only one strong hydrogen bond between the exocyclic amine group of O\textsubscript{6}mG and the O\textsubscript{6} of T (Figure 4.4). The O\textsubscript{6} methyl group was observed to be syn to the N1 of O\textsubscript{6}mG and likely overlapping with the O\textsubscript{6} of T sterically. This clashing would weaken hydrogen bonding between the N1 of O\textsubscript{6}mG and the N3 of T, providing a weaker association overall. In comparison, O\textsubscript{6}mG:C pairs were observed to pair via two hydrogen bonds in a wobble arrangement (Figure 4.4). These data suggested that the preference for O\textsubscript{6}mG to pair with T is controlled by factors apart from thermodynamics.
The original base pairing model had the methyl group of \( \textit{\text{O}}\text{mG} \) in the \textit{anti} conformation (5). Subsequently, this conformer was crystallized and studied using X-ray diffraction analysis (40). This Watson-Crick pairing is stabilized by two hydrogen bonds between \( \textit{\text{O}}\text{mG} \) and T (Figure 4.4), and the mutagenic pattern of \( \textit{\text{O}}\text{mG} \) suggests that this \textit{anti} form of the methoxyl group exists in addition to the \textit{syn} form \textit{in vivo} (41).

\textit{In vitro} replication studies on DNA substrates containing \( \textit{\text{O}}\text{mG} \) also suggest that the lesion can exist in two conformers, only one of which can be bypassed during replication (42). When \( \textit{\text{O}}\text{mG} \) was paired opposite T or C in DNA, about one half of the substrate could be replicated with the Klenow fragment of DNA Pol I. Substrates where the \( \textit{\text{O}}\text{mG} \) had been dealkylated showed no replication blockage, suggesting that only one half of \( \textit{\text{O}}\text{mG} \) is easily bypassed by polymerases.

It has been hypothesized that the \( \textit{\text{O}}\text{mG}:T \) pairs with the \( \textit{\text{O}}\text{mG} \) methyl group \textit{anti} is the replicated form on account of its unperturbed geometry (39;43). Swann suggested the structure of T opposite \( \textit{\text{O}}\text{mG} \), as opposed to C opposite \( \textit{\text{O}}\text{mG} \), provides a more suitable geometry for forming the phosphodiester bond of T to the growing strand (36). Additionally, T opposite the lesion may evade exonucleolytic proofreading better than C (36). This hypothesis was supported by the observations that the rate limiting step for nucleotide incorporation opposite \( \textit{\text{O}}\text{mG} \) by Klenow was the formation of the phosphodiester bond, and bond formation was faster for T than C (44).

The frequency with which \( \textit{\text{O}}\text{mG} \) elicits G:C \( \rightarrow \) A:T transition mutations as a function of the sequence context has been studied. Dosanjh and coworkers first conducted \textit{in vitro} kinetic assays to study the extension of \( \textit{\text{O}}\text{mG} \) paired with T or C when the adduct was flanked by cytosines or thymines (45). They reported that the adduct was replicated more readily when flanked by cytosines and increased G:C \( \rightarrow \) A:T transitions mutations. They also observed that there was a 2-3 fold preference for incorporation of T opposite the lesion as compared to C, regardless of the context. However, extensive investigation of the mutation frequency of \( \textit{\text{O}}\text{mG} \) \textit{in vivo} does not support these findings.
Delaney and Essigmann (46;47) studied the mutation frequency of $\check{O}$mG in the 16 nearest neighbor sequence contexts and observed that the lesion pairs with T nearly 100% of the time, independent of the surrounding bases. The latter assay represented the outcome for replication past $\check{O}$mG by the DNA polymerase holoenzyme in vivo, as opposed to the Klenow fragment used in in vitro, and provides the most accurate picture of $\check{O}$mG mutagenesis in E. coli.

### 4.3. O6-METHYLGUANINE IS CYTOTOXIC

In addition to its mutagenic properties, $\check{O}$mG is cytotoxic in E. coli and in higher organisms. Studies done by Karran and Marinus in the early 1980s identified a number of E. coli mutants that lacked the Dam adenine methylase ($dam$) and were sensitive to agents that produced $\check{O}$mG such as MNNG (48). The Dam protein (see Chapter 1) was known to methylate d(GATC) sequences and provide a signal by which the parental DNA strand could be discriminated from the newly synthesized strand for mismatch repair (MMR) (49-51). Therefore, it was hypothesized that the sensitivity of $dam$ strains was dependent on the inability to distinguish the appropriate strand for MMR. Consistent with this idea, the sensitivity of $dam$ strains was abolished when the cells additionally lacked the mismatch recognition protein, MutS (52). It was later shown that $\check{O}$mG:T and C base pairs were recognized by MutS and substrates for MMR (52;53). Furthermore, subsequent mutagenesis studies revealed that MMR preferentially acted on $\check{O}$mG paired with T rather than C (54), compounding the deleterious effects of $\check{O}$mG.

There are a number of theories of how MMR introduces sensitivity to $\check{O}$mG:T base pairs. One theory posits that replication past the lesion creates $\check{O}$mG:T mismatches that are regenerated when acted on by MMR machinery (55). As a consequence, reiterative attempts by MMR to repair $\check{O}$mG:T pairs could create persistent strand breaks that may arrest DNA replication (54). It has also been hypothesized that two MMR events that are close in proximity could produce overlapping repair tracts on opposite
strands and give rise to a toxic double strand break (56). This theory, however, received criticism on the grounds that mismatches would not occur frequently enough to produce excision tracts that would overlap (57). Years later, the hypothesis that a single mismatch could give rise to a double strand break was vindicated when it was observed that in the absence of d(GATC) methylation, MutH can cleave both DNA strands (58). In higher organisms, MMR intermediates have been observed to give rise to double strand breaks that can trigger the apoptotic pathway (59;60) and in some cases block polymerase replication (61). However, the roles of these events in cellular death are not completely understood. Therefore, while the toxicity of agents which produce O6mG has been established to involve MMR and double strand breaks, the precise mechanism is still under investigation.

4.4. THE ADAPTIVE RESPONSE TO ALKYLATING AGENTS

It was discovered in 1970 that the O6mG lesions created by MNNG were removed from DNA (62), however the mechanism by which the damage was removed was not understood until discovery of the adaptive response. In the mid 1970s, Cairns and Samson were attempting to devise an experimental system to segregate mutated from nonmutated DNA in bacterial cells that were exposed to low doses of alkylating agents for a short time. While these efforts never came to fruition, they realized that their system could be used to examine the mutation rates invoked by low levels of these agents. Subsequent experiments measured the mutation rates affiliated with exposure of cells to low doses of MNNG, and demonstrated that, under certain conditions, mutations would accumulate for the first hour but not afterwards (63).

Additional experiments confirmed this phenomenon and revealed that when E. coli cells were exposed to low levels of MNNG before challenging the cells with a much higher dose, the cells had a significant resistance to the lethal and mutagenic effects of the agent (63-65). This resistance was also discovered to be dependent on protein synthesis, suggesting that this response involves the synthesis of new proteins to protect
the cells from alkylation damage (63). Demonstration that the induction of proteins relates an increase in removal of O₆mG lesions from DNA was consistent with the ideas that this lesion accounted for most MNNG induced mutagenesis and was targeted by the adaptive response for increased repair (66). Indeed, quantitative assessment of the removal of O₆mG pre and post induction revealed that the removal activity increased by approximately 1000-fold (67). This phenomenon is now referred to as the adaptive response to alkylation damage (68;69).

By this time, measuring DNA damage and repair was readily accomplished by monitoring the creation and retention of radiolabeled DNA damage. Typically, cells were treated with a radiolabeled alkylating agent and the high molecular weight DNA was isolated and analyzed for radiosignal. Therefore, the observation that there was less radiosignal from DNA containing O₆mG lesions (where the signal was on the methyl group) from cells that were adapted, as compared to those that were not, was expected (66). It was also expected that released DNA adduct could be recovered in the acid soluble fraction containing excised DNA products, however, this was not the case (67;70). Further investigation revealed that the radioactive methyl groups, released from DNA, were tightly associated with protein (71) that had methyltransferase (MTase) activity (72). Subsequently, the Ada protein was isolated and its ability to demethylate O₆mG was directly demonstrated (73). Additionally, Ada was shown to be the positive regulator of the adaptive response as cells that lacked functional Ada were significantly more sensitive to MNNG treatment (74;75).

In the past fifteen years, our understanding of the adaptive response has developed substantially. In addition to ada, a number of genes involved in the adaptive response were mapped (76) and identified by their contribution to decreasing the mutagenecity or cytotoxicity of alkylating agents. It is now understood that, by a mechanism described below, the Ada protein regulates increased expression of four genes, ada, alkB, alkA, and aidB, upon exposure of E. coli to sub-lethal doses of alkylating agents (Figure 4.5) (77). In total, these genes, which map at 45, 47 and 95 min
on the *E. coli* genetic map (78-80), relate protection against the mutagenecity of methylation, ethylation, propylation, and bulkier adducts, as well (79;81).

In efforts to understand how each repair protein contributes to the protective effects of the adaptive response, mutant strains were studied. The *ada* mutants were found to be sensitive to both the mutagenecity and cytotoxicity of alkylating agents (74) owing to, as discussed above, the mutagenic and cytotoxic effects of O6mG (5;8;9). The number of Ada molecules increases from one or two molecules per cell in an uninduced state (82) to approximately 1000 upon induction (67), in a process that takes about one hour (83). This Ada protein, and how it regulates the adaptive response, is discussed in detail below.

The AidB protein is the product of the *aidB* gene, is so named because of its alkylation inducibility. The 60 kDa protein shows homology with isovaleryl-coenzyme A dehydrogenases (84). Indeed, crude extract from cells bearing over-expressed AidB protein show isovaleryl-coenzymeA dehydrogenase activity (84). Over-expression of the protein leads to reduced mutagenesis after exposure to alkylating agents, and thus the role of AidB has been hypothesized to involve repair of unidentified DNA lesions or the detoxification of the alkylating agents or their reaction intermediates (84-86). Recently, AidB was reported to use flavin adenine dinucleotide (FAD) and bind to DNA directly (86). In light of these observations, another hypothesis posits that AidB is a methylated base dehydrogenase.

Strains deficient in *alkA* were found to be sensitive to killing, but not mutagenesis (87). The AlkA protein was discovered to be a glycosylase that has the ability to excise a wide variety of damaged bases including 3mA, 7mG, 7-methyladenine (7mA), 3-methylguanine (3mG), O6-methylthymine (O6mT), and O6-methylcytosine (O6mC) (88;89). In contrast to the other alkylation lesions, 3mA was found to be biological significance as its ability to block DNA synthesis produces cytotoxicity (90;91). Therefore, upon a 10 to 20-fold induction as part of the adaptive response (77;87), the
primary importance of this glycosylase is to remove the 3mA methylation product. For this reason, AlkA is oftentimes referred to as 3-methyladenine DNA glycosylase.

Strains that are deficient in \textit{alkB} were isolated as early as 1983 \cite{80} however the function of the protein remained unknown for a long time afterwards. While it was established that the bacterial \textit{alkB} gene conferred resistance to MMS (SN2) damage both in bacterial and human cells \cite{92}, it was not until repair studies were conducted with single stranded phage that substrates were proposed. The observation that AlkB proficient and deficient strains processed methylated single stranded DNA differently, suggested that the 1mA and 3mC lesions were AlkB substrates \cite{93}. These lesions are primarily created in single stranded DNA because the 1 position of A and the 3 position of C are not exposed in duplex structures (Figure 4.1). The discovery that AlkB shares a protein fold typical for dioxygenases inspired the idea that the protein could be a \(\alpha\)-ketoglutarate-Fe(II)-dependent dioxygenase \cite{94}. This notion was subsequently verified biochemically, when in the presence of Fe(II) and \(\alpha\)-ketoglutarate, AlkB exhibited repair of 1mA and 3mC substrates by an oxidative demethylation mechanism \cite{95-99}. Subsequently, AlkB has been shown to also repair 3-ethyldeoxycytidine \cite{100}, 1-methyldeoxyguanosine \cite{100}, 3-methylthymine \cite{101}, 1,\(N^6\)-ethenoadenine \cite{102}, and 3,\(N^4\)-ethenocytosine \cite{102}.

\section*{4.5. ADA - IN VIVO AND IN VITRO PROPERTIES}

Ada is the product of the \textit{ada} gene located at 47 min on the \textit{E. coli} genetic map \cite{80;103}, and is so named because of its central role in the adaptive response \cite{104}. Upon induction, levels of Ada increase from approximately one molecule per cell \cite{82} to approximately 1000 molecules per cell \cite{67}. It has also been reported that uninduced cells experience a 20-fold increase in Ada upon entry into stationary phase \cite{105}. The protein displays MTase activity (Figure 4.6) on \(\text{O}^6\text{mG} \) \cite{71}, \(\text{O}^4\text{mT} \) \cite{106} methylphosphetriesters \cite{15;88;107}, and larger alkyl groups, as well \cite{79;108;109}. However, the activity against alkyl groups generally decreases as the size of the groups
increases. For example, the rate of transferase activity for ethyl groups is about one tenth that for methyl groups in vivo (79). Ada additionally appears to repair O6-chloroethyl groups from guanine. These groups are formed by agents such as 1,3-bis(2-chloroethyl)-1-nitrosourea (BCNU) and N-(2-chloroethyl)-N-cyclohexyl-N-nitrosourea (CCNU) and are toxic on account of their ability to form interstrand cross links in DNA (110). The presence of Ada provides protection against the lethality of these agents, suggesting that Ada repairs chloroethyl monoadducts before they can cross link (111;112).

Despite the fact that the O6-methyl bond is stable at neutral pH (77) Ada carries out direct reversal of the methylation in about 1 second at 37 °C (113), or at a rate of k > 10⁻⁸ M⁻¹ sec⁻¹ (114). Additionally, Ada shows no requirement for Mg²⁺ or other cofactors (77). Therefore, this protein appears to be especially tuned to overcome significant energy barriers in its ability to remove alkyl groups from DNA.

Under limiting conditions, Ada was found to be more efficient at repairing O6mG as compared to O6mT (88). This preference was also observed in in vitro mutagenesis studies (115), as the mutation frequency for O6mG (0 - 1.7%) was lower than that for O6mT (12%). The alkyl groups on the O6 position of G and the O4 position of T are both exposed from the major groove of DNA. However, Ada does not repair any other forms of alkylation damage that project into the major groove. Additionally, Ada shows no MTase activity for N methylated residues in DNA (113). Therefore, the specificity of MTase activity on damaged DNA bases is apparently tuned for alkyl groups that project into the major groove of DNA and are oxygen-linked (88).

Ada repairs O6mG in single stranded DNA, albeit at a dramatically reduced rate (113). This finding is especially relevant since significant single stranded DNA regions exist in replication forks. The inefficiency of O6mG repair in these regions may be one of the factors that contributes to the known property of MNNG to generate mutations in clusters at replication forks (116-118). Additionally, MTase activity assays on substrates of varying lengths revealed that the rate of repair increases as the length of the substrate increases (114). MTase repair of the free O6mG base is 10⁻⁷-fold slower than O6mG in
duplex DNA \((114)\). Therefore, the overall structure of DNA must play a role in the recognition and repair process.

Ada activity was originally purified to physical homogeneity from an \textit{E. coli} mutant that constitutively expresses the adaptive response \((119;120)\). The molecular weight of the Ada protein is 39 kDa (354 aa), however, the protein is readily degraded in crude cell extracts to an 19 kDa fragment \((72;110;113;121;122)\) that retains the ability to repair \(\mathcal{O}^\bullet\text{mG} \) \((122)\) and \(\mathcal{O}^\bullet\text{mT} \) \((106)\). For this reason, the Ada protein was originally isolated as the C-terminal domain \((120;123)\), and was not purified as the entire Ada protein until it was expressed recombinantly \((15;124)\).

Endogenous fragmentation of Ada was shown to occur at two sites adjacent to Lys-178 and Lys-129 \((125)\). Interestingly, the two sites show considerable sequence homology to each other \((126)\) and to a sequence in UvrB \((126;127)\). The similarity between the primary structure of the Ada and UvrB sites suggested that a single protease could be responsible for the cleavage at their sites. Subsequently, purified OmpT protease was observed to cut the purified Ada and UvrB proteins into their expected fragment sizes \((77;128)\). Ada was cleaved primarily adjacent to Lys-178 \((129)\). A possible role for this proteolytic cleavage is discussed below.

Treatment of Ada with trypsin, chymotrypsin, subtilisin, and V8 protease results in cleavage within a stretch of 10 amino acids (aa 171 to 181) near the center of the protein \((125)\), consistent with the notion that Ada has two functional domains of similar size connected by a short hinge region. MTase assays demonstrated that the separate C-terminal and N-terminal fragments of Ada retained distinct transferase activities. More specifically, the C-terminal fragment of Ada repaired \(\mathcal{O}^\bullet\text{mG} \) and \(\mathcal{O}^\bullet\text{mT} \), while the N-terminal fragment repaired methylphosphotriesters \((125)\). In accordance with the presence of two distinct MTase domains, when purified Ada was incubated with excess \(\mathcal{O}^\bullet\text{mG} \) and phosphotriester substrates, both substrates were repaired and the protein contained approximately two methyl groups per Ada protein \((15)\).
The transfer of an alkyl group from DNA to Ada results in regeneration of the native DNA structure (73) and a methylated cysteine residue (72). Methylated cysteine residues are stable and therefore, in light that no detectable Ada demethylase activity has been detected (77;120), MTase repair results in a dead end protein product (113). This so called “suicide in activation” has been demonstrated numerous times both in vitro (72;113;122;123) and in vivo (67).

The cysteine residue that participates in MTase repair of O6mG and O4mT adducts is Cys-321 (123). The Cys residue is found in the Pro-Cys-His-Arg active site sequence found in all known O6mG MTases (130). Structural data of the C-terminal domain of Ada revealed that this residue is surprisingly buried (130). Therefore, it was hypothesized that must undergo a conformational change to expose the cysteine to act as a nucleophile (130). An alternative model suggested that the helix-turn-helix (HTH) motif in Ada binds to the DNA to invoke the alkylated base to flip out of the DNA (131). More recent structural data, for active and alkylated human Ada protein, implicate a HTH motif in DNA binding and the use of an “arginine finger” to extrude the alkylated base from the stacked DNA duplex (132).

The known eukaryotic MTases are homologous to the 19 kDa C-terminal domain of E. coli Ada (77;133) and have a high degree of structural similarity (134). Therefore, structure activity relationships for one homolog can assist understanding for the entire family of proteins. Additionally, a nucleotide flipping mechanism has great precedent and is observed in a number of DNA MTases and glycosylases (135). The model that includes repair of an extra-helical base is therefore more mechanistically established.

MTase repair of methylphosphotriesters is carried out in the N-terminal domain of Ada (Figure 4.6) (125). This chemical group can exist in one of two isomeric forms, R and S, and the observation that only one half of methylphosphotriesters are repaired suggested that Ada can only repair one of these forms (15). Quantitative studies subsequently revealed that only the S form is repaired (15;136). Docking models suggest that the ability of Ada to repair the S isomer relies on the accessibility of the methyl
group. The S isomer projects the methyl group into solution whereas the R isomer projects into the major groove of DNA and is not within reach of the active site of Ada (Figure 4.3) (137).

Structural data (138) revealed that the N-terminal domain of Ada is composed of a central β-sheet sandwiched between two α-helices. A tightly bound Zn$^{2+}$ (139) is coordinated by four cysteines (Cys-38, Cys-42, Cys-69, and Cys-72), from two adjacent loops, in a tetrahedral geometry. In addition to the role of Zn$^{2+}$ in activating the nucleophilicity of the acceptor cysteine (140), the metal acts to coordinate the structure of Ada (138). Repair of alkyl groups on the DNA backbone by the N-terminus of Ada converts the protein to a strong transcriptional activator of several genes (77) by eliciting a protein surface that enhances the affinity of the protein for the promoters of the ada-alkB operon and the alkA and aidB genes (Figure 4.5) (137;141-143). Cys-69 was previously assigned as the residue that accepted the alkyl group from the DNA backbone (125), however recent X-ray and NMR structural data (144) indicate that Cys-38 is the methyl acceptor. It has been suggested that methyl transfer to this ligand could cause a reorganization of the Zn$^{2+}$ coordination and subsequent conformational changes in the protein structure (145). However, extended x-ray absorption fine structure and x-ray absorption near edge structure data indicate that the integrity of the thiolate-zinc center is preserved after methytransfer (137). Additionally, solution structures revealed that the transferred methyl group can not make contacts with promoter DNA, suggesting that methyltransfer to Ada causes conformational changes that creates a high affinity DNA binding face (137). However, the precise mechanism by which alkylated Ada converts the protein into a transcriptional activator, however, is unknown (146).

Upon the repair of a methylphosphotriester in DNA by Ada, the N-terminus of the protein binds to a transcriptional regulatory element called the “Ada box” (77;114). The minimal element, identified by deletion studies and site directed mutagenesis, consists of the octanucleotide sequence 5′- AAAGCGCA -3′ just upstream of ada-alkB operon (147). The extended sequence 5′- AAANAAAGCGCA -3′ is found upstream of the alkA gene. These Ada boxes are just upstream of the putative RNA polymerase binding
site (143;148). Therefore, it appears as though alkylated Ada serves to initiate the adaptive response by facilitating the binding of RNA polymerase (77;104;114). The C-terminal domain of Ada has been implicated in the binding of the RNA polymerase (114).

As a consequence to the inability of MTases to act enzymatically on their substrates, the mutagenesis of alkylating agents accelerates after the available MTase molecules have conducted repair and are thus incapacitated. For this reason it is important that the adaptive response has a sensitive mechanism for signaling the cell to cope with excess alkylation damage. The level of methylphosphotriesters produced by alkylating agents is more than that for \( \overset{O}{m}G \) and \( \overset{O}{m}T \) (13), and they therefore do serve as an acute trigger for the adaptive response.

It is not completely understood how the adaptive response switches off upon the withdraw of alkylation pressure. One simple possibility for returning the concentration of repair proteins to their steady state levels is through the diluting effects of cell division (77;114). It has also been hypothesized that proteolysis of Ada could serve to terminate its ability to bind to promoters. As discussed above, the Ada protein is readily cleaved by the endogenous OmpT protease \textit{in vitro} (77;128). The N-terminal product was subsequently shown to bind to promoter sites but have no effect on gene expression (77;114). It follows then that Ada degradation could be responsible for down-regulation of the adaptive response \textit{in vivo} (149). Studies conducted in \textit{ompT} mutant strains, however, showed normal levels of Ada and kinetics of induction (129). Direct evidence for OmpT regulation of the adaptive response therefore is still lacking. Lastly, high concentrations of unmethylated Ada protein have been related to the inhibition of \textit{ada} gene activation by the methylated form (150). This observation provides yet another possibility for how cells return to constitutive levels of repair proteins.

4.6. OGT - \textit{IN VIVO AND IN VITRO} PROPERTIES
In the late 1980s, the isolation of *E. coli* mutants that lacked the function of Ada led to the discovery of a second protein with alkyltransferase activity \((149;151)\). The cellular level of Ogt was found to be independent of the adaptive response \((152)\), as cells that lacked *ada* showed a low threshold for MNNG tolerance that did not increase upon low dose exposure. Approximately 30 molecules of Ogt are present in the cell at all times \((153)\), and therefore provides a constant level of protection and can hold off a low level of damage while the cell adapts for high level repair.

Ogt is the protein product of the *ogt* gene which maps at 29.4 min on the *E. coli* chromosome \((154)\). The protein contains 171 amino acids and a molecular weight of 19 kDa \((153)\). The fact that Ogt is so similar in size and function to the C-terminal domain of Ada was likely to have been the source of some confusion \((82)\). Therefore, it wasn't until the use of mutant strains and recombinant expression techniques that trustworthy studies were possible \((104;153;155)\).

Ogt has been shown to remove methyl groups from O6mG and O4mT, but has no detectable activity for methylphosphotriesters \((149;151;153)\). Additionally, experiments conducted in *E. coli* cell extracts demonstrated that the concentrations of O6mG and O4mT substrates necessary to inactivate 50% of the Ada or Ogt methyltransferase activity were quite different \((156)\). The Ogt protein showed a high affinity for the O6mG substrate (8.1 nM) and an even higher affinity for the O4mT substrate (3 nM). Ada, on the other hand showed good affinity for O4mT (27.5 nM) but very high affinity for O6mG (1.25 nM). These preferences may be evidence that Ada and Ogt divide the responsibility of repairing alkylation damage.

The *ogt* gene has a nucleotide sequence that is distinct from that of *ada*, however the amino acid sequence of Ogt shows extensive regions of homology with the C-terminal domain of Ada \((104;123;153;157)\). The methyl accepting cysteine in Ogt (Cys-139) corresponds with Cys-321 in Ada, and is part of a region that is highly conserved in MTases from a number of species \((153)\). To date, no structural data for Ogt has been reported.
4.7. PROTECTION AGAINST ALKYLATION INDUCED TOXICITY BY ADA AND OGT

The cytotoxic effects of MNNG have been studied in *E. coli* strains with a variety of repair proficiencies to define the relative contributions to alkylation protection. In cells that were not pre-exposed to alkylating agent, strains of (wild type) wt and *ogf* backgrounds were similarly resistant whereas the *ada* strain was sensitive and experienced a high level of killing (152;154). These results suggested that even in conditions that are not optimal for inducing the adaptive response (66;74) Ada provides protection from alkylation damage. While both the *ada* and the *ogf* *ada* double mutant were sensitive to MNNG, Ogt over-expression was sufficient to confer resistance in these strains to wt level (154). These results indicate that O6mG is the major alkylation product involved in cell death.

The overlapping of O6mG repair by Ada and Ogt was also seen in MTase double mutants. Cells that lacked Ada and Ogt were more sensitive to alkylation treatment than cells lacking just Ogt (154). The *ada* *ogf* double mutant was reported to be more sensitive than the *ogf* mutant in one case (152) but not another (154). Interestingly, *ada* cells showed a low threshold to mutation induction owing to the presence of Ogt. On the other hand, mutation induction in the *ada* *ogf* double mutant was graded and linear (152). Clearly, the two proteins tag-team MTase repair of O6mG to evade the lethal effects of the lesion.

As described above, toxicity data indicate that O6mG achieves killing via MMR. Studies done by Karran and Marinus in the 1980s established that cells that are defective in the Dam methylase are sensitive to agents that produce O6mG such as MNNG (48). This effect is due to the O6mG lesions which evade MTase repair and are subsequently recognized by the mismatch repair proteins and inaccurately processed. Accordingly, while *dam* mutants are sensitive to alkylation pressure, *dam ada* double mutants are
more sensitive (52). Additionally, cells that lack Ada in addition to Dam and MutS experience more lethality from MNNG treatment (52).

In contrast with the view that O-alkyl repair by Ada provides direct protection against the cytotoxic effects of MNNG, some reports have posited that Ada provides indirect resistance to alkylation damage by inducing AlkA which then initiates BER on toxic lesions other than O6mG (77;87;158;159). However, Ada has been shown to have a direct effect of protecting cells against MNNG induced toxicity.

Comparison of alkA− and alkA− ada− strains showed that the double mutant was more sensitive to MNNG (152). This result would not be expected if the sole protection effect of Ada was through the adaptive response. Therefore, while it is possible Ada may provide some indirect protection to cells by AlkA induction, Ada also protects cells directly.

4.8. REPAIR OF O6-METHYL GUANINE BY NUCLEOTIDE EXCISION REPAIR

In addition to the ability of MMR and the Ada and Ogt MTases to act on O6mG and O4mT, nucleotide excision repair (NER) has also been shown to process these lesions (83;160-164). Removal of these lesions by NER was originally monitored in a range of host strains using monoclonal antibodies (83), and results suggested that NER provided a low level of protection against alkylation induced mutation. It was subsequently shown that O6mG is repaired by NER much less efficiently than other types of DNA damage (165), and therefore MTase repair is likely the primary repair mechanism for this lesion. On the other hand, repair of O4mT by NER appeared to be quite efficient and actually blocked by the presence of MTases (164;166). Therefore, repair of the O6mG and O4mT lesions are independently balanced by a number of distinct repair systems.
Figure 4.1. Sites of methylation on the DNA bases
Figure 4.2. Conversion of MNNG to the active methanediazonium ion by cysteine. Adapted from reference (146).

\[ \text{HOOC} \quad \text{H}_2\text{N} \quad \text{SH} \quad \text{Cysteine} \]

\[ + \]

\[ \text{H}_3\text{C} \quad \text{N} \quad \text{N} \quad \text{NO}_2 \quad \text{MNNG} \]

\[ \text{NH}_3 \quad \text{NH}_2 \text{NH} \quad \text{NO}_2 \quad \text{methanediazonium} \]
**Figure 4.3.** The R and S stereoisomers of methylphosphotriesters
Figure 4.4. Base pairing models for $\overset{\omega}{m}G:T$ and $\overset{\omega}{m}G:C$. Shown are the models for (A) $\overset{\omega}{m}G$ in the anti conformation bound with T, (B) $\overset{\omega}{m}G$ in the syn conformation bound with T, and (C) $\overset{\omega}{m}G$ in the syn conformation bound with C in a wobble arrangement.
Figure 4.5. Genetic structure and regulation of the adaptive response in *E. coli*

![Diagram](image)

- **Transcriptional Activation**
  - *ada*
  - *alkB*
  - *alkA*
  - *aidB*

- MTase
- Oxidative demethylase
- Glycosylase
- Dehydrogenase

**N terminal of Ada**
- Unmethylated

**Ada with methylated N-terminus**
Figure 4.6. Direct reversal of base damage by Ada

[Diagram showing the process of base damage reversal by Ada protein, involving S-methyl phosphotriester and phosphodiester bonds.]
Table 4.1. Relative amounts of DNA alkylated products following reaction of double stranded DNA with MMS (S_N2) and MNU (S_N1) alkylating agents. Adapted from Singer and Grunberger (13).

<table>
<thead>
<tr>
<th></th>
<th>MMS (S_N2)</th>
<th>MNU (S_N1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>in vitro</td>
<td>in vivo</td>
</tr>
<tr>
<td>Adenine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N1</td>
<td>3.8</td>
<td>1.3</td>
</tr>
<tr>
<td>N3</td>
<td>10.4</td>
<td>9.0</td>
</tr>
<tr>
<td>N7</td>
<td>1.8</td>
<td>1.7</td>
</tr>
<tr>
<td>Guanine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N3</td>
<td>0.6</td>
<td>0.8</td>
</tr>
<tr>
<td>N7</td>
<td>83.0</td>
<td>67.0</td>
</tr>
<tr>
<td>O⁺</td>
<td>0.3</td>
<td>6.3</td>
</tr>
<tr>
<td>Thymine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O⁺</td>
<td>ND</td>
<td>0.11</td>
</tr>
<tr>
<td>N3</td>
<td>0.08</td>
<td>0.3</td>
</tr>
<tr>
<td>O⁺</td>
<td>ND</td>
<td>0.4</td>
</tr>
<tr>
<td>Cytosine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O⁺</td>
<td>ND</td>
<td>0.1</td>
</tr>
<tr>
<td>N3</td>
<td>&lt;1</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Phosphotriester</td>
<td>0.8</td>
<td>16</td>
</tr>
</tbody>
</table>
4.9. REFERENCES


182


CHAPTER 5

Effect of Mismatch Repair Proteins on Methyltransferase Repair of $\textit{O}^6$-Methylguanine \textit{in Vivo}
5.1. ABSTRACT

DNA repair is essential for fighting the adverse effects of base damage to the cell. One example of base damage, 6-methylguanine (6mG), stably pairs with thymine during replication and thereby creates a promutagenic 6mG:T mismatch. This mismatch has also been linked with cellular toxicity. Therefore, in the absence of repair, 6mG:T mismatches can kill the cell, or, upon the next round of replication, result in G:C → A:T transition mutations. The Ada and Ogt methyltransferase (MTase) proteins can both directly reverse the 6mG base damage to yield guanine. When the 6mG is in a G:C pair, this reaction poses little problem for the cell. However, if the 6mG has already been replicated, and an 6mG:T pair exists, the consequences for the cell are much more complicated. If the 6mG:T pair is biochemically converted to a G:T pair, the G:T mispair must still be corrected to G:C in order to avoid mutation. Two pathways in *E. coli* can convert G:T mispairs to the native G:C base pairing. These are the Methyl-directed Mismatch Repair (MMR) and the Very Short Patch Repair (VSPR) pathways. The use by the cell of VSPR has distinct advantages. In this work, we hypothesized that *E. coli* may coordinate MTase repair with VSPR as an efficient means to restore harmful 6mG:T mismatches to normal G:C base pairs. To test this hypothesis, we analyzed the relative efficiencies of MTase repair of 6mG (by Ada, Ogt, and both) in cells that lacked functional mismatch repair proteins MutS, MutL, MutH, or Vsr. We found that MTase repair of 6mG is positively influenced by the MutS, MutL, and Vsr proteins. In contrast, cells that lacked MutH showed wild type levels of MTase repair. These results are consistent with the notion that VSPR positively influences MTase repair of 6mG:T mismatches to avoid the detrimental effects of this mispairing.
5.2. INTRODUCTION

Alkylating agents produce a variety of DNA lesions by reacting at a variety of positions on the DNA bases. The cellular impact of these lesions will depend on their mutagenic and cytotoxic properties, as well as the extent to which they are repaired. One particular type of DNA alkylation damage, $\sigma^\theta$-methylguanine ($\sigma^\theta$mG), has been given much attention due to its mutagenicity (1-4) and cytotoxicity (5-7). Extensive in vivo mutagenesis studies done in our laboratory established that $\sigma^\theta$mG codes as an adenine during replication in *E. coli* nearly 100% of the time (8). As a result, this lesion is responsible for G:C $\rightarrow$ A:T transition mutations following a second round of replication. The $\sigma^\theta$mG:T base pair has also been shown to be a substrate for methyl-directed mismatch repair (MMR) (9-11). It is hypothesized that the toxicity of $\sigma^\theta$mG results from reiterative and futile attempts to replace the mismatched T with a base that does not recreate a substrate for another round of MMR (12). This process can lead to persistent repair gaps and, ultimately, toxic double strand breaks. Therefore, repairing $\sigma^\theta$mG is paramount to avoiding the lethality of alkylating agents.

In *E. coli*, repair of $\sigma^\theta$mG is primarily the responsibility of two methyltransferase (MTase) proteins Ada and Ogt (6). Both of these proteins can repair $\sigma^\theta$mG directly by transferring the methyl group to an active site cysteine residue (13-16). This methyltransfer is not reversible and therefore inactivates the protein (17). For this reason, the MTase proteins are not true enzymes and are commonly referred to as “suicide proteins”. In the case of the Ada protein, repair of methylphosphotriesters, which are normally formed alongside $\sigma^\theta$mG, leads to upregulation of Ada through a process called the adaptive response (18). This response enables the repair of high levels $\sigma^\theta$mG associated with alkylation pressure (14). In the case where MTase repair of $\sigma^\theta$mG follows one round of replication, a mismatched G:T base pair results. To prevent mutation, this mismatch must be further processed before replication.

G:T mismatches are a substrate for two different types of Mismatch Repair (MR), Methyl-directed Mismatch Repair (MMR) (19) and Very Short Patch Repair (VSPR)
Both systems are stimulated by the activities of the MutS and MutL mismatch recognition proteins, although the downstream mechanisms of repair initiation and completion are quite different.

MMR is only active for a short time period following replication, owing to the hemimethylated state of Dam d(GATC) sites. Once the Dam methylase has had enough time to methylate the newly synthesized strand, MutH can no longer initiate repair by incising the DNA backbone. Therefore, MMR is time limited. Additionally, incision by MutH at an adjacent d(GATC) site initiates repair as far as two kilobases away from the mismatch. Following incision by MutH, strand removal is accomplished through the concerted efforts of many proteins including DNA helicase II, single stranded DNA binding protein, and Exo I, Exo VII, Exo X, or Rec J. DNA Polymerase III must then resynthesize the DNA strand from the d(GATC) site to the mismatch. DNA Ligase completes the repair event by sealing the DNA nick. In total, repairing the G:T mismatch via MMR is of significant metabolic cost to the cell.

VSPR of G:T mismatches is initiated by the Vsr endonuclease, which in all cases creates a single stranded nick on the 5' side of the mismatched T. Pol I then removes and replaces a small number of bases (<10) 3' to the nick utilizing its 5' to 3' exonuclease and polymerase activities. DNA Ligase then seals the nick to complete the restoration of the native G:C pairing. Given that the repair patch created by VSPR is much smaller than that for MMR, the VSPR pathway is significantly more efficient.

The primary purpose of VSPR is to repair G:T mismatches that arise through the deamination of the Dcm product, 5-methylcytosine. Dcm methylates the 5 position of the second cytosine of the sequence 5'-CCAGG-3' (28). Therefore Vsr most commonly initiates repair on G:T mismatches in the 5'-CTAGG-3' sequence context (where the T is mismatched with a G). However, Vsr has been shown to have significant endonuclease activity for G:T mismatches in a variety of sequence contexts (29;30). Therefore, VSPR could play a role in repairing G:T mismatches that arise through MTase repair of O6mG:T base pairs, as well.
Upon inspection, it appears the most efficient route to repairing $\text{O}^6\text{mG}:T$ base pairs is through the methytransferase repair of $\text{O}^6\text{mG}$ followed by VSPR of the resulting G:T mismatch. Additionally, it would seem that the proposed competition between MutS and Ada for the $\text{O}^6\text{mG}:T$ pair would be inefficient, and perhaps unlikely, for cell survival. We therefore hypothesized that the MTases might act in conjunction with proteins in the VSPR pathway, to restore cooperatively $\text{O}^6\text{mG}:T$ mismatches to G:C base pairs.

To test our hypothesis that MTase repair can be positively affected by MR proteins, we utilized a M13 based *in vivo* MTase assay in a number of cell strains. We constructed a matrix of mutant *E. coli* that are defective in MTase activities (Ada, Ogt, or both) and a single mismatch repair protein (MutS, MutL, Vsr, and MutH). In total 20 cell strains were assayed for their relative MTase repair efficiency, allowing for a comprehensive view of how MMR and VSPR proteins affect the repair of $\text{O}^6\text{mG}$ by MTase proteins. Remarkably, both MutS and MutL were observed to play a cooperative role with Ada and Ogt. Additionally, while cells deficient in MutH showed no difference to wt, cells lacking Vsr showed a significant decrease in MTase repair efficiency. These data represent the first direct examination of how MTase repair is affected by MR proteins, and assist in our understanding of how repair systems interact *in vivo*.
5.3. EXPERIMENTAL PROCEDURES

Preparation of E. coli Genomic DNA. Preparation of E. coli genomic DNA was carried out essentially as described in Current Protocols (31). Briefly, 1 mL of cells from an overnight culture was spun down and the resulting cell pellet was resuspended in 570 µL TE 8.0 (10 mM Tris, 1 mM EDTA). After the addition of 30 µL of 10% SDS and 1.8 units of Proteinase K (Roche), the solution was incubated at 37 °C for 1 h. This solution was then supplemented with 100 µL of 5 M NaCl, 80 µL of a 10% CTAC/0.7 M NaCl solution, and incubated at 65 °C for 10 minutes. This solution was extracted twice from 750 µL of a 25:24:1 ratio of phenol to chloroform to isoamyl alcohol (Invitrogen). DNA was precipitated from the final aqueous layer with 420 µL of isopropanol, and the DNA pellet was washed with 250 µL of 70% Ethanol. The resulting DNA was dried in vacuo for 5 minutes, resuspended in 100 µL TE, and quantified using UV spectrometry.

Preparation of Recombination Substrates. The Tn10 marker, which codes for TetR_N and TetR_C, was amplified from GM5555 (obtained from M. Marinus, U. Mass. Worcester) genomic DNA using primers that contained flanking sequences to match the target gene (Table 5.1). The PCR reaction was carried out according to Stratagene’s Opti-prime PCR Optimization Kit. The combination of Taq polymerase (NEB), buffering condition #11, and 2% DMSO as a reaction adjunct provided the best results. The thermocycling parameters were as follows: 94 °C for 90 sec, then thirty rounds of 95 °C for 30 sec, 56 °C for 30 sec, 72 °C for 60 sec, and finally 10 min at 72 °C. The 2 kb product, identified by separating a fraction of the reaction products on a 1% agarose gel, was purified using the QIAquick PCR Purification kit.

Transformation of Recombination Substrates and Isolation of Recombinants. KM22 cells (obtained from M. Marinus, U. Mass Worcester) (32) were prepared for transformation by diluting 100 µL of an overnight culture in 10 mL LB plus 1 mM IPTG, growing (roller drum, 37 °C) until OD<sub>600</sub> = 0.6, washing the cells twice with 15 mL cold water, and resuspending the final cell pellet in 75 µL 1 mM MOPS/20% glycerol. Approximately 100 µL of cells and 500 ng of recombination substrate was combined in a chilled
electroporation cuvette and a 2.5 kV pulse at 129 Ohms was applied using a Electro Cell Manipulator 600 electroporation system (Bantex). Immediately after electroporation, cells were transferred to a culture tube containing 3 mL LB and 1 mM IPTG and grown at 37 °C as above. After three hours, cells were harvested by centrifugation, resuspended in 100 μL LB, plated on a LB plate containing 4 μg/mL Tetracycline, and incubated over night at 37 °C. The next day, successful transductants were picked and grown over night in liquid culture with 12 μg/mL Tetracycline. Cultures were then analyzed individually for the correct genotype by isolating their genomic DNA and performing PCR as described above using primers to amplify the gene of interest (Table 5.2).

**Preparation of Knock Out P1vir Lysates (aka “Up Infection”).** A 1:100 dilution of KM22 donor cells, which were grown over night in LB containing 12 μg/mL Tetracycline and 1 mM IPTG, was grown to mid log phase (OD<sub>600</sub>=0.5) in 5 mL LB supplemented with 5 mM CaCl₂. Serial dilutions of P1vir (gift of Walker Lab, MIT), ranging from 10<sup>9</sup> to 10<sup>-5</sup> of stock (10<sup>9</sup> pfu/μL), were prepared in LB. P1 infections (six total) were prepared to contain 100 μL of cells and 10 μL of one of the five P1vir dilutions. A control which had no virus was done in parallel for comparison. These mixtures were incubated in a capped 1.5 mL tube at 37 °C for 20 minutes. Infections were then mixed with 3.5 mL of R-top agar (51 °C) (per L: 10g Bacto tryptone, 1g Bacto yeast extract, 8g Difco agar, 8g NaCl, add 2 mL 1 M CaCl₂ and 5 mL 20% glucose after autoclaving) and immediately spread on R-plates (as R-top agar but 12 g Difco agar). The agar layer solidified after 10 minutes and the plates were incubated at 37 °C over night. The following day, virus was harvested from the plate displaying the most discernable plaques. A spatula was used to scrape and transfer the agar layer to a 15 mL centrifuge tube, and two 1 mL volumes of LB were used to collect all virus from the plate. The tube containing the agar was supplemented with 300 μL chloroform, vortexed on high for 30 seconds to elute the virus, and spun at 7,000 xg for 10 minutes to pellet the insoluble material. Approximately 1.75 mL of P1vir was transferred to a sterile 2 mL glass vial, supplemented with 50 μL chloroform, and stored at 4 °C. Serial dilutions of this virus (10<sup>9</sup> to 10<sup>-5</sup>) were used to infect recipient cells.
**Generation of Knock Out Transductants (aka "Down Infection").** A 1:100 dilution of recipient cells was grown to mid log phase (OD$_{600}$=0.5) in 5 mL LB, supplemented with 500 μL of a solution containing 1 M MgCl$_2$ and 50 mM CaCl$_2$, vortexed, and incubated at room temperature for 20 minutes. For infection, 100 μL of cells and 10 μL of a viral dilution were incubated at 37 °C for 20 minutes in a capped 1.5 mL tube. A control which had no virus was done in parallel for comparison. An equal volume of 250 mM sodium citrate was added to stop the infection, and cells were allowed to recover at 37 °C for 3 hours. The entire contents of each tube (200 μL) was spread on a LB plate containing 4 μg/mL Tetracycline and incubated over night at 37 °C. The next day, successful transductants were picked and grown over night in liquid culture with 12 μg/mL Tetracycline. The genomic DNA from individual cultures was then analyzed for the correct genotype using PCR, agarose gel separation, and sequencing techniques as described above. One stock, possessing the correct genotype, was chosen for all subsequent experiments.

**Preparation of M13mp7L2 (M13) DNA.** A 5 mL culture of NR9050 cells grown over night in LB was diluted 1:5 in 5 mL of 2xYT and grown (roller drum, 37 °C) for 1.5 h. Once the cells reached mid-log phase, 2 mL were used to inoculate 1 L of 2xYT, in a 2 L baffled flask, which then grew (275 rpm, 37 °C) for 2.5 h. A 2 mL volume of saturated M13mp7L2 progeny phage (gift of Sarah Delaney) was added to the cells which then grew for an additional 9 h. To isolate progeny phage, the mixture was incubated at 0 °C for 10 min, the cells were pelleted by centrifugation (9.5 krpm, 10 min), and the supernatant was decanted into sterile tubes (6 tubes, 165 mL/tube). The phage in each supernatant was precipitated by adding ¼ volume (42 mL) of 20% PEG/2.5 M NaCl and incubating over night at 4 °C. Precipitated phage was collected by centrifugation (9.5 krpm, 10 min), resuspended in a total volume of 19 mL TE (10 mM Tris-HCl, 1 mM EDTA (pH 8.0)), split into thirds, and extracted 5 times with 3 mL phenol:chloroform:isoamyl alcohol (25:24:1 v/v/v, Invitrogen). The aqueous phase (~18 mL) was applied to a hydroxylapatite column (1 g DNA grade resin (BioRad)) conditioned with 5 mL TE, and eluted using moderate air pressure. The column was washed with 5 mL TE and M13 DNA was eluted using 20 mL total of 140 mM
potassium phosphate buffer. Each of the 1 mL fractions was analyzed for DNA content by separating 10 μL on a 1% agarose gel in TAE, staining with ethidium bromide, and visualizing the gel over a transilluminator. The 6 fractions containing the most amount of M13 DNA were concentrated and dialyzed through 3 Centricon-100 columns (Amicon) by centrifugation followed by 3 x 2 mL TE washes. The M13 DNA was collected in a final volume of ~250 μL TE. The concentration of genome was quantified by spectrophotometry using the conversion factor 14 pmol/mL = 1 A_{260} unit. Typical yields of M13 DNA were approximately 600 pmoles/L of culture.

Construction of O6mG - M13 Viral Genome. Genomes were constructed on a 24 pmole scale. Single-stranded M13 DNA (2.4 pmole/μL) was linearized by cleavage with EcoRI (70 U, NEB) for 8 h at 23 °C in a total of 50 μL of the supplier’s buffering conditions. Cleavage occurs at a hairpin containing a single EcoRI site (33). The linearized M13 DNA was annealed with 1.25 equivalents of each scaffold 5’- GGT CTT CCA CTG AAT CAT GGT CAT AGC -3’ and 5’- AAA ACG ACG GCC AGT GAA TTG GAC GC -3’ (sequences that are partially complementary to the 5’ and 3’ sides of the insert and the genomic DNA termini) by heating the mixture to 50 °C for 5 min and cooling linearly to 0 °C over 50 min (33). The O6mG lesion containing insert, 5’- GAA GAC CTX GGC GTC C -3’ was phosphorylated on the 5’-terminus with T4 PNK (10 U, NEB) using the supplier’s buffering conditions supplemented with an excess of ATP (Roche). The phosphorylated insert (30 pmoles), the annealed scaffold mixture (24 pmol), 1 μL of 100 mM ATP (Roche), and 2 μL of T4 DNA Ligase (800 U, NEB) were mixed in a final reaction volume of 55 μL and incubated at 16 °C for 20 h to recircularize the genome (33-35). The scaffold DNA was removed using the exonuclease activity of T4 DNA polymerase (36,37) by incubating the mixture at 37 °C for 6 h after the addition of 3 μL enzyme (10.5 U, NEB). The genome construct mixture was brought to a final volume of 100 μL with water, extracted twice with 100 μL phenol:chloroform:isoamyl alcohol (25:24:1 v/v/v, Invitrogen), and desalted using G-50 Sephadex Quick Spin Columns (Roche).
Quantification of Viral Genome. Quantification of genome constructions was accomplished by annealing 15 pmol of the 5′[32P]-labeled 30 mer probe 5′- TCC CAG TCA CGA CGT TGT AAA ACG ACG GCC -3′ to 5 μL of genome construct and a range (3.9, 7.8, 15.6, 31.2, 125, 250, and 500 fmoles) of wt M13 DNA standards. The probe was designed to anneal in a region of the M13 remote from the cloning/insert site. The annealing solutions consisted of 100 mM NaCl, 4.2% Ficoll (Amersham), 0.042% bromophenol blue (ICN), and 0.042% xylene cyanol FF (ICN) in a volume of 15.5 μL. After incubating at RT for 1 h, the genome probe mixtures were run on a 1% agarose gel in 1x TAE buffer for 4 h at 100 V, after which the free probe front was excised. The gel was run for an additional 3 h and then transferred onto a glass plate and dried under a box fan for 36 h. The amount of circular genome was quantified by using phosphorimagery to compare its signal intensity to the signal intensities for the standards. Typical yields were in the range of 5% (1.2 pmoles), which was enough for hundreds of electroporations.

Determination of Relative Rates of O6mG Repair in Strains of Differing Genotype. To prepare cells for electroporation, an over night grown in LB (containing 50 μg/mL Kanamycin, 12 μg/mL Chloroamphenicol, or 12 μg/mL Tetracycline where applicable) was diluted 1:50 in 10 mL LB (4 replicates), and grown to mid-log phase (OD600 = 0.5). Cells were transferred into 15 mL tubes, harvested by centrifugation (6,000 xg, 10 min), washed twice with 15 mL 4 °C water, and resuspended in 75 μL 10% glycerol. Chilled electrocuvettes were then supplemented with 100 μL of competent cells (~2 x 10⁹ cells) and ~3.5 fmoles of the O6mG-M13 viral genome for transformation.

The cell/genome mixtures were electroporated (2.5 kV, 129 ohms in a 0.2-cm cuvette) using a Electro Cell Manipulator 600 electroporation system (Bantex), and transferred to 10 mL LB. Fractions of this dilution (20, 75, and 200 μL) were transferred to a tube containing 3.75 mL overlay mix, which was then vortexed and immediately spread over the surface of a B-broth plate (per L; 10g tryptone, 8g NaCl, 20g bacto agar).
Overlay mix was prepared by adding 375 μL of a solution containing 300 μL plating bacteria (NR9050 over night diluted 1:5 in 2 x YT and grown on a roller drum at 37 °C for 1.5 h), 25 μL 1% thiamine, 10 μL of 100 mM IPTG, and 40 μL of 40 mg/mL X-gal in DMF to 3 mL B-broth soft agar (10 g/L tryptone, 8 g/L NaCl, and 6 g/L) Bacto agar just before use.

Once completing all the electroporations and platings, the soft agar mix on the top of each plate was allowed to cool for an additional 10 min on the bench. Plates were then put at 37 °C, up side down, over night. After incubation for 16-20 h, the number of light blue and dark blue plaques were counted and the ratio of darks to total was used to calculate percent repair. Of the three dilutions plated for each electroporation, the one which yielded the highest number of discernable plaques was counted (approximately 100 to 800 total plaques (Average = 421).
5.4. RESULTS

*Genotypic Characterization of E. coli Strains.* Before constructing the MMR protein mutants, the parental C215 (wt), C216 (ogf::kan), C217 (ada-,alkB::Cam), and C218 (ogf::kan, ada::Cam) strains (8;38-40) (as FC215 derivatives) were characterized genotypically. PCR was conducted on genomic DNA using primer sets (Table 5.2) that annealed to ogf or ada gene sequences outside of the locations altered by antibiotic resistance marker insertion (Figures 5.1 and 5.2). The PCR products obtained were separated via 1% agarose gel in TAE and compared against their expected fragment sizes (Table 5.3). In all cases the size of the fragments obtained by PCR matched our expectations.

Resulting PCR fragments were also sequenced using the 5' primer sequence utilized for amplification. Open reading frames within the sequencing data were identified using NEB cutter (41) and analyzed using BLAST software (42). BLAST results confirmed the presence of native gene sequences in the appropriate strains. For the ogf strains, BLAST results yielded conserved domains matching the aminoglycoside 3'-phosphotransferase family, a member of which confers resistance to Kanamycin. For the ada strains, BLAST results yielded conserved domains matching the chloramphenicol acetyltransferase protein.

Genotypic characterization for the mutH, mutL-, mutS, and vsr strains was conducted as above. PCR was conducted on genomic DNA using primer sets (Table 5.2) that annealed to mutH, mutL-, mutS, and vsr gene sequences outside of the locations altered by antibiotic resistance marker insertion (Figures 5.3, 5.4, 5.5 and 5.6). In each case, the size of the PCR product (Table 5.3) was easily matched to that of the native gene (analyzed in parallel as a negative control) or the insertion deletion mutant. PCR products from successful recombinants were analyzed, as above, by sequencing and BLAST analysis. Results indicated that, as expected, the majority of the target gene was indeed replaced by conserved domains matching the TetR_N and TetR_C proteins.
Rates of \( \overset{\circ}{m}G \) Methyltransferase Repair. The twenty cell strains assayed for their relative rates of \( \overset{\circ}{m}G \) repair showed a wide range of MTase activity (Tables 5.4, 5.5, 5.6, and 5.7). Firstly, the data from the C215, C216, C217, and C218 parental strains (Figure 5.7, black columns) demonstrates good correlation between percent repair of \( \overset{\circ}{m}G \) and MTase status of the cell. As expected, the C215 wt strain that expresses both functional methyltransferases shows a high amount of lesion repair (96 ± 1%). The C216 strain which lacks Ogt protein shows significantly less MTase ability (69 ± 3%), while the C217 strain which lacks Ada protein shows near wt levels of MTase repair (95 ± 1%). The asymmetry in the results for these single knock outs were not surprising as the normal cell contains roughly thirty Ogt molecules (15) and one molecule of Ada (43). Therefore, cells which lack the predominant MTase under normal conditions, Ogt, are expected to display the more significant decrease in MTase ability. Accordingly, the C218 double mutant shows very poor repair of \( \overset{\circ}{m}G \) (6 ± 1%).

The parental strains bearing the mutH mutation (Figure 5.7, light grey columns) showed MTase repair results nearly identical to those for the parental strains (Figure 5.7, black columns). The percent repair data for C215H, C216H, C217H, and C218H strains were 97 ± 1, 70 ± 2, 96 ± 2, and 6 ± 1, respectively. These data indicate that the MutH protein does not affect the MTase repair of \( \overset{\circ}{m}G \) by Ada or Ogt.

In contrast, mutS, and mutL− strains displayed levels of repair significantly lower than those for the C215, C216, and C217 parental strains. The effect of knocking out mutL− from the C215, C216, and C217 parental strains (Figure 5.7, grid columns) resulted in a decrease in repair to ~70% that of wild type (70, 75, and 66% of wt, respectively). Though there was variation between the mutL− mutants, there appeared to be no significant differences between them. On the other hand, results for the mutS knock outs showed more variation between the MTase backgrounds. Mutation of mutS (Figure 5.7, dark gray columns) in the C215 background caused a modest decrease in percent repair to 71% that for the parent strain. A more dramatic decrease was observed by mutating mutS in the C216 background. Here the mutS mutant showed repair 62% of
that for the parental strain. The largest effect of mutating this protein was observed in the C217 background where mutating mutS caused a drop in repair to only 41% that for the parent cells. Cells lacking the functions of MutL and MutS clearly are less efficient at repairing O6mG.

Remarkably, the vsr strains (Figure 5.7, white columns) also displayed less efficient repair of O6mG as compared to their parent lines. Knocking out vsr from the C215 background conferred a small decrease in repair to 84% that for the parent. Absence of Vsr from the MTase single knockouts however dropped the levels of O6mG repair more significantly. Mutating vsr in the C216 background dropped MTase repair to 59% that for the parent. The level of MTase repair in the vsr mutant form of C217 was 75% of that for the parent.

The level of O6mG repair for the C218 derivatives was, in all but one case, consistently low. The exception was the C218L strain that displayed 44 ± 4% repair of O6mG. If reproducible, this result would be interesting from the standpoint of how the absence of MutL protein confers a dramatic increase in repair of O6mG in cells lacking both MTase proteins. An experiment to confirm of this result is underway.
PCR mediated gene replacement (32) was used to construct sixteen mutant forms of *E. coli* that contained differing MTase capacities and were deficient in either *mutH*, *mutL*, *mutS*, or *vsr*. We chose to construct these mutants such that the vast majority of the target gene would be replaced by the marker for tetracycline resistance. We designed the recombination sequences such that approximately 150 bases on each end of the target gene would be preserved (Figures 5.3 to 5.5, from the end of the gene to the opposite end of recombination sequence). The genetic information central to the recombination sequences, however, would be replaced by Tet^R^ upon recombination with the targeted substrate generated by PCR (Figure 5.8) (32). Recombination, between the target gene in the cell and the PCR generated substrate, is carried out by KM22 cells that have *recBCD* replaced with *λ red* (32). These cells facilitate the production of mutants because no prior cloning of the target gene is necessary, and the frequency of recombination is 10 to 100-fold higher than in *rec* proficient strains (44). Employing this method allowed us to generate our mutant KM22 donor cells relatively quickly. Subsequent transfer our alleles into recipient C215, C216, C217, and C218 strains was carried out by generalized P1vir techniques as described in the experimental procedures section.

The assay we used to measure the relative efficiencies of MTase repair has been described previously (8;33). Upon transformation of the ϕmG-M13 genome into *E. coli*, the ability of the cell to repair the lesion prior to replication determines if the codon sequence containing the lesion encodes for a Trp or a stop codon (Figure 5.9). If repair of ϕmG does not precede replication, the replication machinery interprets the highly mutagenic lesion as an A within a TAG stop codon. Subsequently, transcription of the α-fragment of β-Galactosidase is terminated. If, however, ϕmG is repaired before replication, the codon in which it is located, TGG (Trp), allows read through and the α-fragment of β-Galactosidase is transcribed and translated. Subsequently, this fragment associates with the ω-fragment of the protein, provided by the host cell, and gives rise to active β-Galactosidase. This protein cleaves the chromogenic substrate X-gal to yield a
The efficiency to which an O6mG lesion is repaired in *E. coli* depends on the number of MTase molecules present in the cell and the affinities of those MTases for the lesion. During exponential growth, approximately thirty molecules of Ogpt (15) and one molecule of Ada (43) are present. On this basis, one would expect to observe a significantly lower level of repair in the C216 background (ogt-::kan) as compared to the C217 background (ada-::Cam). The affinity of Ada for O6mG, however, is greater than that for the Ogpt protein (45) and therefore the difference in repair observed for the ada and ogt mutants reflects a multitude of parameters. Therefore, the mutant strains were grouped according to the MTase background of the parent cell.

Interestingly, we observed data that were consistent with a small amount of O6mG repair in the C218 (ada-, ogt) double mutant strain bearing no MTase activity. There are two potential ways to rationalize this result. First, nucleotide excision repair (NER) has been observed to repair these lesions (38,46-50) especially in the absence of MTase activity. Alternatively, the repair signal observed could be due to a small population of M13 genome containing a G instead of O6mG. The synthetic route used to produce the 16-mer ODN containing the O6mG lesions utilizes a deprotection step that is known to yield small amounts of the unmodified oligonucleotide (51). A small amount of viral genome containing G at the lesion site would manifest itself as “repair” by yielding a dark blue plaque. Mass spectrometry analysis of the lesion containing 16-mer confirmed the presence (~5%) of a species consistent with G at the lesion site. Therefore, a small amount of systematic error is present in the repair percentages calculated, and results in slightly elevated numbers for percent repair of O6mG.

Strains lacking *mutS* displayed a decrease in MTase repair. These results were interesting as MutS binds to O6mG base pairs (10) and can presumably shield the lesion from MTase repair. The mechanism by which the presence of MutS confers increased MTase repair of O6mG therefore likely involves (1) identification of the lesion by MutS
followed by (2) subsequent dissociation of MutS and (3) signaling the presence of a mismatch to downstream repair proteins. The MutL protein has been shown to assist MutS in mismatch presentation by stimulating the formation of an α-shaped DNA loop containing the mismatch (52). The observation that mutL mutants display decreases in MTase repair similar to the decreases for mutS mutants is in line with previous observations that mutating either gene results in an identical defect in mismatch recognition (53). It therefore appears that a novel role of the MutS and MutL mismatch recognition proteins in assisting MTase repair has been discovered. More studies are necessary to examine the exact mechanism of this effect.

It could be the case, for example, that MutS and MutL recruit Vsr to the 5hmG lesion site, and Vsr is responsible for stimulating MTase repair (Figure 5.10). Consistent with this idea, strains that lack vsr showed a decrease in MTase repair (Figure 5.7, white columns). A possible role for Vsr dependent stimulation of MTases could be to increase the chances that Vsr initiates repair on the G:T mismatches that result from MTase repair of a 5hmG:T base pair. Subsequently, VSPR, and not MMR, restores mismatch to the native G:C pairing with high efficiency. This model seems particularly attractive from multiple perspectives. First, it is the most energetically favorable mechanism by which to process 5hmG:T mismatches to G:C base pairs. Second, it might be that the cell has no other choice. Given that 5hmG:T mismatches arise through the replication of the lesion containing strand, MMR would be targeted to the strand opposite the lesion and reiteratively regenerate a MMR substrate. Eventually, if the cell does not die, the DNA becomes fully methylated at the d(GATC) sites and MMR becomes inactive. At this point, correct repair of 5hmG:T mismatches depends on MTase repair followed by VSPR. Third, it is likely that the efficiency of these two repair systems has helped define cellular survival. If cells have been continually faced with this scenario, it seems plausible that cooperation between the MTase and VSPR proteins would evolve over time to produce a cell that is more fit.
5.6. FUTURE DIRECTIONS

Bacterial Two-Hybrid Studies between MTases and VSPR Proteins

A question raised from the data in Figure 5.7 is how the presence of MutS, MutL, and Vsr confers increased repair of $\eth mG$. One may suggest that MutS and MutL assist MTase proteins in identification of $\eth mG$ lesions by presenting the lesion for repair in the loop shaped structure reported by Allen (52). If this were the case, MutS and MutL would stimulate (1) MMR by identifying mismatches on DNA and subsequently activating the latent endonuclease activity of MutH at an unmethylated d(GATC) site adjacent (by up to thousands of base pairs) to the mismatch (22), (2) VSPR by identifying G:T mismatches on DNA and facilitating their recognition by Vsr for incision next to the mismatched T (54), and (3) MTase repair by identifying $\eth mG$ mispairs and facilitating their recognition by a MTase for direct reversal of the damage. In the first scenario, the endonuclease activity of MutH is activated directly by MutS and MutL. In contrast, the second two situations involve a mechanism whereby MutS and MutL assist the efficiency, but not activity, of the downstream repair protein, and are consistent with the observations that Vsr and the MTases are active in the absence of MutS and MutL. As an alternative to presenting the lesion for recognition, MutS and MutL may assist the efficiency of Vsr and/or the MTases directly through protein-protein interactions to aid localization of these proteins at damaged DNA substrates.

One technology to study the extent to which proteins directly interact with each other in vivo is the two-hybrid system (55). One such bacterial two-hybrid system utilizes the reconstitution of the catalytic domain of Bordetella pertussis adenylate cyclase (Figure 5.11) (56). In this system, the two proteins of interest are genetically fused to the complementary T25 and T18 enzyme fragments that associate to produce cAMP. This molecule then triggers transcriptional activation of genes such as lacZ which are involved in carbohydrate fermentation and produce a characteristic phenotype. In the absence of binding affinity between the two proteins of interest the T25 and T18 fragments do not associate for cAMP synthesis. However, as the binding affinity between the two proteins...
of interest increases, the amount of cAMP produced increases, as well. Therefore, this system provides a tool to study protein-protein interactions qualitatively.

The Cupples laboratory has successfully utilized this system to study the extent to which the proteins involved in VSPR interact with each other (57;58). Using β-galactosidase units as a measure, it was determined that MutL interacts with the N-terminus of Vsr. It was also determined that the interaction between MutL and Vsr inhibits the ability of MutL to dimerize, to interact with MutS and MutH, and to mediate a previously unknown interaction between MutS and MutH.

In efforts to understand the extent to which these VSPR proteins interact with the Ada and Ogt MTases, we have established a collaborative project with the Cupples Lab. The T18 and T25 fusion vectors, each containing Ada and Ogt, will be constructed and investigated as previously described (57;58) to address the possibility that these proteins interact with MutS, MutL, and Vsr in vivo.

Toxicity Studies on Cells Proficient and Deficient in Vsr Protein

As previously described, O6mG is cytotoxic to E. coli in a manner that is dependent on MR. A number of studies have been done to address the contribution of the Dam, MutS, MutL, and MutH proteins in alleviating the lethality of the lesion. However no studies have reported the effect of Vsr on cell survival, and one could imaging multiple outcomes. First, it could be that Vsr contributes to the lethality of O6mG:T mismatches. Vsr may incise 5’ to the mismatched T to initiate a round of VSPR that would give rise to the original O6mG:T mismatch. Here, reiterative rounds of attempted repair could continue indefinitely because VSPR does not become inactive like MMR does once the DNA is fully Dam methylated. This futile cycling could in turn lead to death of the cell through some unknown mechanism. A second, and more likely, scenario involves Vsr combating the cytotoxic properties of O6mG. Given Vsr increases the efficiency of O6mG MTase repair, one may hypothesize the presence of Vsr would
correlate with more efficient repair of the lethal \( \text{6mG} \) lesion. A third possibility is that the presence or absence of Vsr has no effect on the toxicity of \( \text{6mG} \).

As described in the discussion section, once MMR is inactive, MTase repair of \( \text{6mG:T} \) to G:T is only useful (in avoiding a mutation during the next round of replication) if the subsequent G:T can be corrected to G:C by VSPR. Therefore, there is a rational basis for why the MTase proteins might be more inclined to repair \( \text{6mG:T} \) if Vsr is present for a concerted effort to restore the normal G:C pairing.
Figure 5.1. Sequence of *ogt* and location of PCR primers. Primer sequences used for analysis are shaded in gray.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Primer Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATGCTGAGAT TACTTGAAGA AAAAATTGCC ACGCCACTGG G<del>C</del>CTGTG~GGTGATTTG</td>
<td><strong>TCCACTGTTG GGTGATTG</strong></td>
</tr>
<tr>
<td>GATGAGCAAT TTCGCCCTGG GGGCGTTGAA TGGGAAGAGT ACAGCGAACG CATGCGTCAG</td>
<td></td>
</tr>
<tr>
<td>CGGCTAGACA TCCATTTTGG CAAAGAAGGC TATGAGCGCA TTTCTGCCAC CAATCCAGGC</td>
<td></td>
</tr>
<tr>
<td>CCCACTGCTA CCACTTTCTA CCGTAATATTT GCCGGTAATC TTAGCATTAT TGATACGCTT</td>
<td></td>
</tr>
<tr>
<td>GGGCGGCTG AGCGAGATT GCGGTACAG GTGCTGCTGAC CTGGCTGGCC</td>
<td></td>
</tr>
<tr>
<td>AGGTTATGTC GCAAGCTTCG TGAATATTT GCCGGTAATC TTAGCATTAT TGATACGCTT</td>
<td></td>
</tr>
</tbody>
</table>
**Figure 5.2.** Sequence of *ada/alkB* and location of PCR primers. Primer sequences used for analysis are shaded in gray. The last base in *ada* is the first base of *alkB*. The codons that overlap are underlined.
Figure 5.3. Sequence of mutH and location of PCR primers. Primer sequences used for analysis are shaded in gray. Primer sequences used in production of recombination substrate are shaded in black.

1 ATGTCCGCAAC CTGGCCCGACT
61 GCCGCCACAC TTTCCTCTTT TACATTTGGGA GAAGTGGCAG
121 CTAGAAGATT TAAAACGGGA AAGGCTGGTG TACTGGAGAT
181 GCCAGCGCAG GGGATAAACCG TGGAGAAGAT TTTGCTGCTC
241 ATCCCCTGTTG ATAGTCTCTTG TCGTCGCCGT GAAAACACAT
301 ACGGCGAATA CGCGGCTGAC CTGGGAAACC AGCCACGTGC GGCAAAACT
361 CTGTTGCAATC CGGTGCAAGG CGGCCGCAAG ATCCGGCTTT GCGAGTTGCG
421 CGGTGCCTGT GGAGGGCCAA TGAGAGGAA GACGGCGTGC TCGGTGATCA
481 TTGAATGGTA TAATGTTTCT CGGTGCTGTT GACGGAGTTA CGGTCTGCTC
541 TTACGCTGAAAA GACGGCAAGT AAGCGCCTTA CGCGATGCT
601 GCCGCGCGGA TTGCAAGCCT GCCGCGCGCG TTTTATTTGA AGAGAGATTG
661 CTAACCGGG CTCATTTTCT GATTGCTAG
Figure 5.4. Sequence of \textit{mutL} and location of PCR primers. Primer sequences used for analysis are shaded in gray. Primer sequences used in production of recombination substrate are shaded in black.
Figure 5.5. Sequence of mutS and location of PCR primers. Primer sequences used for analysis are shaded in gray. Primer sequences used in production of recombination substrate are shaded in black.

```plaintext
1  ATGAGTGGAA~TAGAAAA~~CGACGCeGAT ACGCCCATGA TGCAGCAGTA TCTCAGGCTG
61  AAAGCCCAGC ATCCCGAGAT CCTGCTGTTT TACCGGATGG GTGATTTTTA TGAACTGTTT
121  TATCAATACG TATGCTGGGG GTGTAATCTA CTACCGACGC TCTACTTTGT TGAGTTCAAC
181  TCGCGGAGAG ACCGATCTCC GAAGCGGCAG ATACCGGCTT GCGAGAAGTA TGGGATTTCT
241  GCCAAACTGC TGAATCAGGG AGAGTCCGTT GCCATCTGCG AACAAATTGG CGATCCGGCG
301  ACCAGCAAAG GTCCGGTTGA GCGCAAAGTT GTGCGTATCG TTACGCCAGG CACCATCAGC
361  GATGAAGCCC TGTTGCAGGA GCGTCAGGAC AACCTGCTGG CGGCTATCTG GCAGGACAGC
421  AAAGGTTTCG GCTACGCGAC GCTGGATATC AGTTCCGGGC GTTTTCGCCT GAGCGAACCG
481  GCTGACCGCG AAACGATGGC GGCAGAACTG CAACGCACTA ATCCTGCGGA ACTGCTGTAT
541  GCAGAAGATT TTGCTGAAAT GTCGTTAATT GAAGGCCGTC GCGGCCTGGC GCTGCGCGCC
601  GTGCTTGAGC GCCAGCAAAC TATTGGCGCA TTGCAGGATT TCACCGCCGG GCTACAGCCG
661  GTACTGCGTC AGGTCGGCGA CCTGGAACGT ATTCTGGCAC GTCTGGCTTT ACGAACTGCT
721  CGCCCACGCG ATCTGGCCCG TATGCGCCAC GCTTTCCAGC AACTGCCGGA GCTGCGTGCG
781  GAACGTGAGC AGGACAGCAT CATTATGGAT GCCGCGACGC GTCGTAATCT GGAAATCACC
841  TACCGCGCGA GACGCGTTCC CAGACGAAGG AGGACGCTTC CTGGTGCGTT TGGGCTGGCC
901  GCCAGAAGAT TGGACTTCTC TCAAGCAAGC AGAGACGCAG TCTGCTGCGC CATGTACAGC
961  TACATCATTC CAGAGCTAAA AGAGTACGAA GATAAAGTTC TCACCTCAAA AGGCAAAGCA
1021 GTACTGCGTTC AGGTGACTGC CCTGCAAAGTC ATCCGTCCTT CACCGTAAAT AGGCGCGGCC
1081 CCAGCAAGTC GTGTGCGCAGA CACCGGAGCT CGGCACTGCG GCTGGTACGC
1141 CGCAGAGCTC CCGCGCGAGC CCAAAGCTGC GTGAGCTGGA AAGCATTTCG CCGAACGCGA
1201 CTTGCTGACG TGTGCGTCCG GCAAGTCTGT TGGCAGTCTT CACCGTAAA AAGTGGATTT
1261 GCCGGTCTGT TTTATCCGCAG CGCGTATCTC GTGCTGCTGG GTGCTTTTAT CTGCAGCTGT
1321 CACGCGGCGA CCAATTAATC GACCGGCTG TGGCAGATGC GCTCAGCAGC GTGCGCGCGC
1381 TACCGCGCGA AAGTGGCTTT TAATGGCGCG CACCGCTACT ACATTTCAAT CACGGCGCGC
1441 CACGCGGCGA TCGCCGCAGA CAATACCCAG GTGCTGAGCC CCAGGAAAGA GCGCGCGGCC
1501 TACATCATTC CAGAGCTAAA AGAGTACGAA GATAAAGTTC TCACCTCAAA AGGCAAAGCA
1561 CTGCGACTCG TAAAACACGT TTTAAGAAGG TCTGCTGACT TGCCGAAATG CTGAGTCTG
1621 GCATCGTGGC CACCGGAGCT CGGCACTGCG GCTGGTACGC
1681 CGGCGCTATA CAACTCGCACT CGACTGCGGC ACGCGATTTG CCAGGCGGGG CACCGCGCCG
1741 ACCGAAGTC GCGCGCGCTT GTGTCGGCGA TTGGCGTTCG GTGCTGCTGG GTGCTTTTAT CTGCAGCTGT
1801 GTCTAATCTG CCGCCAGGCG CGCGCGCTGT TGGCAGTCTT CACCGTAAA AAGTGGATTT
1861 ATGCACCTTA TGCGCGCAGC CGCACTGCTG CCGCTTGGTG CGCAGTCTTT CAGTACAGCA
1921 TCGCCGCTAA AAGTCACTGG TGGCAGATTG TCGCCTGAGT ATGGCCTGAG CACCGGCGCC
1981 GTATGACCTGG CTGCTGCGGC ATATCTGGCA GTGCTGCTGG GTGCTTTTAT CTGCAGCTGT
2041 TTACATACGC CGGCGGAGCT GGAGCTGCTG GCTGCTGCTG ATGGCCTGAG CACCGGCGCC
2101 ACCTCAGATG GTGCTGCTCT GTGCGGAGCT GTGCTGCTGG GTGCTTTTAT CTGCAGCTGT
2161 CGGTGCTATG GTGCTGCTCT GTGCGGAGCT GTGCTGCTGG GTGCTTTTAT CTGCAGCTGT
2221 ATGGCGCTATG GTGCTGCTCT GTGCGGAGCT GTGCTGCTGG GTGCTTTTAT CTGCAGCTGT
2281 AGCTCGCGTA ATGGCGCTATG GTGCTGCTCT GTGCGGAGCT GTGCTGCTGG GTGCTTTTAT CTGCAGCTGT
2341 CTGCGCAAGG AGATTTTTAA CCGCGCGCGA CCAAGCGGCG CGAGCAGCGC GTGCTGCTGG GTGCTTTTAT CTGCAGCTGT
2401 CCGCGGAAGG CGCGCGGAGG CAGCGGAGGT AGCGGAAATGT AGCTGCTGTAT
2461 CAGCTGCTGCTG CAGCTGCTGCTG CAGCTGCTGCTG CAGCTGCTGCTG
2521 CAGGCGGAGG AGGCTGCTGCTG CAGCTGCTGCTG CAGCTGCTGCTG
```
Figure 5.6. Sequence of \textit{vsr} and location of PCR primers. Primer sequences used for analysis are shaded in gray. Primer sequences used in production of recombination substrate are shaded in black.

\begin{verbatim}
1 ATGGCCCAAG TTTCACGAAG GGCCACTCGC AGCAAAAATA TGCGCGCGAT TGCCACGCGT
2 61 GATACCGCGA TAGAGAAGCC GTGGCCCATG TTGGACCGGT TGGATTTTCA AGATTTTTCA
121 GTTCAGCGAC CGACTCTGCC CGAACGTCGG GATTTGTTGG TTGATGAATA TCGCTGCTGG
181 ATATTTACCC ATGGCTGCTT CTGGCATCAT CATCACTGCT ATCTGTTTAA AGTGCCTGCG
241 ACTGGAACCC AGTTCATGGT GTGGAGAGATA GGTAAAAAGT TTGAGCGCGA TGCCCGCGAT
301 ATCAGTGGCT TGCGAGGACT CGCTGGCGCT GATTTGATTG TCTGGGAGTG CGCGTTACGT
361 CGGCTGAGGA AGCTGAACGA TGAGGGGTAG ACCAGCGTC TGGAAGAGTG GATCTGCAGC
421 GAAGCTGCA GGGCGAGAT CGACACCGAG GGGATTCATT TACTCGGTTG A
\end{verbatim}
**Figure 5.7.** Relative efficiencies of MTase repair in MR defective strains

![Graph showing the relative efficiencies of MTase repair in MR defective strains.](image)

<table>
<thead>
<tr>
<th>Parent Cell Type</th>
<th>% Repair C215</th>
<th>% Repair C216</th>
<th>% Repair C217</th>
<th>% Repair C218</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td>96</td>
<td>69</td>
<td>95</td>
<td>6</td>
</tr>
<tr>
<td>vsr</td>
<td>81</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>mutH</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>97</td>
</tr>
<tr>
<td>mutL</td>
<td>67</td>
<td>70</td>
<td>63</td>
<td>44</td>
</tr>
<tr>
<td>mutS</td>
<td>88</td>
<td>71</td>
<td>39</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parent Cell Type</th>
<th>% of Parental C215</th>
<th>% of Parental C216</th>
<th>% of Parental C217</th>
<th>% of Parental C218</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>vsr</td>
<td>84</td>
<td>3</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>mutH</td>
<td>101</td>
<td>101</td>
<td>101</td>
<td>101</td>
</tr>
<tr>
<td>mutL</td>
<td>70</td>
<td>52</td>
<td>62</td>
<td>62</td>
</tr>
<tr>
<td>mutS</td>
<td>71</td>
<td>43</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

- Chapter 5 -
Figure 5.8. PCR mediated gene replacement

- Chapter 5 -

- Chapter 5 -
Figure 5.9. *In vivo* assay for MTase repair
**Figure 5.10.** Hypothetical mechanism for stimulation of MTase repair by VSPR proteins. (A) MutS and MutL identify the $\phi$mG mismatch and (B) present it in a looped structure. (C) MutL stimulates the affinity of Vsr for the DNA. In the absence of a substrate for Vsr, the protein assists in localizing the MTase to the $\phi$mG lesion site.
**Figure 5.11.** Bacterial two-hybrid system utilizing domains of adenylate cyclase. (A) The fused T18 and T25 fragments of *Bordetella pertussis* adenylate cyclase exhibit a constitutive enzymatic activity that results in cAMP synthesis. (B) However, when the two fragments are co-expressed as independent polypeptides are unable to interact and cAMP synthesis does not occur. (C) When the two fragments are co-expressed as fusion proteins with X and Y, cAMP synthesis results from binding between X and Y.
Table 5.1. PCR primers used for construction of recombination substrates.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer 5' Sequence</th>
<th>Primer 3' Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>vsr</td>
<td>5'-CCT CGC CAG TCT GTT AAC CGG GCA GGG CCT GGC ATT TCG CCT CGA CAT CTT GGT TAC CGT GAA G-3'</td>
<td>5'-AAG CGC TTC ATC CGT CAG CTT CTC GCG CCC ACG TAA CGC GCG CGG AAT AAC ATC ATT TGG TGA C-3'</td>
</tr>
<tr>
<td>mutS</td>
<td>5'-GAC GAC GCA AAA CGC GCG TCG CAA CTG CTG GAT ATT TCA CTG ACC AAC TCG ACA TCT TGG TTA CCG TGA AG -3'</td>
<td>5'-GGT ACT GAC AGC AAA GAC ATT TGC GTA CCA TCC ACT TGC GTA GCG GCG CGC GGA ATA ACA TCA TTT GGT GAC -3'</td>
</tr>
<tr>
<td>mutL</td>
<td>5'-GCC TCG ATG CAG GTG CGA CGC GTA TCG ATA TTG ATA TCG AAC GCC TCG ACA TCT TGG TTA CCG TGA AG -3'</td>
<td>5'-CAG GGT TAT GGC CTG TGG GTA CTG CAT TGA CCA CTG CGC ATG TTC GCT CGC GGA ATA ACA TCA TTT GGT GAC -3'</td>
</tr>
<tr>
<td>mutH</td>
<td>5'-CAC TTG TCG GCC TGG TTA CGC CAG AGA ATT TAA AAC GCG ATA AAG CTC GAC ATC TTG GTT ACC GTG AAG -3'</td>
<td>5'-ATG GCT TCG GTA AGC GCT TTC GCA TTC GCT TTC GGT CGT ATC CGC GGA ATA ACA TCA TTT GGT GAC -3'</td>
</tr>
</tbody>
</table>
**Table 5.2.** PCR primers used for analyzing strain genotypes.

<table>
<thead>
<tr>
<th>Gene</th>
<th>5' Primer</th>
<th>3' Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>ada/alkB</td>
<td>5' - TGC TTA ACT GAC GAT CAA CGC TGG -3’</td>
<td>5' - TTA CCT GCC TGA CGG AAT GTC AG -3’</td>
</tr>
<tr>
<td>mutH</td>
<td>5' - ATG TCC CAA CCT CGC CCA CT -3’</td>
<td>5' - ACT GGC CCG TCA TTT TCT GAT CCA GTA G -3’</td>
</tr>
<tr>
<td>mutL</td>
<td>5' - ATG CCA ATT CAG GTC TTA CCG CC -3’</td>
<td>5' - CAT CTT TCA GGG CTT TTA TCG CCG G -3’</td>
</tr>
<tr>
<td>mutS</td>
<td>5' - GAG TGC AAT AGA AAA TTT CGA CGC CC -3’</td>
<td>5' - ACA CCA GGC TCT TCA AGC GAT AAA TC -3’</td>
</tr>
<tr>
<td>ogt</td>
<td>5' - TCC ACT GTG GGT GAT TTG C</td>
<td>5' - TCA TGG CGC AAT AAC CAC TC -3’</td>
</tr>
<tr>
<td>vsr</td>
<td>5' - TTT GAA TTC ATG ATG GCC GAC GTT CAC GAT -3’</td>
<td>5' - GAC AAG CTT TCA AGC GAG TAA ATG AAT CCC -3’</td>
</tr>
</tbody>
</table>
Table 5.3. Sizes of gene and PCR amplification products.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene Size</th>
<th>PCR Product Size*</th>
<th>KO Product Size</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>ada</em></td>
<td>1065</td>
<td>~1700</td>
<td>~2000</td>
</tr>
<tr>
<td><em>mutH</em></td>
<td>690</td>
<td>~700</td>
<td>~2200</td>
</tr>
<tr>
<td><em>mutL</em></td>
<td>1848</td>
<td>~1850</td>
<td>~2200</td>
</tr>
<tr>
<td><em>mutS</em></td>
<td>2562</td>
<td>~2550</td>
<td>~2200</td>
</tr>
<tr>
<td><em>ogt</em></td>
<td>516</td>
<td>~450</td>
<td>~2100</td>
</tr>
<tr>
<td><em>vsr</em></td>
<td>471</td>
<td>~450</td>
<td>~2200</td>
</tr>
</tbody>
</table>

* - Note that the expected size of the PCR product may not match the size of the gene if the primer sequences used do not match the 3' and 5' most ends of the target gene. See Figure 5.1 to 5.6 for locations of the sequences used for amplification.
Table 5.4. Results from MTase assay on C215 derivatives.

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Volume Plated</th>
<th>Dilution Factor</th>
<th>Light Blue</th>
<th>Dark Blue</th>
<th>Events Plated</th>
<th>Total Events</th>
<th>% Repaired</th>
<th>Average</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C215-1</td>
<td>200</td>
<td>50</td>
<td>19</td>
<td>395</td>
<td>414</td>
<td>20700</td>
<td>95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C215-2</td>
<td>200</td>
<td>50</td>
<td>25</td>
<td>568</td>
<td>593</td>
<td>29650</td>
<td>96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C215-3</td>
<td>200</td>
<td>50</td>
<td>15</td>
<td>376</td>
<td>391</td>
<td>19550</td>
<td>96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C215-4</td>
<td>200</td>
<td>50</td>
<td>16</td>
<td>549</td>
<td>565</td>
<td>28250</td>
<td>97</td>
<td>96</td>
<td>1</td>
</tr>
<tr>
<td>C215V-1</td>
<td>75</td>
<td>133</td>
<td>64</td>
<td>295</td>
<td>359</td>
<td>47867</td>
<td>82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C215V-2</td>
<td>75</td>
<td>133</td>
<td>82</td>
<td>297</td>
<td>379</td>
<td>50533</td>
<td>78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C215V-3</td>
<td>75</td>
<td>133</td>
<td>76</td>
<td>390</td>
<td>466</td>
<td>62133</td>
<td>84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C215V-4</td>
<td>75</td>
<td>133</td>
<td>90</td>
<td>322</td>
<td>412</td>
<td>54933</td>
<td>78</td>
<td>81</td>
<td>3</td>
</tr>
<tr>
<td>C215H-1</td>
<td>200</td>
<td>50</td>
<td>25</td>
<td>656</td>
<td>681</td>
<td>34050</td>
<td>96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C215H-2</td>
<td>200</td>
<td>50</td>
<td>10</td>
<td>411</td>
<td>421</td>
<td>21050</td>
<td>98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C215H-3</td>
<td>200</td>
<td>50</td>
<td>23</td>
<td>602</td>
<td>625</td>
<td>31250</td>
<td>96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C215H-4</td>
<td>75</td>
<td>133</td>
<td>9</td>
<td>306</td>
<td>315</td>
<td>42000</td>
<td>97</td>
<td>97</td>
<td>1</td>
</tr>
<tr>
<td>C215L-1</td>
<td>75</td>
<td>133</td>
<td>133</td>
<td>288</td>
<td>421</td>
<td>56133</td>
<td>68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C215L-2</td>
<td>75</td>
<td>133</td>
<td>145</td>
<td>285</td>
<td>430</td>
<td>57333</td>
<td>66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C215L-3</td>
<td>75</td>
<td>133</td>
<td>146</td>
<td>310</td>
<td>456</td>
<td>60800</td>
<td>68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C215L-4</td>
<td>75</td>
<td>133</td>
<td>157</td>
<td>319</td>
<td>476</td>
<td>63467</td>
<td>67</td>
<td>67</td>
<td>1</td>
</tr>
<tr>
<td>C215S-1</td>
<td>200</td>
<td>50</td>
<td>175</td>
<td>350</td>
<td>525</td>
<td>26250</td>
<td>67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C215S-2</td>
<td>75</td>
<td>133</td>
<td>141</td>
<td>314</td>
<td>455</td>
<td>60667</td>
<td>69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C215S-3</td>
<td>200</td>
<td>50</td>
<td>193</td>
<td>409</td>
<td>602</td>
<td>30100</td>
<td>68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C215S-4</td>
<td>75</td>
<td>133</td>
<td>130</td>
<td>281</td>
<td>411</td>
<td>54800</td>
<td>68</td>
<td>68</td>
<td>1</td>
</tr>
</tbody>
</table>
### Table 5.5. Results from MTase assay on C216 derivatives.

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Volume Plated</th>
<th>Dilution Factor</th>
<th>Light Blue Plated</th>
<th>Dark Blue Plated</th>
<th>Events Total</th>
<th>% Repaired</th>
<th>Average</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C216-1</td>
<td>200</td>
<td>50</td>
<td>137</td>
<td>346</td>
<td>483</td>
<td>24150</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>C216-2</td>
<td>200</td>
<td>50</td>
<td>144</td>
<td>373</td>
<td>517</td>
<td>25850</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>C216-3</td>
<td>200</td>
<td>50</td>
<td>152</td>
<td>315</td>
<td>467</td>
<td>23350</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>C216-4</td>
<td>200</td>
<td>50</td>
<td>161</td>
<td>320</td>
<td>481</td>
<td>24050</td>
<td>67</td>
<td>69 3</td>
</tr>
<tr>
<td>C216V-1</td>
<td>200</td>
<td>50</td>
<td>168</td>
<td>107</td>
<td>275</td>
<td>13750</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>C216V-2</td>
<td>200</td>
<td>50</td>
<td>319</td>
<td>240</td>
<td>559</td>
<td>27950</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>C216V-3</td>
<td>200</td>
<td>50</td>
<td>439</td>
<td>274</td>
<td>713</td>
<td>35650</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>C216V-4</td>
<td>200</td>
<td>50</td>
<td>372</td>
<td>272</td>
<td>644</td>
<td>32200</td>
<td>42</td>
<td>41 2</td>
</tr>
<tr>
<td>C216H-1</td>
<td>200</td>
<td>50</td>
<td>48</td>
<td>111</td>
<td>159</td>
<td>7850</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>C216H-2</td>
<td>200</td>
<td>50</td>
<td>102</td>
<td>226</td>
<td>328</td>
<td>16400</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>C216H-3</td>
<td>200</td>
<td>50</td>
<td>133</td>
<td>317</td>
<td>450</td>
<td>22500</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>C216H-4</td>
<td>200</td>
<td>50</td>
<td>131</td>
<td>343</td>
<td>474</td>
<td>23700</td>
<td>72</td>
<td>70 1</td>
</tr>
<tr>
<td>C216L-1</td>
<td>200</td>
<td>50</td>
<td>283</td>
<td>255</td>
<td>538</td>
<td>26900</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>C216L-2</td>
<td>75</td>
<td>133</td>
<td>302</td>
<td>375</td>
<td>677</td>
<td>90267</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>C216L-3</td>
<td>75</td>
<td>133</td>
<td>217</td>
<td>251</td>
<td>468</td>
<td>62400</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>C216L-4</td>
<td>75</td>
<td>133</td>
<td>185</td>
<td>206</td>
<td>391</td>
<td>52133</td>
<td>53</td>
<td>52 3</td>
</tr>
<tr>
<td>C216S-1</td>
<td>200</td>
<td>50</td>
<td>139</td>
<td>82</td>
<td>221</td>
<td>11050</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>C216S-2</td>
<td>200</td>
<td>50</td>
<td>145</td>
<td>127</td>
<td>272</td>
<td>13600</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>C216S-3</td>
<td>200</td>
<td>50</td>
<td>228</td>
<td>189</td>
<td>417</td>
<td>20850</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>C216S-4</td>
<td>200</td>
<td>50</td>
<td>169</td>
<td>129</td>
<td>298</td>
<td>14900</td>
<td>43</td>
<td>43 4</td>
</tr>
</tbody>
</table>
Table 5.6. Results from MTase assay on C217 derivatives.

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Volume Plated</th>
<th>Dilution Factor</th>
<th>Light Blue</th>
<th>Dark Blue</th>
<th>Events Plated</th>
<th>Total Events</th>
<th>% Repaired</th>
<th>Average</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C217-1</td>
<td>200</td>
<td>50</td>
<td>19</td>
<td>326</td>
<td>345</td>
<td>17250</td>
<td>94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C217-2</td>
<td>200</td>
<td>50</td>
<td>25</td>
<td>473</td>
<td>498</td>
<td>24900</td>
<td>95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C217-3</td>
<td>200</td>
<td>50</td>
<td>15</td>
<td>338</td>
<td>353</td>
<td>17650</td>
<td>96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C217-4</td>
<td>200</td>
<td>50</td>
<td>21</td>
<td>345</td>
<td>366</td>
<td>18300</td>
<td>94</td>
<td>95</td>
<td>1</td>
</tr>
<tr>
<td>C217V-1</td>
<td>200</td>
<td>50</td>
<td>176</td>
<td>472</td>
<td>648</td>
<td>32400</td>
<td>73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C217V-2</td>
<td>200</td>
<td>50</td>
<td>165</td>
<td>423</td>
<td>588</td>
<td>29400</td>
<td>72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C217V-3</td>
<td>200</td>
<td>50</td>
<td>258</td>
<td>545</td>
<td>803</td>
<td>40150</td>
<td>68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C217V-4</td>
<td>200</td>
<td>50</td>
<td>269</td>
<td>510</td>
<td>779</td>
<td>38950</td>
<td>65</td>
<td>70</td>
<td>3</td>
</tr>
<tr>
<td>C217H-1</td>
<td>200</td>
<td>50</td>
<td>7</td>
<td>382</td>
<td>389</td>
<td>19450</td>
<td>98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C217H-2</td>
<td>200</td>
<td>50</td>
<td>9</td>
<td>181</td>
<td>190</td>
<td>9500</td>
<td>95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C217H-3</td>
<td>200</td>
<td>50</td>
<td>6</td>
<td>184</td>
<td>190</td>
<td>9500</td>
<td>97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C217H-4</td>
<td>200</td>
<td>50</td>
<td>17</td>
<td>316</td>
<td>333</td>
<td>16650</td>
<td>95</td>
<td>96</td>
<td>2</td>
</tr>
<tr>
<td>C217L-1</td>
<td>200</td>
<td>50</td>
<td>106</td>
<td>178</td>
<td>284</td>
<td>14200</td>
<td>63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C217L-2</td>
<td>200</td>
<td>50</td>
<td>67</td>
<td>113</td>
<td>180</td>
<td>9000</td>
<td>63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C217L-3</td>
<td>200</td>
<td>50</td>
<td>88</td>
<td>128</td>
<td>216</td>
<td>10800</td>
<td>59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C217L-4</td>
<td>200</td>
<td>50</td>
<td>202</td>
<td>386</td>
<td>588</td>
<td>29400</td>
<td>66</td>
<td>63</td>
<td>3</td>
</tr>
<tr>
<td>C217S-1</td>
<td>200</td>
<td>50</td>
<td>116</td>
<td>58</td>
<td>174</td>
<td>8700</td>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C217S-2</td>
<td>200</td>
<td>50</td>
<td>234</td>
<td>178</td>
<td>412</td>
<td>20600</td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C217S-3</td>
<td>200</td>
<td>50</td>
<td>144</td>
<td>90</td>
<td>234</td>
<td>11700</td>
<td>38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C217S-4</td>
<td>200</td>
<td>50</td>
<td>364</td>
<td>259</td>
<td>623</td>
<td>31150</td>
<td>42</td>
<td>39</td>
<td>4</td>
</tr>
</tbody>
</table>
Table 5.7. Results from MTase assay on C218 derivatives.

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Volume Dilution</th>
<th>Light Blue</th>
<th>Dark Blue</th>
<th>Events Plated</th>
<th>Total Events</th>
<th>% Repaired</th>
<th>Average</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C218-1</td>
<td>200</td>
<td>50</td>
<td>418</td>
<td>31</td>
<td>449</td>
<td>22450</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>C218-2</td>
<td>200</td>
<td>50</td>
<td>474</td>
<td>37</td>
<td>511</td>
<td>25550</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>C218-3</td>
<td>200</td>
<td>50</td>
<td>234</td>
<td>12</td>
<td>246</td>
<td>12300</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>C218-4</td>
<td>200</td>
<td>50</td>
<td>407</td>
<td>27</td>
<td>434</td>
<td>21700</td>
<td>6</td>
<td>6 1</td>
</tr>
<tr>
<td>C218V-1</td>
<td>200</td>
<td>50</td>
<td>460</td>
<td>30</td>
<td>490</td>
<td>24500</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>C218V-2</td>
<td>200</td>
<td>50</td>
<td>430</td>
<td>42</td>
<td>472</td>
<td>23600</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>C218V-3</td>
<td>200</td>
<td>50</td>
<td>506</td>
<td>35</td>
<td>541</td>
<td>27050</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>C218V-4</td>
<td>200</td>
<td>50</td>
<td>369</td>
<td>28</td>
<td>397</td>
<td>19850</td>
<td>7</td>
<td>7 1</td>
</tr>
<tr>
<td>C218H-1</td>
<td>200</td>
<td>50</td>
<td>258</td>
<td>20</td>
<td>278</td>
<td>13900</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>C218H-2</td>
<td>200</td>
<td>50</td>
<td>218</td>
<td>13</td>
<td>231</td>
<td>11550</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>C218H-3</td>
<td>200</td>
<td>50</td>
<td>320</td>
<td>23</td>
<td>343</td>
<td>17150</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>C218H-4</td>
<td>200</td>
<td>50</td>
<td>436</td>
<td>26</td>
<td>462</td>
<td>23100</td>
<td>6</td>
<td>6 1</td>
</tr>
<tr>
<td>C218L-1</td>
<td>200</td>
<td>50</td>
<td>66</td>
<td>66</td>
<td>132</td>
<td>6600</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>C218L-2</td>
<td>200</td>
<td>50</td>
<td>58</td>
<td>47</td>
<td>105</td>
<td>5250</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>C218L-3</td>
<td>200</td>
<td>50</td>
<td>92</td>
<td>69</td>
<td>161</td>
<td>8050</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>C218L-4</td>
<td>200</td>
<td>50</td>
<td>136</td>
<td>91</td>
<td>227</td>
<td>11350</td>
<td>40</td>
<td>44 4</td>
</tr>
<tr>
<td>C218S-1</td>
<td>200</td>
<td>50</td>
<td>149</td>
<td>9</td>
<td>158</td>
<td>7900</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>C218S-2</td>
<td>200</td>
<td>50</td>
<td>475</td>
<td>32</td>
<td>507</td>
<td>25350</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>C218S-3</td>
<td>200</td>
<td>50</td>
<td>525</td>
<td>38</td>
<td>563</td>
<td>28150</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>C218S-4</td>
<td>200</td>
<td>50</td>
<td>498</td>
<td>22</td>
<td>520</td>
<td>26000</td>
<td>4</td>
<td>6 1</td>
</tr>
</tbody>
</table>
5.7. REFERENCES


225


226


CHAPTER 6
Cloning, Expression, and Purification of Truncated Hormone Receptors
6.1. INTRODUCTION

The final two projects are ones that are not related to mismatch repair. My skills in protein purification were needed on another project involving the synthesis of agents that kill cells by blocking DNA repair. My advisor asked me to include these chapters, as they represent important contributions to the lab research portfolio.

6.2. NUCLEAR HORMONE RECEPTORS

Nuclear hormone receptors (NHR) are ligand inducible transcription factors that regulate gene expression (1). Members of this family include the estrogen receptor (ER), androgen receptor (AR), progesterone receptor (PR), and glucocorticoid receptor (GR) proteins. In the absence of hormones, these receptors are associated with heat shock proteins and are not active (1;2). However, upon binding to their cognate ligand, these proteins undergo a series of events including conformational changes, dissociation from heat shock complexes, dimerization, phosphorylation, nuclear translocation, and binding to promoter sequences in DNA called hormone response elements (1-4). Promoter binding is followed by the recruitment of coactivators, corepressors, and transcriptional machinery for regulation of target genes (2).

The observation that a number of cancer types over express hormone receptors makes these proteins good therapeutic targets (5-9). Indeed, many cancer treatments act by interrupting hormone dependent transcription and cell proliferation, and a number of these treatments are highly cancer specific because of the fidelity of ligand binding by certain hormone receptors. Other cancer drugs produce their cytotoxic effects through covalent damage to DNA (10;11). Our work aims to combine the specificity of steroid binding with the toxicity of DNA damage to produce a potent drug that is selectively lethal to cancer cells (Figure 6.1).
In principle, tethering a synthetic ligand for a cancer specific receptor to a chemical functionality that crosslinks DNA allows a bifunctional approach to slowing cancer progression. First, the bifunctional compound may bind the cognate hormone receptor and antagonize the receptors role in malignant growth. Second, the ability of the compound to recruit a receptor to the site of DNA damage and render the damage undetectable, allows time for interstrand crosslinks to form. We hypothesize that toxic crosslinks will form preferentially on DNA in cancer cells due to the abundance of camouflaging receptor proteins. In contrast, non-cancer cells containing relatively less receptor protein will selectively detect and repair the monoadduct before formation of a toxic crosslink.

The ability of such bifunctional agents to possess selective toxicity, on the basis of the cell receptor status, has been demonstrated by our lab (12-15). The ability of lead compounds, both in solution and bound to duplex DNA, to bind hormone receptors was also investigated biochemically. This section describes my work towards these ends.

Four truncated human receptor genes were cloned and three receptor ligand binding domains were expressed in S21 insect cells using a baculovirus. This chapter contains the methods by which these genes were cloned and these proteins expressed and purified. Chapters 7 and 8 contain the publications to which my work contributed.
6.3. EXPERIMENTAL PROCEDURES

Preparation of the hER-LBD. The human estrogen receptor ligand binding domain (hER-LBD) was obtained by PCR amplification of a segment of the hER gene contained within the plasmid pSV2NeoCMV-hER (16). Primers 5'-GGA TCC ATG TCT GCT GGA GAC ATG AGA GCT GCC -3' and 5'-CCG CTC GAG TCA GAC TGT GGC AGG GAA ACC CTC -3' were used to amplify a 0.94 kb segment containing amino acids 282 to 595 of the hER. Using the BamH I and Xho I restriction sites contained within these primers the amplified segment was inserted into complementary sites of pSK278 (17) (provided by Sang Seok Koh, Whitehead Institute, Figure 6.2) using T4 DNA ligase. Following transformation into competent E. coli DH5α and restriction analysis of DNA from isolated clones, the recombinant transfer vector (pSK278-hER-LBD, Figure 6.3) and pBacPAC6 viral DNA (Clonetech, Palo Alto, CA) were cotransfected into S21 cells (Invitrogen, Carlsbad, CA). Recombinant virus was collected, purified via plaque assay and used to reinfect S21 cells from which cell extracts were prepared. Expression of the expected 34 kDa protein containing the hER-LBD fused to a FLAG® peptide was confirmed by Western blot analysis. The hER-LBD was purified from S21 cell lysates by immunoaffinity chromatography using ANTI-FLAG® M2-affinity gel (Sigma). Following adsorption to the gel, the 34 kDa hER-LBD protein was recovered by elution with FLAG® -peptide (Sigma). The estradiol binding activity of the purified hER-LBD was confirmed using a competitive assay with [3H]-17β-estradiol (18).
Preparation of the hAR-LBD. The human androgen receptor ligand binding domain (hAR-LBD) was obtained by PCR amplification of a segment of the hAR gene contained within the plasmid pCMV3.1 AR (19). Primers 5'-AAT GCG GCC GCT ACT CTG GGA GCC CGG AA -3' and 5'-GGG TTA ATT AAG GCT CAC TGG GTG TGG AAA TAG AT -3' were used to amplify a 0.89 kb segment containing amino acids 624 to 919 of the hAR. Using the Not I and Pac I restriction sites contained within these primers the amplified segment was inserted into complementary sites of pSK278 (17) (provided by Sang Seok Koh, Whitehead Institute, Figure 6.2) using T4 DNA ligase. Following transformation into competent E. coli DH5α and restriction analysis of DNA from isolated clones, the recombinant transfer vector (pSK278-hAR-LBD, Figure 6.4) and pBacPAC6 viral DNA (Clonetech, Palo Alto, CA) were cotransfected into S21 cells (Invitrogen, Carlsbad, CA). Recombinant virus was collected, purified via plaque assay and used to reinfect S21 cells from which cell extracts were prepared. Expression of the expected 33 kDa protein containing the hAR-LBD fused to a FLAG® peptide was confirmed by Western blot analysis. The hAR-LBD was purified from S21 cell lysates by immunoaffinity chromatography using ANTI-FLAG® M2-affinity gel (Sigma). Following adsorption to the gel, the 33 kDa hAR-LBD protein was recovered by elution with FLAG® -peptide (Sigma). The R1881 binding activity of the purified hAR-LBD was confirmed using a competitive assay with [17α-methyl-3H]-(R1881), analogous to that for the estrogen receptor (18).
**Preparation of the hPR-LBD.** The human progesterone receptor ligand binding domain (hPR-LBD) was obtained by PCR amplification of a segment of the hPR gene contained within the plasmid pCR®3.1 hPR-B (provided by Weigel Lab, Baylor College of Medicine). Primers 5'- AAC GGA TCC ATG GAA GGG CAG CAC AAC TAC -3' and 5'-CCG CTC GAG GCC TCA CTT TTT ATG AAA GAG -3' were used to amplify a 1.0 kb segment containing amino acids 595 to 933 of the hPR. Using the BamH I and Xho I restriction sites contained within these primers the amplified segment was inserted into complementary sites of pSK278 (17) (provided by Sang Seok Koh, Whitehead Institute, Figure 6.2) using T4 DNA ligase. Following transformation into competent *E. coli* DH5α and restriction analysis of DNA from isolated clones, the recombinant transfer vector (pSK278-hPR-LBD, Figure 6.5) and pBacPAC6 viral DNA (Clonetech, Palo Alto, CA) were cotransfected into Sf21 cells (Invitrogen, Carlsbad, CA). Recombinant virus was collected, purified via plaque assay and used to reinflect Sf21 cells from which cell extracts were prepared. Expression of the expected 38 kDa protein containing the hPR-LBD fused to a FLAG® peptide was confirmed by Western blot analysis. The hPR-LBD was purified from Sf21 cell lysates by immunoaffinity chromatography using ANTI-FLAG® M2-affinity gel (Sigma). Following adsorption to the gel, the 38 kDa hPR-LBD protein was recovered by elution with FLAG®-peptide (Sigma). The progesterone binding activity of the purified hPR-LBD was confirmed using a competitive assay with [³H]-progesterone, analogous to that for the estrogen receptor (18).
Preparation of the hGR-LBD Transfer Vector. The human glucocorticoid receptor ligand binding domain (hGR-LBD) was obtained by PCR amplification of a segment of the hGR gene contained within the plasmid OB7 (20). Primers 5'- AAT GCG GCC GCT GAA GCT CGA AAA ACA AA -3' and 5'- GGC TTA ATT AAC GCG TCA CTT TTG ATG AAA C -3' were used to amplify a 0.87 kb segment containing amino acids 489 to 777 of the hGR. Using the Not I and Pac I restriction sites contained within these primers the amplified segment was inserted into complementary sites of pSK278 (17) (provided by Sang Seok Koh, Whitehead Institute, Figure 6.2) using T4 DNA ligase. The recombinant transfer vector (pSK278-hGR-LBD, Figure 6.6) was transformed into competent E. coli DH5α and restriction analyses were conducted to identify successful clones.
Figure 6.1. Model for mechanism of bifunctional toxicants
Figure 6.2. Restriction map of pSK278
Figure 6.3. Restriction map of pSK278-hER-LBD
Figure 6.4. Restriction map of pSK278-hAR-LBD
Figure 6.5. Restriction map of pSK278-hPR-LBD
Figure 6.6. Restriction map of pSK278-hGR-LBD

hGR-LBD in pSK278
6452 bp
6.4. REFERENCES


CHAPTER 7

A Rationally Designed Genotoxin that Selectively Destroys Estrogen Receptor-Positive Breast Cancer Cells

Kaushik Mitra, John C. Marquis, Shawn M. Hillier, Peter T. Rye, Beatriz Zayas, Annie S. Lee, John M. Essigmann, and Robert G. Croy

Department of Chemistry and Division of Bioengineering and Environmental Health, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received October 22, 2001
7.1. ABSTRACT

We describe a novel strategy to increase the selective toxicity of genotoxic compounds. The strategy involves the synthesis of bifunctional molecules capable of forming DNA adducts that have high affinity for specific proteins in target cells. It is proposed that the association of such proteins with damaged sites in DNA can compromise protein function and/or DNA repair resulting in increased toxicity. We describe the synthesis of a bifunctional compound consisting of an aniline mustard linked to the 7α position of estradiol. This novel compound can form covalent DNA adducts that have high affinity for the estrogen receptor. Breast cancer cells that express high levels of the estrogen receptor showed increased sensitivity to the cytotoxic effects of the new compound.
The persistence of genetic damage produced by alkylating agents (1-8) as well as the antagonism of essential biochemical processes such as transcription can have lethal consequences for malignant cells (9;10). Both mechanisms have been identified in studies to uncover the reasons for the efficacy of cisplatin in the treatment of several cancers (11-19). We describe a synthetic strategy to create bifunctional molecules that produce DNA adducts capable of binding the estrogen receptor (ER), which is aberrantly expressed in many breast cancer cells (20). It is speculated that DNA adducts that form complexes with the ER will be poorly repaired in these cells because they are camouflaged from detection by DNA repair enzymes. Consequently, the DNA lesions persist. Furthermore, the DNA adducts would be expected to act as "molecular decoys" capable of displacing the ER from its natural targets and antagonizing its role in malignant growth. In healthy cells, where the abundance of the ER is minimal, no such ER-DNA adduct complexes will be present, and the cell should survive (21).

In this report we describe the design and synthesis of compound 1, a bifunctional agent that can form covalent DNA adducts capable of binding the ER with high affinity and specificity. We show that 1 has selective toxicity toward ER+ breast cancer cells compared to ER- cells in vitro.

Compound 1 consists of a bis-chloroethyl aniline mustard as the DNA alkylating unit tethered to estradiol, the natural ligand for the ER. The site of substitution of estradiol in 1 was based on reports that relatively large alkyl groups can be attached at the 7α position with retention of high affinity for the ER (22-24). The synthetic strategy for 1 is shown in Scheme 7.1. Compound 7, a key compound in the synthesis, was prepared by a modification of a published strategy (25;26). Briefly, 3 was functionalized with a 6-carbon chain at the 7-position in α-stereochemistry to provide the alkenyl steroid 4. Efficient reduction of the 6-oxo group in 4 was achieved with Et₃SiH/BF₃·Et₂O; however, this treatment also caused the loss of 3,17-tetrahydropyranoxy (THP) groups producing diol 5. The 3,17-OHs of 5 were reprotected with THP groups to afford 6, followed by oxidation of the alkene at the terminus of the linker to provide alcohol 7. Steroid alcohol
7 was converted to bromide 8, which was subsequently allowed to react with a protected ethanolamine to give 9. Compound 9 was desilylated with tetrabutylammonium fluoride (TBAF) and converted to a carbonate intermediate with p-nitrophenyl chloroformate. The carbonate was coupled to (N,N-bis-2-chloroethylaminophenyl)propylamine that was prepared from chlorambucil via the Curtius reaction[27]. Deprotection of the product in HCl/methanol furnished 1.

The affinity of 1 for the ER was first determined. Using a competitive binding assay [28] with [3H]-17β estradiol, compound 1 was found to have a relative binding affinity (RBA) for the calf uterine ER of 30; RBA of estradiol = 100.

Next, the ability of 1 to modify DNA covalently was investigated. Plasmid DNA was incubated with 100 μM [14C]-1 (obtained by reacting the carbonate intermediate with 4-(N,N-bis-2-chloroethylaminophenyl)-[1-14C] propylamine in step xii in Scheme 6.1; additional details in Supporting Information) at 37 °C for up to 6 h. After unbound 1 was removed by phenol-CHCl₃ extraction and ethanol precipitation, the radioactivity associated with DNA was measured. The amount of radioactivity bound to DNA increased at a constant rate over the 6-h period indicating the formation of covalently bound 1 (see Supporting Information). On the basis of previous studies on DNA alkylation by nitrogen mustards [29-31], it is likely that covalent adducts of 1 are formed primarily at the N7 atom of guanines. To investigate the identity of the covalent adducts, 1-modified DNA was treated with 0.1 N HCl, the major product from the digested DNA was isolated by reversed-phase HPLC and analyzed by full scan electrospray mass spectrometry to yield a prominent molecular ion signal at M + H⁺/2 813.5051. This mass is consistent with a chemical structure in which one ethylene arm of the mustard of 1 is attached to guanine and the other arm contains a -OH substituted for the Cl atom. DNA modified with d₄-1 (deuterium labeled 1) and analyzed using the same conditions for hydrolysis, HPLC and mass spectrometry revealed a molecular ion signal at M+H⁺/2 817.5. This result is consistent with d₄-1 attached to guanine as described. The preparation of d₄-1 is described in Supporting Information.)
The affinity of DNA adducts of 1 for the ER was investigated using an Electrophoretic Mobility Shift Assay (EMSA). Substrates for this assay were prepared by reaction of 5'[32P]-labeled self-complementary oligonucleotide 5'-d(ATTATTGGCCAATAAT) with 1 for 4 h at 37 °C. To assess quantitatively the level of covalent modification under these conditions, the DNA was treated with piperidine (1 M, 90 °C, 1 h) and the products were separated on a denaturing 20% polyacrylamide gel (Figure 6.1A). This analysis found that 50% of the 16-mer oligo was cleaved, forming two products identical with those produced by the Maxam-Gilbert sequencing reaction for guanine (data not shown). EMSA was performed by combining the modified [32P]-labeled 16-mer with a 30 kD fragment of the ER containing the ligand binding site for estradiol (ER-LBD). The 30 kD ER-LBD was obtained by expression of a gene fragment containing amino acids 282 to 595 of the human estrogen receptor in Sf21 cells using the baculovirus expression system. See Supporting Information). Analysis on a 4% non-denaturing polyacrylamide gel (0.5X TBE) revealed that approximately 50% of the 16-mer formed a slowly migrating band (Figure 6.1B; lane 2). The amount of retarded band corresponded closely with the level of modification of the 16-mer as revealed by piperidine treatment. Furthermore, no slowly migrating band was seen with unmodified 16-mer (Figure 7.1B, lane 1) or with 16-mer modified with the nitrogen mustard chlorambucil (see Supporting Information). To examine the identity of the complex, increasing amounts of 17β-estradiol were added to the mixture containing the ER-LBD and 1-modified 16-mer. As shown in Figure 7.1B (lanes 3-8), addition of 17β-estradiol resulted in disruption of the ER-LBD-DNA complex. These in vitro results indicated that ER-LBD selectively forms a complex with 1-modified oligo, and raises the strong possibility of formation of similar complexes in cells.

Clonal survival assays were performed in ER+ and ER- breast cancer cell lines to determine if ER status affected the sensitivity of cells to the toxicity of 1. To reveal clearly the effect of DNA adducts, cells were exposed to 1 for 2 h after which fresh drug-free medium was added to the cell cultures. This procedure minimized the possibility that unreacted 1 could function as a receptor antagonist and thereby inhibit growth or survival. Under these conditions 1 was found to possess a significantly lower EC50 in the
ER+ cell line MCF-7 (EC₅₀ = 3.5 µM) as compared to the ER- cell line MDA-MB231 (EC₅₀ = 9.2 µM) (Figure 7.2). In contrast to these results, chlorambucil under the same conditions did not show any difference in EC₅₀ between these two cell lines (Figure 7.2).

The fact that ER status affected the EC₅₀ of 1, but not of chlorambucil, is consistent with the role of the ER as an effector of selective toxicity of 1. Covalent 1-guanine adducts have been identified in DNA isolated from treated cells (unpublished). Therefore, 1 is stable to cell culture conditions and can form DNA adducts in cells identical with those formed in vitro. We are currently investigating the role of 1-DNA adducts in vivo to determine if they are less efficiently repaired in ER+ cells and to evaluate whether the association of the ER with DNA adducts of 1 is directly responsible for the greater sensitivity of MCF-7 cells.

Acknowledgment

We thank John Wishnok and W. Sara Gardinar-Stillwell for helpful discussions. We are grateful to the National Cancer Institute (1-RO1-CA77743-04) for support of this work, to the Susan G. Komen Foundation for a fellowship to K.M. (PDF00-000679), to the National Institutes of Health for support of the NMR facility (1S10RR13886-01), and to the BEH Mass Spectrometry Laboratory for use of their instruments.

Supporting Information Available

Procedures for the synthesis of 1, hER-LBD expression and purification, RBA experiments, DNA-damaging studies, gel shift experiments and toxicity experiments (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.
**Scheme 7.1.** Synthesis of compound 1. \( ^{\text{a}}^{\text{a}} \) Conditions: (i) DHP, pyridinium p-toluenesulfonate, reflux, 92%; (ii) n-BuLi, KOtBu, B(OMe)\(_3\), H\(_2\)O\(_2\); (iii) PCC, 59%; (iv) 6-iodohexene, KOtBu, Et\(_3\)B, 47%; (v) CH\(_3\)COCl/MeOH, Et\(_3\)SiH, BF\(_3\).Et\(_2\)O, 74%; (vi) DHP, pyridinium p-toluenesulfonate, 77%; (vii) BH\(_3\).THF, KOH/H\(_2\)O\(_2\), 66%; (viii) methane sulfonyl chloride, LiBr, 86%; (ix) Ph\(_2\)P(O)NH-CH\(_2\)-CH\(_2\)-OtBDMS, NaH, catalyst tetra-n-butylammonium bromide, 73%; (x) TBAF, 73%; (xi) p-nitrophenylchloroformate, DIEA; (xii) 4-(N,N-bis-2-chloroethylaminophenyl)-propylamine, DIEA; (xiii) H\(\text{+}\), 60%.
Figure 7.1. (A) Piperidine treatment of self-complementary deoxy oligonucleotide 5'-d(AATATTGGCCAAATATT) treated with 1. Lane 1: untreated oligonucleotide, Lane 2: oligonucleotide + 200 μM 1. (B) Retarded mobility of oligonucleotide-1 in the presence of ER-LBD illustrated by EMSA, and disappearance of the retarded band by competition with estradiol. Lane 1: untreated oligonucleotide + ER-LBD, Lane 2: oligonucleotide modified by 200 μM 1 + ER-LBD, Lanes 3-8: modified oligonucleotide + ER-LBD + 10-300 nM estradiol.
Figure 7.2. Survival of MCF-7 (ER+) (■, □) and MD-MB231 (ER-) (○, ◦) cells treated with 1 (closed symbols) or chlorambucil (open symbols).
7.3. REFERENCES


CHAPTER 8

Design, synthesis, and evaluation of estradiol-linked genotoxicants as anti-cancer agents


Department of Chemistry and Biological Engineering Division, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received 27 February 2004; Revised 21 April 2004; accepted 21 April 2004. Available online 2 June 2004.
8.1. ABSTRACT

A series of bifunctional compounds was prepared consisting of $17\beta$ estradiol linked to a DNA damaging $N,N$-bis-(2-chloroethyl)aniline. The objective of our studies was to determine the characteristics of the linker that permitted both reaction with DNA and binding of the resultant covalent adducts to the estrogen receptor. Linker characteristics were pivotal determinants underlying the ability of the compounds to kill selectively breast cancer cells that express the estrogen receptor.
8.2. INTRODUCTION

Many anti-tumor drugs produce their cytotoxic effect through covalent damage to DNA (1,2). Repair enzymes that remove such covalent lesions, restoring the integrity of genetic information, limit the effectiveness of these drugs. Inhibition of DNA repair in cancer cells is an attractive yet not significantly explored strategy to potentiate the therapeutic effects of alkylating chemotherapeutic drugs (3,4). We have previously described the synthesis and biological activity of novel bifunctional compounds that form covalent DNA adducts with high affinity for the estrogen receptor (ER) (5,6). The compounds were designed to test the hypothesis that adduct-ER complexes would be concealed from DNA repair proteins and therefore be refractory to repair. The compounds would thus exhibit greater toxicity in cancer cells that over express the ER protein.

The original series of compounds consisted of an N,N-bis-chloroethylaniline connected to a 2-phenylindole (2PI) group by alkyl-amino-carbamate linkers of various lengths (7). Investigations with these derivatives identified several molecular characteristics of the linker that were important for the compound to react with DNA and bind to the ER. A compound with greater affinity for the ER was obtained by replacing the 2PI group with that of 17β-estradiol connected via the 7α position (8). Both the 2PI and estradiol-based compounds showed increased toxicity toward breast cancer cells that express the ER. Evidence was presented that the differential toxicity of the DNA damaging agents was not due to an anti-hormonal mechanism.

We investigated the molecular features of the linker that are essential for the biological activities of the estradiol-linked compound 1 (Figure 8.1) by varying the structure of the linker connecting the N,N-bis-(2-chloroethyl)aniline and estradiol moiety. The roles of the amino and carbamate groups in the linker were of particular interest because their substitution with different chemical groups might lead to a less complex synthetic process and increase overall yields. The modified linkers retained the six carbon alkyl chain that appears, based on our earlier work, to be essential for the attached ligand to fit into the estradiol binding site of the ER (9). Amino, amido, and guanidino...
groups were incorporated into the region connecting the hexanyl-substituted estradiol and the aniline group. We report here the synthesis and physical properties of these molecules and their cytotoxic effects toward breast cancer cells.
8.3. EXPERIMENTAL PROCEDURES

The synthesis of compound 1 was described in our previous report (10). The syntheses utilized 3,17β-bis-(2-tetrahydropyranoyloxy)-7α-(6-hydroxyhexan-1-yl)-estr-1,3,5(10)triene 2 as the starting compound (see box in Scheme 7.1); its preparation has also been described (11). Construction of linkers proceeded by linear additions to 2 with final addition of the N,N-bis-(2-chloroethyl)aniline moiety.Compound 5 was prepared by conversion alcohol 2 to the bromide, which was allowed to react with a protected ethanolamine providing 3 as described in Ref. (12). The Mitsunobu reaction (13) then was used to couple 1,3-bis-(tert-butoxycarbonyl)-2-methyl-2-thiopseudourea with 3. Reaction of the resulting product 4 with excess (N,N-bis-2-chloroethylaminophenyl)-propylamine followed by acid deprotection produced 5. Procedures described by Linney et al. (14) were applied to incorporate an N,N-disubstituted guanidine moiety into the linker. The preparation of 7 proceeded with the initial reaction of 2 with 1,3-bis(tert-butoxycarbonyl)-2-methyl-2-thiopseudourea under Mitsunobu conditions (15). The resulting product 6 was then allowed to react with an excess of (N,N-bis-2-chloroethylaminophenyl)-propylamine followed by acid deprotection to furnish 7. Compound 9 was prepared by conversion of 2 to the p-nitrophenyl carbonate 8, which was then allowed to react with (N,N-bis-2-chloroethylaminophenyl)-propylamine (16).

Removal of THP groups under acidic conditions produced 9. Compound 10, the starting material for compounds 12–15, was prepared by hydrazinolysis of the phthalimide of 2 that was formed under Mitsunobu conditions (17). Compound 12 containing two amide groups was synthesized by first reacting 10 with the NHS ester of 4(tert-butoxycarbonylamino)butyric acid. Following removal of the THP and Boc groups the terminal amino group was allowed to react with the NHS ester of chlorambucil producing 12. Compound 13 in which the linker contains two amino groups was produced by reduction of 12 with borane dimethylsulfide complex (18). The preparation of 14 was also accomplished by conversion of 2 to the phthalimide via Mitsunobu conditions with subsequent hydrazinolysis to obtain the amine 10. The NHS ester of chlorambucil was then allowed to react with the terminal amine, producing 14. Compound 15, containing a secondary amino group in the linker, was prepared by reduction of the amide in 14 using borane dimethylsulfide complex (18).
8.4. RESULTS AND DISCUSSION

In the initial characterization of the biochemical properties of new compounds 5, 7, 9, 12–15 we evaluated their affinities for the ER. A radiometric competitive binding assay (19) with the rabbit uterine ER was used to determine the relative binding affinity (RBA) of each compound for the ER as compared with estradiol; RBA=100. All of the compounds exhibited some affinity for the ER. The data in Table 7.1 show that the new compounds have RBA values for the ER ranging from 6 to 40. Among the new compounds, 15 containing a single amino group in the linker had an RBA of 40, which is comparable to 1.

Although it is apparent that the original combination of the positively charged secondary amine with the neutral carbamyl group (compound 1) results in a bifunctional compound with excellent affinity for the rabbit uterine ER, compounds 7, 12, and 13 also have good affinities. These molecules were viewed as valuable assets as we move ahead toward probing structure–activity relationships and the biochemical mechanisms underlying the biological activity of 1. It is likely that our 7α-linked estradiol compounds adopt a binding mode similar to that identified for the 7α-undecylamide estradiol analog ICI 164,384 (20). The positioning and orientation of the estradiol moiety of ICI-164,384 within the hydrophobic binding cavity of the ER is directed by its 7α side chain, which protrudes out of a hydrophobic channel extending from the binding pocket. At the surface of the LBD, a 90° flexion of the undecyl chain enables the remainder of the linker to track closely with the surface contours of the LBD (21). The low RBAs of compounds 5, 9, and 14 may result from surface interactions adopted by the linkers in these molecules that do not permit optimal alignment of the estradiol moiety within the binding cavity.

The reactivity of each compound with DNA was assessed by its ability to produce piperidine labile sites in the self complementary deoxyoligonucleotide 5'-d(AATATTGGCCAATATT). The results in Table 8.1 (column 3) indicate the percent of the oligomer that was cleaved by piperidine. Compound 9 in which the alkyl linker contains a single carbamyl group produced the lowest level of modification (i.e., 3%
cleaved by piperidine). Compound 14 containing an amido instead of the carbamyl group produced approximately five times the number of DNA adducts (14% cleaved by piperidine). High levels of reactivity toward DNA were observed with compounds with linkers containing secondary amino groups. The combination of the amino and carbamyl groups in the linker of 1 resulted in a 10-fold increase in reactivity over 9 in which the linker contains only the carbamyl group. The reactivity of 1 was similar to that of 15 in which the linker contains a single secondary amine suggesting that the charged amino group is the major determinant of reaction rate. Compound 13 in which the linker contains a diamine -NH-(CH₂)₄-NH-CH₂- was the most reactive (79% cleaved by piperidine). The same is likely the case for molecules 5 and 7 in which the strongly basic guanidino groups would be cationic under assay conditions. It is likely that the cationic nature of these molecules gives them a high reactivity with DNA by localizing the reactive alkylating group in the vicinity of nucleophilic atoms. A similar result has been reported for a conjugate of chlorambucil with the polyamine spermidine (22).

Using an electrophoretic gel mobility shift assay (23), we observed that covalent DNA adducts of 1, 5, 7, 13, and 15 form complexes with the portion of the ER containing the ligand binding domain (ER-LBD) (Table 1, column 4). Under conditions that allowed complex formation, addition of the ER to the modified DNAs resulted in the appearance of a slowly migrating band by electrophoresis that was eliminated by addition of excess competitor, estradiol (data not shown). The results in Table 1 (column 4) indicate that the extent of complex formation for 1, 7, 13, and 15 were correlated with the RBAs of the unreacted compounds. The exception was compound 5 in which the linker contained both amino and guanidino groups. In this case, despite its low RBA, virtually all of the modified oligonucleotide formed a slowly migrating band. We do not know the basis for this unexpected finding.

Log $P$ and logD values can be predictive of aqueous solubility, absorption, and permeability (24). The lipophilicities of 1, 5, 7, 9, 12–15 were assessed using an HPLC method to estimate the logP of the neutral form of each compound. LogD values at pH 7.4 were estimated using an equation derived by Horváth et al. (25) for basic compounds. The logD values in Table 1 indicated that the aqueous solubilities of the
eight compounds span approximately a 2500-fold range under physiological conditions. The compounds with logD values >5 (compounds 9 and 14) had both low affinities for the ER and low reactivity with DNA. Compounds containing charged groups with calculated logD values <3 generally had the highest affinities for the ER along with the greatest reactivities toward DNA. These relationships, however, did not prove to be reliable predictors of biological activities in cytotoxicity assays against breast cancer cells.

The lethal effects of our new compounds were investigated in the MCF-7 (ER+) and MDA-MB231 (ER−) breast cancer cell lines. The data shown in Figure 8.2 and Table 8.1 indicate that most but not all of the modifications that were introduced in the linker resulted in decreased toxicity toward both cell lines. The low toxicity of 5 and 7, which contain guanidinium groups, may be related to either their poor uptake by cells or their rapid excretion once absorbed (26). Despite showing reactivity toward DNA in vitro, neither compound showed significant toxicity at the highest dose; that is, 20 μM. Lack of uptake may also be responsible for the low toxicity of 9, 12, and 14, which have high logD values that are not predictive of good absorption (24). Further work is warranted to determine if cellular uptake is indeed limiting for these compounds.

As previously reported, 1 was significantly more toxic toward MCF-7 cells than MDA-MB231 cells (23). Compounds 13 and 15 containing amino groups showed toxicity similar to that of 1. Both of these compounds also showed greater toxicity toward the ER-positive MCF-7 cells than toward the ER-negative MDA-MB231 cells. This result was consistent with our intended mechanisms, since the RBAs and reactivities with DNA of 13 and 15 imply greater toxicity on ER-positive cells. It is interesting that the results of the electrophoretic mobility shift assay indicate that DNA adducts of 1 have the greatest affinity for the ER-LBD; compound 1 also shows the largest differential toxicity between the two cell lines.
8.5. CONCLUSION

We designed and synthesized a series of estradiol-aniline mustard-linked bifunctional molecules that differ in their relative affinities for the ER and their capacities to covalently modify DNA. The selective cytotoxic effects of 3 of these compounds (1, 13, 15) toward ER-positive breast cancer cells correlated with their ability to react with DNA and their affinity for the ER, as well as favorable solubility. We are directing research to determine if the interaction of the ER with DNA adducts formed by these compounds is responsible for their selective effects.
8.6. ACKNOWLEDGEMENTS

Research was supported by NIH grant CA77743, the Natural Sciences and Engineering Research Council of Canada (PDF-242490-2001 Fellowship to AND), an Award from the Susan B. Komen Foundation, and NIH grant 1SS10RR13386-01 for support of the NMR and the mass spectrometry facility at MIT.
Figure 8.1. Structure and molecular features of lead compound.
Table 8.1. RBA=relative binding affinity for the rabbit uterine ER as compared to estradiol (RBA=100); % Oligo MODIFIED=percent of 16-mer cleaved by piperidine after treatment with test compound for 24 h at 37 °C; % Oligo SHIFTED=percent of covalently modified 16-mer that formed a slowly migrating complex with the ER-LBD under electrophoresis; ED30=concentration of compound that resulted in 30% clonal survival of cells exposed for 2 h.

<table>
<thead>
<tr>
<th>Compound</th>
<th>RBA</th>
<th>Oligo modified (%)</th>
<th>Oligo shifted (%)</th>
<th>log P</th>
<th>log D</th>
<th>ED30 (μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MCF-7</td>
</tr>
<tr>
<td>1</td>
<td>46</td>
<td>45</td>
<td>93</td>
<td>5.32</td>
<td>2.22</td>
<td>5.1</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>29</td>
<td>93</td>
<td>6.27</td>
<td>3.12</td>
<td>&gt;20</td>
</tr>
<tr>
<td>7</td>
<td>28</td>
<td>58</td>
<td>38</td>
<td>5.67</td>
<td>2.52</td>
<td>&gt;20</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>3</td>
<td>ND</td>
<td>5.61</td>
<td>5.61</td>
<td>&gt;20</td>
</tr>
<tr>
<td>12</td>
<td>29</td>
<td>39</td>
<td>ND</td>
<td>4.85</td>
<td>4.85</td>
<td>&gt;20</td>
</tr>
<tr>
<td>13</td>
<td>29</td>
<td>79</td>
<td>38</td>
<td>5.32</td>
<td>2.22</td>
<td>5.1</td>
</tr>
<tr>
<td>14</td>
<td>13</td>
<td>14</td>
<td>ND</td>
<td>5.07</td>
<td>5.07</td>
<td>&gt;20</td>
</tr>
<tr>
<td>15</td>
<td>40</td>
<td>44</td>
<td>67</td>
<td>5.96</td>
<td>2.87</td>
<td>5.1</td>
</tr>
</tbody>
</table>
Scheme 8.1. Synthesis of estradiol-linked aniline mustards. The starting compound 2 is shown in the box (center). Reagents and conditions: (a) methanesulfonyl chloride, DIEA, THF; (b) LiBr, TEA, DMF, 60 °C; (c) Ph₂P(O)NHCH₂CH₂OTBDMS, NaH, TBAB, Ph-H, 60 °C; (d) TBAF, THF; (e) Boc-NHC(=N-Boc)SCH₃, PPh₃, DIPAD, THF; (f) 4-(N,N-bis-2-chloroethylamino-phenyl)-propylamine, THF/H₂O (90:10), reflux; (g) HCl/dioxane, CH₂Cl₂; (h) p-nitrophenylchloroformate, DIEA, THF; (i) phthalimide, DIPAD, PPh₃, THF; (j) Hydrazine, EtOH, reflux; (k) 4-(N-Boc)n-butyl(N-hydroxysuccinimide)ester, TEA, DMF; (l) 4-(N,N-bis-chloroethylamino-phenyl)butyl(N-hydroxysuccinimide)ester, TEA, DMF; (m) BH₃S(CH₃)₂, THF, HCl; (n) HCl, THF. DIEA = diisopropylethyl amine; TEA = triethyl amine; TBAB = tetrabutylammonium bromide; TBAF = tetrabutylammonium fluoride; DIPAD = diisopropylethyl azadicarboxylate; DMF = dimethylformamide; THF = tetrahydrofuran.

268
Figure 8.2. Survival of MCF-7 (ER+) and MDA-MB231 (ER−) breast cancer cells after 2 h exposure to estradiol-linked toxicants.
8.7. REFERENCES


271


EDUCATION
Massachusetts Institute of Technology - Department of Chemistry
Doctor of Philosophy in Chemistry, June 2006. GPA: 4.8/5.0

Syracuse University - College of Arts and Sciences
B.S. Biochemistry, May 1999. Magna cum laude with honors. GPA: 3.8/4.0
Minors: Quantitative Analysis and Philosophy

PUBLICATIONS


Intermetal Coupling in [(η⁵-C₅R₅)Fe(dppe)]₂(μ-CH=CHCH=CH) and in their Dication and Monocation, Mixed-Valence Forms. Chung, M.-C.; Gu, X.; Etzenhouser, B. A.; Spuches, A. M.; Rye, P. T.; Seetharaman, S. K.; Rose, D. J.; Zubieta, J.; Sponsler, M. B. Organometallics (2003) 22, 3485-3494.


HONORS/AWARDS

Karl Taylor Compton Prize for outstanding contributions to the MIT community, May 2006.
Best Poster Award, The 7th Annual Northeastern Student Chemistry Research Conference, April 2005.
Fellowship funding from The National Institutes of Health, September 2000 - January 2005.
Outstanding Teaching Awards from the MIT Chemistry Department, Fall 1999 and Spring 2000.
Phi Beta Kappa National Honors Society, September 1998.
Phi Eta Sigma National Honor Society, September 1996.

LEADERSHIP/EXTRACURRICULAR

Elected to the MIT Corporation Joint Advisory Committee, October 2004.
Secured funding for, and conducted, a series of conflict negotiation seminars, 2002.
Directed a workshop for Northeastern University's Center for Effective Teaching, September 2001.
Instructed specific lectures for MIT summer biochemistry programs, July 2001.
Certified as a mediator for the Commonwealth of Massachusetts, June 2000.
President of Watson Hall at Syracuse University, January - May 1997.

273