Construction of Fluorescently-Tagged and Adenosine Nucleotide-Binding Mutations of the Human MutS Homolog Heterodimer MSH2-MSH3

Senior Honors Thesis

Presented in Partial Fulfillment of the Requirements for graduation
with research distinction in Biochemistry in the undergraduate
colleges of The Ohio State University

by
Chris Cook

The Ohio State University
August 2009

Project Advisor: Dr. Richard Fishel, Department of Molecular Virology, Immunology and Medical Genetics

Chapter Outline

1. Overview
   i) General mismatch repair background
   ii) Research description
2. Mismatch Repair and Disease
   i) Microsatellite instability and HNPCC
   ii) Tri-nucleotide repeats and Huntington's Disease
3. Mutation of the ATP binding domain of MSH2 and MSH3
4. FRET and construction of a CFP-MSH3 hybrid
5. Overexpression and Purification of MSH2-MSH3
6. ATPase activity of MSH2-MSH3
7. Real time binding-dissociation of hMSH2-hMSH3 using surface plasmon resonance
8. Discussion
9. References

Introduction

DNA mismatch repair is a highly conserved system for correcting mispaired or extra helical nucleotide lesions in DNA. Lesions may result from misincorporation errors during DNA replication, from chemical and physical damage, or from genetic recombination between parental DNA strands lacking perfect homology. In particular, this pathway helps to maintain genomic integrity via the repair of polymerase misincorporation errors by increasing the fidelity of DNA synthesis 1,000 fold[1]. The importance of the mismatch repair pathway is underscored by the fact that mutations in key mismatch repair genes are the cause of certain types of cancers as well as neurodegenerative diseases[8,9].

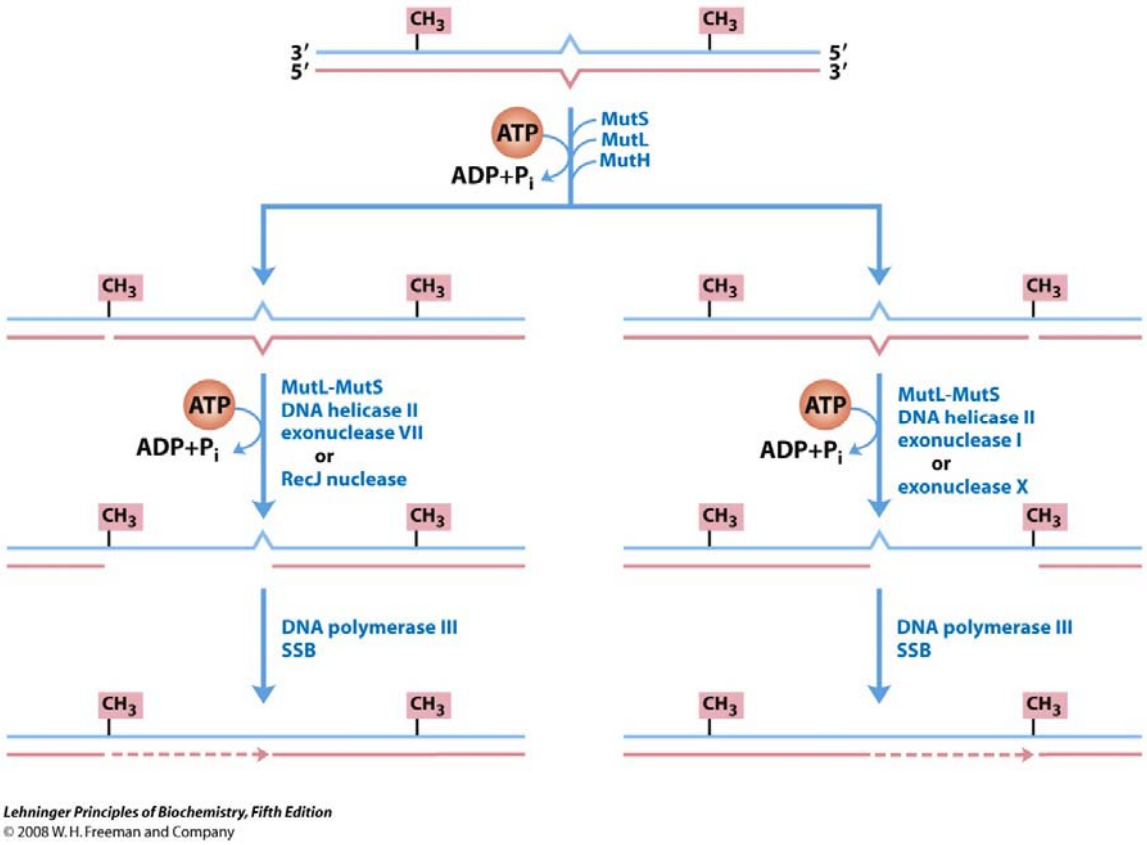
Examination of the mismatch repair system in *Escherichia coli* provides an important model for elucidating similar pathways in eukaryotes. The proteins responsible for initiating mismatch repair in *E. coli* are MutS, MutL, and MutH. Single base mismatches and insertion-deletion loops up to four nucleotides are recognized by the MutS homodimer[2,3]. The mechanism of mismatch repair is controversial[53]. However, the majority of data are consistent with the conclusion that ADP-bound MutS specifically recognizes these lesions which provoke ADP→ATP exchange and loading of multiple MutS sliding clamps on the DNA[4]. ATP-dependent recruitment of MutL by lesion-bound MutS is thought to transmit a signal downstream to MutH, an endonuclease that recognizes and nicks unmethylated GATC sequences in the daughter strand following DNA replication. The nick directs unwinding of the DNA to the site of the mismatch by UvrD helicase, while exonucleolytic cleavage of the newly synthesized strand is accomplished by one of four single-strand specific proteins (RecJ, ExoVII, ExoX, and

ExoI)[5-7]. The resulting gap is resynthesized and sealed by polymerase III and DNA ligase (Fig. 1).

The mismatch repair reaction in eukaryotes is significantly more complex. Many homologs of the MutS and MutL proteins have been discovered and partially characterized (Table 1). No MutH homologs have been identified in higher organisms, which strongly suggests that strand discrimination in the eukayotic mismatch repair pathway occurs via an alternate route, possibly from single-stranded nicks or gaps in newly replicated DNA that have not been formed by DNA ligase. The major MutS homologs (MSH) responsible for the recognition and initiation of mismatch repair in humans are MSH2, MSH3, and MSH6. MSH2 forms heterodimers with both MSH3 and MSH6. MSH2-MSH6 and MSH2-MSH3 have different but overlapping lesion specificities[10]. MSH2-MSH6 functions primarily in the repair of base-base mismatches, and to a lesser extent small insertion-deletion loops (IDL)[11]. MSH2-MSH3 also can recognize and bind single base mismatches, but has a preference for larger IDLs generally less than 16 nucleotides[12,13]. While the eukaryotic mismatch repair system has been extensively studied, many questions remain regarding its fundamental mechanism including the role of activated nucleotides, downstream signaling, strand discrimination, and many others. Answers to these questions will provide a more detailed understanding of how the mismatch repair system works, and how defects can have debilitating pathological consequences such as cancer and neurodegenerative disease.

This research project has focused primarily on human MSH2-MSH3, and to a lesser extent MSH2-MSH6. In the forthcoming chapters, I explain my work on constructing useful mutations that will be used to characterize the ATP and DNA-binding

mechanics of MSH2-MSH3. I will discuss a detailed purification scheme that results in active wild-type MSH2-MSH3. I will detail the development of mutations made in the ATP-binding domain of both MSH2 and MSH3 and explain their significance. I will discuss the development of a fluorescently-tagged CFP-MSH3 hybrid which allows understanding of the conformational changes upon IDL binding using fluorescence resonance energy transfer, and discuss the substrates created and used in all the experiments with MSH2-MSH3.



Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Figure 1. **The Prokaryotic Mismatch Repair Model.** MutS binds to mismatch DNA which recruits MutL and signals downstream to MutH in an ATP dependent process. Excision of the DNA tract between the nick and the mismatch occur by identical mechanisms, but require different nucleases depending on the location of the nick.

MutS and MutL Homologs

Bacteria	Yeast	Human a	Function
MutS	MSH1	?	mitochondrial MMR
	MSH2	***hMSH2***	nuclear MMR
	MSH3	hMSH3	nuclear MMR (IDL)
	MSH4	hMSH4	meiosis
	MSH5	hMSH5	meiosis
	MSH6	***hMSH6***	nuclear MMR
MutL	MLH1	***hMLH1***	nuclear MMR
	MLH2	hPMS1	?
	MLH3	hMLH3	nuclear MMR (IDL)
	PMS I	***hPMS2***	nuclear MMR

Table 1. **Currently identified MutS and MutL homologs** (Table reproduced from Fishel et al. 2000.)

Mismatch Repair and Disease

Hereditary non-polyposis colorectal cancer (HNPCC) is the most common form of colorectal cancer and is intimately associated with mismatch repair[14]. Germline mutations in any one of five mismatch repair genes, *MSH2*, *MSH3*, *MSH6*, *PMS2* and *PMS1* increases susceptibility to HNPCC[15-19]. The majority (80%) of HNPCC-associated mismatch repair mutations occur in *MSH2* and MLH1, with other mismatch repair genes being responsible for only a fraction of cases[14, 20]. HNPCC carriers have a 90% chance of developing cancer by the time they reach seventy years of age[21].

One of the clues that helped link colorectal cancer and mismatch repair was the observation that a fraction of colorectal tumors display microsatellite instability (MSI)[22].

Microsatellites are short repetitive sequences in DNA usually 1-6 base pairs occurring throughout the genome, the most common repeat being $(CA)_n$. The reason these tracts of DNA become unstable is thought to occur from polymerase slippage during replication, resulting in either an extension or shortening of these sequences if left unrepaired by the mismatch repair pathway[23](Fig. 2). In a mismatch repair-deficient environment, microsatellites become unstable and display changes in the length of the respective repeat sequences. Genes containing microsatellites in their coding regions are susceptible to inactivation through frameshift mutations. Inactivation of such genes may ultimately lead to cancer.

Interestingly, while mutations in mismatch repair genes are an underlying cause of HNPCC and other cancers, other fully functional mismatch repair genes appear to be implicated in the pathogenesis of trinucleotide expansion of CAG and CTG sequences found in at least eight hereditary and progressive neurodegenerative and muscular disorders[24-26]. One of these disorders, Huntington's disease (HD) contains repeats of CAG bases found within the coding sequence of the Huntingtin gene. Normal unaffected individuals have 6-29 CAG repeats, carriers of the disease have 29-35, while affected individuals have 36-120 repeats[27]. The expansion of these repeats in the coding region of the gene produces long polyglutamine tracts (CAG codes for glutamine) in the protein, disrupting normal protein function and destroying brain cells.

The underlying cause for CAG expansion has been linked to MSH2-MSH3 mismatch repair proteins. Experiments in mouse models have suggested that functional MSH2 and MSH3 are necessary for the expansion of CAG repeats in HD[29]. It seems discrepant that MSH2-MSH3, which is responsible for repairing DNA, is facilitating

mutation and causing disease. The nature of this relationship is still under investigation, and a possible mechanism is thought to reside in the CAG repeat secondary structure and its interactions with MSH2-MSH3[27]. Long stretches of CAG repeats form hairpin loops with a mismatched base every third nucleotide. These loops form stable hairpin structures that fail to be repaired *in vivo*, while loops comprised of triplets not capable of forming secondary structure were effectively removed[30]. Incorporation of these stabilized-loop intermediates into DNA leads to trinucleotide repeat expansion in a mechanism similar to that of microsatellite instability. MSH2-MSH3 is thought to assist in this process by either failing to correct post-replicative slippage errors or by further stabilizing CAG loop intermediates[27]. It is interesting to note that binding of MSH2-MSH3 to CAG hairpin complexes has been suggested to alter many of its biochemical properties, including its affinity for nucleotides and its tertiary structure[29]. More specifically, binding of MSH2-MSH3 to trinucleotide hairpins causes a conformational transition that inhibits its ATPase activity and likely inhibits its ability to signal the mismatch repair complex downstream[29]. A detailed study of the conformation of MSH2-MSH3 binding to CAG hairpin structures compared to normal IDLs, as well as controlled studies on the effect of such binding on adenosine nucleotide processing would be crucial to elucidate the exact role of MSH2-MSH3 in trinucleotide repeat expansion.

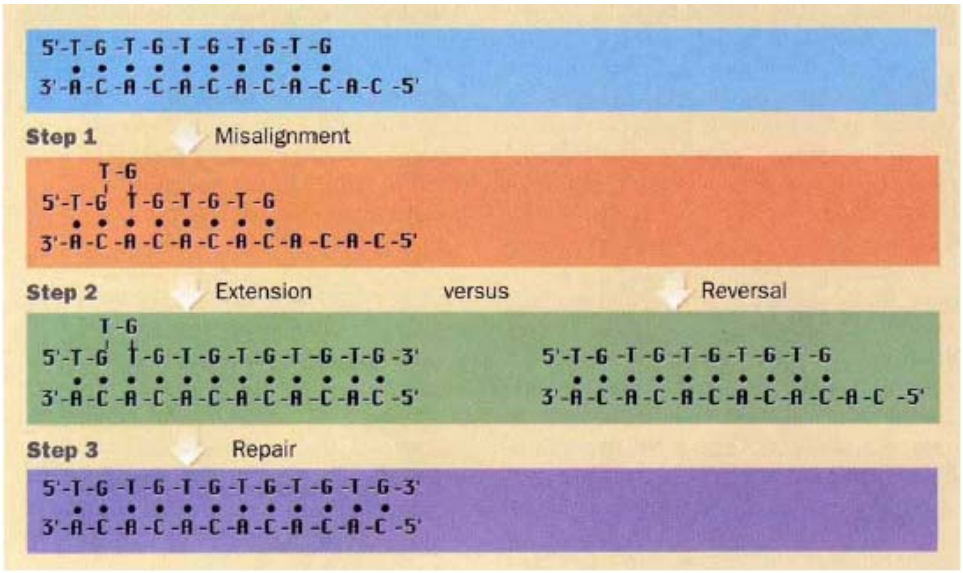


Figure 2. **Polymerase Slippage on Microsatellite Tracts During Replication.** During replication, strand slippage can form intermediates stabilized by downstream base pairs. If left uncorrected by mismatch repair, the following generation contains either an extension or reversal of the repeat tract, causing frameshift mutations in certain genes (Figure reproduced from T. Kunkel, 1993. *Nature* **365:** 207-208)

Results

Mutation of the ATP binding domain of MSH2 and MSH3

Many of the processes of mismatch repair are not well understood, but crystal structures of both MutS and MSH2-MSH6 binding to mismatch DNA have provided clues to the structural and functional basis for mismatch recognition[33,34]. All of the MSH proteins have a conserved Walker A motif near the C terminus which binds adenine nucleotides ATP and ADP[32]. While MSH2-MSH6 and MSH2-MSH3 are structurally very similar (both contain conserved DNA-binding domains, ATP domains, and share a

common MSH2 subunit), the mechanism by which these proteins recognize DNA lesions and bind ATP have been suggested to be different[35]. The MSH2 and MSH3 subunits appear to bind ATP or ADP with similar affinity in the absence of DNA. Upon loop recognition, MSH2 hydrolyzes ATP and binds ADP while the ATP-binding domain of MSH3 remains empty[35]. The lesion-bound MSH3 subunit then binds and hydrolyzes ATP, which promotes ADP→ ATP exchange in MSH2[35]. Similar to MSH2-MSH6, it is reasonable to assume that the ATP-MSH2-MSH3-ADP intermediate would undergo hydrolysis-independent diffusion along the DNA, possibly to signal other components of mismatch repair[13,36]. Because no ADP/ATP-binding-defective mutations were examined, this hypothesis awaits experimental verification.

To provide a more detailed study regarding the role of adenine nucleotide in lesion recognition and repair by MSH2-MSH3, two mutations were made in the ATP-binding domain of MSH3, and co-purified with a homologous mutant of MSH2. The mutants constructed in MSH3 were lysine to arginine and lysine to alanine mutations at amino acid 893 (K893R and K893A, respectively). Similarly, the mutations used in MSH2 were K675R and K675A.

Mutations were introduced using an overlapping PCR method, the general schematic of which appears in Figure 3. The human *MSH3* gene was provided cloned into the pFastBac1 plasmid between the *Nde* I and *Hind*III restriction sites (a gift from Nidhi Punja). Two unique restriction sites were identified flanking the region of the mutation: a *BstE*II site 5' and an *Nco*I site 3' of the desired mutation target sequence. Six primers were synthesized, one contained the *BstE*II restriction sequence designed to anneal to the non-coding strand (BstEII-f; 5'-AAA-GAA-GTG-GGT-GAC-CCA-3',

where f and r denote forward and reverse primers); a mutation primer containing the desired mutation base sequence to anneal to the coding strand (K893R-r; 5'-GTA-GGA-GCT-TCT-TCC-ACC-CAT-3'); a primer complementary to K893R-r (K893R-f; 5'-ATG-GGT-GGA-AGA-AGC-TCC-TAC-3') and a primer containing the Nco I restriction sequence designed to anneal to the non-coding strand (NcoI-r; 5'-CGA-GCT-CCC-ATG-GTC-ATA-AT-3'). Two primers containing the K893A mutation were designed similar to the K893R mutation primers, K893A-f (5'-ATG-GGT-GGA-GCG-AGC-TCC-TAC-3') and K893A-r (5'-GTA-GGA-GCT-CGC-TCC-ACC-CAT-3'). Four separate PCR reactions were set up utilizing the following primer combinations; BstEII-f & K893R-r, NcoI-r & K893R-f; BstEII-f & K893A-r, NcoI-r & K893A-f. The separate PCR mutation products were gel purified, combined, and used as templates for another set of PCR reactions utilizing the outside restriction sequence primers (BstEII-f & NcoI-r). The resulting 1.8 kb band was digested with BstEII and NcoI (New England Biolabs) and gel purified. pFastBac1 with MSH3 was also digested with the same enzymes, and the resulting fragment was ligated to the PCR product overnight at 16 ºC. The mixture was then transformed into XL-10 Gold (Stratagene) competent cells and plated on LB plates containing ampicillin. Colonies were screened for the desired insert by linearization with NcoI and analyzed by gel electrophoresis. Integrity of the final mutation product was confirmed by sequencing.

Expression of Mutant Protein via Recombinant Bacmid

Following confirmation of the mutant gene product, 5 ng of mutant (MSH3 (K893R) and MSH3(K893A)) plasmid were transformed into DH10Bac *E. coli* cells and

plated on LB agar plates containing 50 μg/mL kanamycin, 7 μg, mL gentamicin, 10 μg/mL tetracycline, 100 μg/mL Bluo-gal, and 40 μg/mL isopropyl β-D-1-thiogalactopyranoside. Colonies containing recombinant bacmid appeared white while those without recombination appeared blue. The plates were incubated for 48 hours at 37 ºC. Five white colonies were re-streaked on fresh plates and allowed to grow overnight. Single colonies were inoculated in 5 mL of LB media containing 50 μg/mL kanamycin, 7 μg, mL gentamicin, and 10 μg/mL tetracycline and grown for 16 hours at 37 ºC. Recombinant bacmid DNA was isolated from the overnight culture by miniprep and resuspended in buffer TE to a final concentration of 4 mg/mL.

For bacmid transfections, 9 x 10^5 Sf9 insect cells were seeded in a 6-well tissue culture plate containing 2 mL of Sf-900 II SFM media with 2 μg/mL gentamicin. Cells were allowed to attach to the plate for 1 hour at 27 ºC. During this period, 6 μg bacmid DNA was diluted in 600 μL of Sf-900 II media and combined with 600 μL of media containing 36 μL Cellfectin Reagent (Invitrogen) and incubated for 45 minutes at room temperature. 5.4 mL unsupplemented media was added to the Cellfectin/Bacmid mixture and mixed well. Media was then removed from the 6-well plate containing cells and 1 mL of the Cellfectin/Bacmid mixture was placed in each well and allowed to incubate for 5 hours at 27 ºC. The media was removed from the wells and 2 mL of Sf-900 II media supplemented with 2 μg/mL gentamicin and GlutaMAX (20 mM L-alanyl-L-glutamine; Invitrogen) was added and incubated for 72 hours at 27 ºC.

Cells were visually inspected after 72 hours for signs of viral infection which include cessation of cell growth, detachment from the plate, and a granular appearance when compared to a cell-only control. After verification of successful infection, the

media from each well was collected and placed in a 15 mL conical vial and subjected to centrifugation at 500 x g for 5 minutes to pellet cells and debris. The supernatant was collected, stored at 4 ºC, and labled P1 (Phase-1) viral stock. The P1 stock was used to generate high-titer P2 and P3 viral stocks. 0.5 mL of P1 viral stock was used to inoculate a 10 mL Sf-9 cell culture suspension at 2 x 10^6 cells/mL. The cells were incubated at 27 ºC for 72 hours then centrifuged at 500 x g for 5 minutes. The supernatant was collected and stored as P2 viral stock. 0.5 mL of P2 was used to infect 15 mL of cells at 2 x 10^6 cells/mL. The resulting culture was incubated for 72 hours at 27 ºC, centrifuged, and the supernatant stored as P3 viral stock.

P3 viral stocks of all MSH2 and MSH3 mutants (lysine-to-arginine and lysine-to-alanine) were created. P3 viral stock for MSH2 (K675A), MSH3 (K893A) and wild-type MSH2 and MSH3 were used to infect 200 mL of Sf-9 cells yielding the following permutations of expressed protein; MSH2(wt)-MSH3(wt), MSH2(wt)-MSH3(K893A), MSH2(K675A)-MSH3(wt), and MSH2(K675A)-MSH3(K893A) where 'wt' denotes wild-type. To check for expression, 1 mL of cells was collected in a microcentrifuge tube following 48 hours of infection and centrifuged for 5 minutes at 1000 x g. The pellet was resuspended in 150 µL H_20, 100 uL PBS, and 50 µL SDS dye (350 mM Tris-HCl (pH 6.8), 30% glycerol, 10% SDS (w/v), 600 mM DTT and a tinch of bromophenol blue), boiled for 10 minutes and centrifuged for 10 minutes at 12,000 x g. 10 µL of the solution was loaded on an 8% polyacrylamide gel to verify expression of the protein (Fig. 10). Following verification, infected Sf9 cells were harvested and suspended in wash buffer A (150 mM NaCl, 25 mM HEPES-NaOH (pH 8.1), 10% glycerol, 0.5 mM EDTA and a protease inhibitor cocktail (0.5 mM phenylmethyl sulfonyl fluoride (PMSF), 0.8

ug/mL pepstatin and 0.8 ug/mL leupeptin), pelleted and resuspended in freeze buffer B (300 mM NaCl, 20 mM imidazole, 25 mM HEPES-NaOH (pH 8.1), 10% glycerol, and the protease inhibitor cocktail followed by flash freezing in liquid nitrogen and storage at -80 ºC .

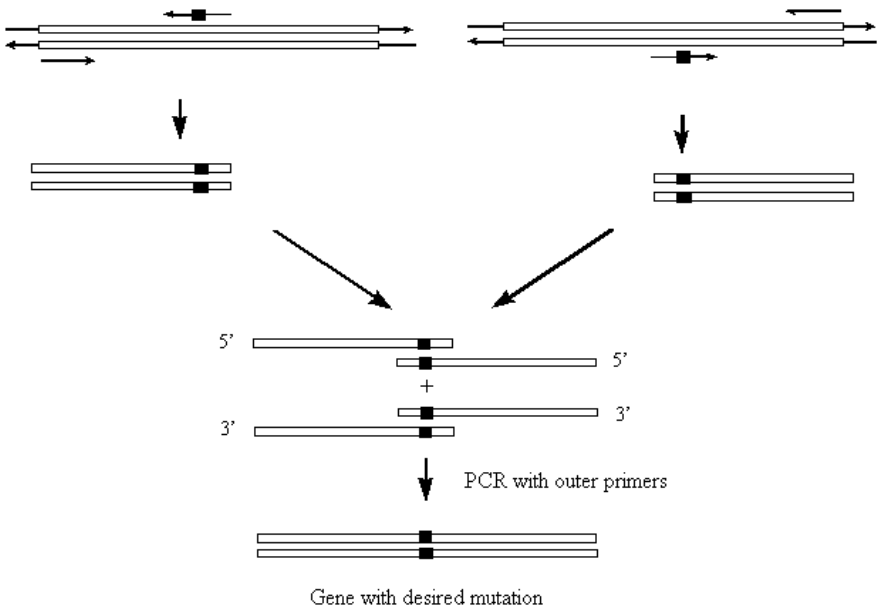
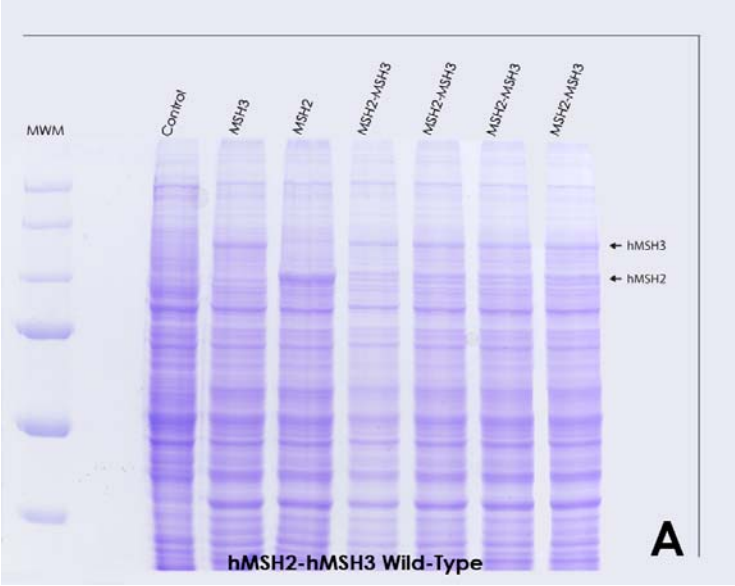
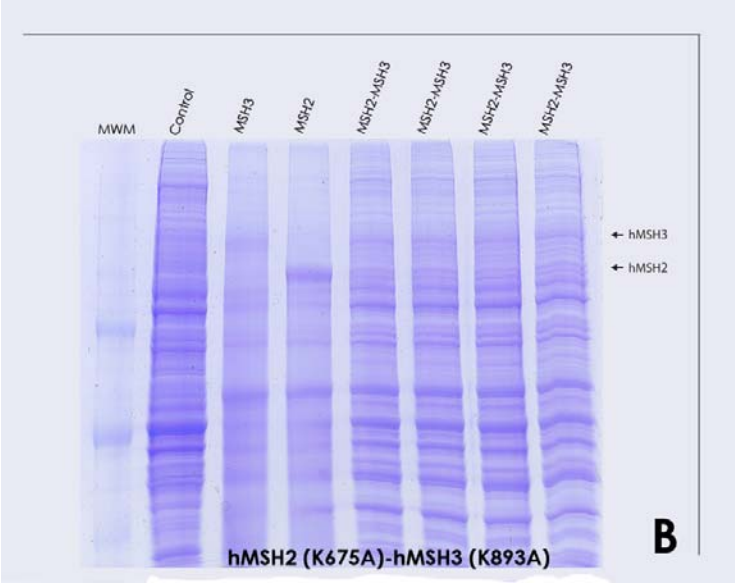


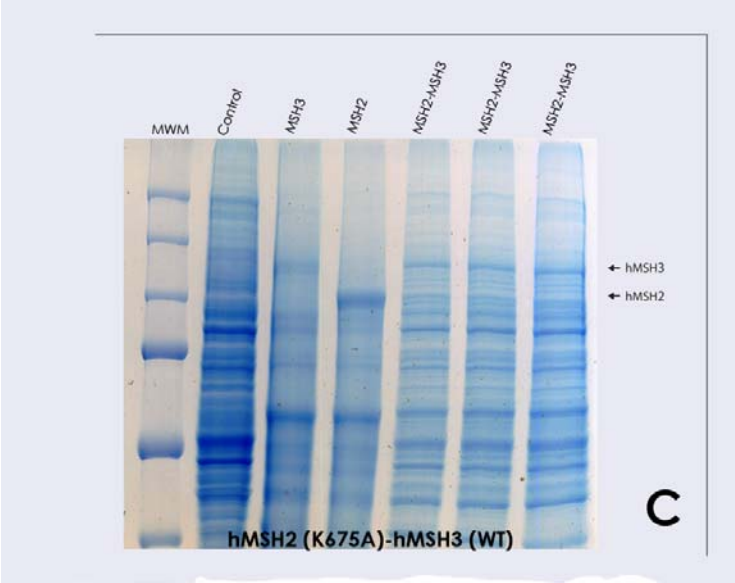
Figure 3. **Schematic of Site Directed Mutagenesis by Overlapping PCR.** Outside primers carrying unique restriction sequences are combined with internal primers flanking the mutation site with the desired codon change in a PCR reaction. The products of which are used in equal proportion as template for a PCR reaction utilizing the outside primers. The result is a section of the gene carrying the desired mutation which is then ligated to the original gene. (Figure reproduced from M. Blaber, Florida State University

The mutations created in MSH2 and MSH3 were done specifically with the intention of studying how each subunit of the dimer binds and hydrolyzes ATP upon IDL recognition in protein-crosslinking experiments. Each mutation was made in the Walker A nucleotide-binding motif previously described. These mutations will provide

information regarding which subunit binds which nucleotide (ATP or ADP) and how these change before and after lesion binding. The lysine-to-arginine mutations in the Walker A region allow for nucleotide binding but prevent ATP hydrolysis, while the lysine-to-alanine mutations are deficient in nucleotide binding altogether.







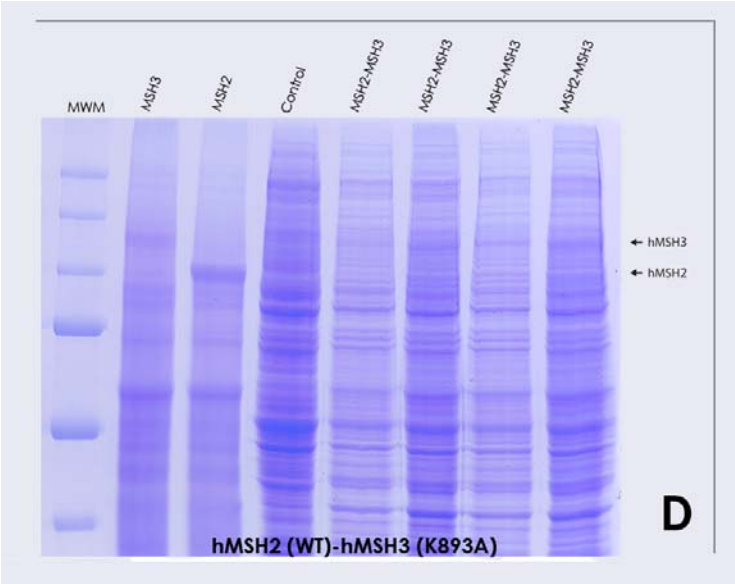
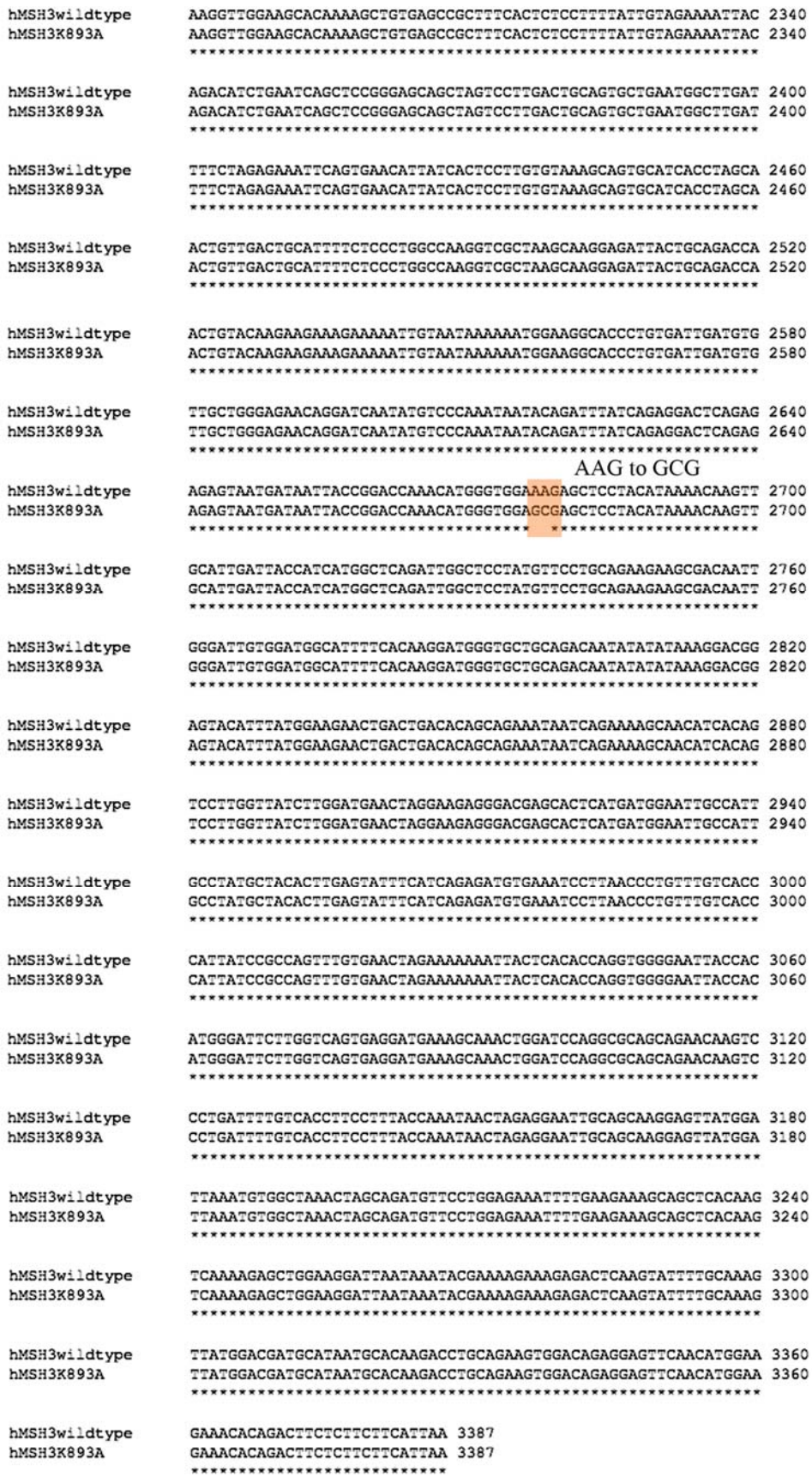


Figure 10. **Expression of Mutant MSH2-MSH3.** 8% Polyacrylamide gels showing protein expression in Sf9 cells 48 hours after infection with P3 viral stock. **(A)** MSH2-MSH3 wild type; **(B)** MSH2(K675A)-MSH3 (K893A); **(C)** MSH2(K675A)-MSH3(WT); **(D)** MSH2(WT)-MSH3(K893A)

```
CLUSTAL 2.0.11 multiple sequence alignment


hMSH3wildtype       ATGTCTCGCCGGAAGCCTGCGTCGGGCGGCCTCGCTGCCTCCAGCTCAGCCCCTGCGAGG 60
hMSH3K893A          ATGTCTCGCCGGAAGCCTGCGTCGGGCGGCCTCGCTGCCTCCAGCTCAGCCCCTGCGAGG 60
                    ************************************************************

hMSH3wildtype       CAAGCGGTTTTGAGCCGATTCTTCCAGTCTACGGGAAGCCTGAAATCCACCTCCTCCTCC 120
hMSH3K893A          CAAGCGGTTTTGAGCCGATTCTTCCAGTCTACGGGAAGCCTGAAATCCACCTCCTCCTCC 120
                    ************************************************************

hMSH3wildtype       ACAGGTGCAGCCGACCAGGTGGACCCTGGCGCTGCAGCGGCCGCAGCGCCCCCAGCGCCC 180
hMSH3K893A          ACAGGTGCAGCCGACCAGGTGGACCCTGGCGCTGCAGCGGCCGCAGCGCCCCCAGCGCCC 180
                    ************************************************************

hMSH3wildtype       GCCTTCCCGCCCCAGCTGCCGCCGCACGTAGCTACAGAAATTGACAGAAGAAAGAAGAGA 240
hMSH3K893A          GCCTTCCCGCCCCAGCTGCCGCCGCACGTAGCTACAGAAATTGACAGAAGAAAGAAGAGA 240
                    ************************************************************

hMSH3wildtype       CCATTGGAAAATGATGGGCCTGTTAAAAAGAAAGTAAAGAAAGTCCAACAAAAGGAAGGA 300
hMSH3K893A          CCATTGGAAAATGATGGGCCTGTTAAAAAGAAAGTAAAGAAAGTCCAACAAAAGGAAGGA 300
                    ************************************************************

hMSH3wildtype       GGAAGTGATCTGGGAATGTCTGGCAACTCTGAGCCAAAGAAATGTCTGAGGACCAGGAAT 360
hMSH3K893A          GGAAGTGATCTGGGAATGTCTGGCAACTCTGAGCCAAAGAAATGTCTGAGGACCAGGAAT 360
                    ************************************************************

hMSH3wildtype       GTTTCAAAGTCTCTGGAAAAATTGAAAGAATTCTGCTGCGATTCTGCCCTTCCTCAAAGT 420
hMSH3K893A          GTTTCAAAGTCTCTGGAAAAATTGAAAGAATTCTGCTGCGATTCTGCCCTTCCTCAAAGT 420
                    ************************************************************

hMSH3wildtype       AGAGTCCAGACAGAATCTCTGCAGGAGAGATTTGCAGTTCTGCCAAAATGTACTGATTTT 480
hMSH3K893A          AGAGTCCAGACAGAATCTCTGCAGGAGAGATTTGCAGTTCTGCCAAAATGTACTGATTTT 480
                    ************************************************************

hMSH3wildtype       GATGATATCAGTCTTCTACACGCAAAGAATGCAGTTTCTTCTGAAGATTCGAAACGTCAA 540
hMSH3K893A          GATGATATCAGTCTTCTACACGCAAAGAATGCAGTTTCTTCTGAAGATTCGAAACGTCAA 540
                    ************************************************************

hMSH3wildtype       ATTAATCAAAAGGACACAACACTTTTTGATCTCAGTCAGTTTGGATCATCAAATACAAGT 600
hMSH3K893A          ATTAATCAAAAGGACACAACACTTTTTGATCTCAGTCAGTTTGGATCATCAAATACAAGT 600
                    ************************************************************

hMSH3wildtype       CATGAAAATTTACAGAAAACTGCTTCCAAATCAGCTAACAAACGGTCCAAAAGCATCTAT 660
hMSH3K893A          CATGAAAATTTACAGAAAACTGCTTCCAAATCAGCTAACAAACGGTCCAAAAGCATCTAT 660
                    ************************************************************

hMSH3wildtype       ACGCCGCTAGAATTACAATACATAGAAATGAAGCAGCAGCACAAAGATGCAGTTTTGTGT 720
hMSH3K893A          ACGCCGCTAGAATTACAATACATAGAAATGAAGCAGCAGCACAAAGATGCAGTTTTGTGT 720
                    ************************************************************

hMSH3wildtype       GTGGAATGTGGATATAAGTATAGATTCTTTGGGGAAGATGCAGAGATTGCAGCCCGAGAG 780
hMSH3K893A          GTGGAATGTGGATATAAGTATAGATTCTTTGGGGAAGATGCAGAGATTGCAGCCCGAGAG 780
                    ************************************************************

hMSH3wildtype       CTCAATATTTATTGCCATTTAGATCACAACTTTATGACAGCAAGTATACCTACTCACAGA 840
hMSH3K893A          CTCAATATTTATTGCCATTTAGATCACAACTTTATGACAGCAAGTATACCTACTCACAGA 840
                    ************************************************************

hMSH3wildtype       CTGTTTGTTCATGTACGCCGCCTGGTGGCAAAAGGATATAAGGTGGGAGTTGTGAAGCAA 900
hMSH3K893A          CTGTTTGTTCATGTACGCCGCCTGGTGGCAAAAGGATATAAGGTGGGAGTTGTGAAGCAA 900
                    ************************************************************

hMSH3wildtype       ACTGAAACTGCAGCATTAAAGGCCATTGGAGACAACAGAAGTTCACTCTTTTCCCGGAAA 960
hMSH3K893A          ACTGAAACTGCAGCATTAAAGGCCATTGGAGACAACAGAAGTTCACTCTTTTCCCGGAAA 960
                    ************************************************************

hMSH3wildtype       TTGACTGCCCTTTATACAAAATCTACACTTATTGGAGAAGATGTGAATCCCCTAATCAAG 1020
hMSH3K893A          TTGACTGCCCTTTATACAAAATCTACACTTATTGGAGAAGATGTGAATCCCCTAATCAAG 1020
                    ************************************************************

hMSH3wildtype       CTGGATGATGCTGTAAATGTTGATGAGATAATGACTGATACTTCTACCAGCTATCTTCTG 1080
hMSH3K893A          CTGGATGATGCTGTAAATGTTGATGAGATAATGACTGATACTTCTACCAGCTATCTTCTG 1080
                    ************************************************************

hMSH3wildtype       TGCATCTCTGAAAATAAGGAAAATGTTAGGGACAAAAAAAAAGGGCAACATTTTTATTGGC 1140
hMSH3K893A          TGCATCTCTGAAAATAAGGAAAATGTTAGGGACAAAAAAAAAGGGCAACATTTTTATTGGC 1140
                    ************************************************************
```

```
hMSH3wildtype       ATTGTGGGAGTGCAGCCTGCCACAGGCGAGGTTGTGTTTGATAGTTTCCAGGACTCTGCT 1200
hMSH3K893A          ATTGTGGGAGTGCAGCCTGCCACAGGCGAGGTTGTGTTTGATAGTTTCCAGGACTCTGCT 1200
                    ************************************************************

hMSH3wildtype       TCTCGTTCAGAGCTAGAAACCCGGATGTCAAGCCTGCAGCCAGTAGAGCTGCTGCTTCCT 1260
hMSH3K893A          TCTCGTTCAGAGCTAGAAACCCGGATGTCAAGCCTGCAGCCAGTAGAGCTGCTGCTTCCT 1260
                    ************************************************************

hMSH3wildtype       TCGGCCTTGTCCGAGCAAACAGAGGCGCTCATCCACAGAGCCACATCTGTTAGTGTGCAG 1320
hMSH3K893A          TCGGCCTTGTCCGAGCAAACAGAGGCGCTCATCCACAGAGCCACATCTGTTAGTGTGCAG 1320
                    ************************************************************

hMSH3wildtype       GATGACAGAATTCGAGTCGAAAGGATGGATAACATTTATTTTGAATACAGCCATGCTTTC 1380
hMSH3K893A          GATGACAGAATTCGAGTCGAAAGGATGGATAACATTTATTTTGAATACAGCCATGCTTTC 1380
                    ************************************************************

hMSH3wildtype       CAGGCAGTTACAGAGTTTTATGCAAAAGATACAGTTGACATCAAAGGTTCTCAAATTATT 1440
hMSH3K893A          CAGGCAGTTACAGAGTTTTATGCAAAAGATACAGTTGACATCAAAGGTTCTCAAATTATT 1440
                    ************************************************************

hMSH3wildtype       TCTGGCATTGTTAACTTAGAGAAGCCTGTGATTTGCTCTTTGGCTGCCATCATAAAATAC 1500
hMSH3K893A          TCTGGCATTGTTAACTTAGAGAAGCCTGTGATTTGCTCTTTGGCTGCCATCATAAAATAC 1500
                    ************************************************************

hMSH3wildtype       CTCAAAGAATTCAACTTGGAAAAGATGCTCTCCAAACCTGAGAATTTTAAACAGCTATCA 1560
hMSH3K893A          CTCAAAGAATTCAACTTGGAAAAGATGCTCTCCAAACCTGAGAATTTTAAACAGCTATCA 1560
                    ************************************************************

hMSH3wildtype       AGTAAAATGGAATTTATGACAATTAATGGAACAACATTAAGGAATCTGGAAATCCTACAG 1620
hMSH3K893A          AGTAAAATGGAATTTATGACAATTAATGGAACAACATTAAGGAATCTGGAAATCCTACAG 1620
                    ************************************************************

hMSH3wildtype       AATCAGACTGATATGAAAACCAAAGGAAGTTTGCTGTGGGTTTTAGACCACACTAAAACT 1680
hMSH3K893A          AATCAGACTGATATGAAAACCAAAGGAAGTTTGCTGTGGGTTTTAGACCACACTAAAACT 1680
                    ************************************************************

hMSH3wildtype       TCATTTGGGAGACGGAAGTTAAAGAAGTGGGTGACCCAGCCACTCCTTAAATTAAGGGAA 1740
hMSH3K893A          TCATTTGGGAGACGGAAGTTAAAGAAGTGGGTGACCCAGCCACTCCTTAAATTAAGGGAA 1740
                    ************************************************************

hMSH3wildtype       ATAAATGCCCGGCTTGATGCTGTATCGGAAGTTCTCCATTCAGAATCTAGTGTGTTTGGT 1800
hMSH3K893A          ATAAATGCCCGGCTTGATGCTGTATCGGAAGTTCTCCATTCAGAATCTAGTGTGTTTGGT 1800
                    ************************************************************

hMSH3wildtype       CAGATAGAAAATCATCTACGTAAATTGCCCGACATAGAGAGGGGACTCTGTAGCATTTAT 1860
hMSH3K893A          CAGATAGAAAATCATCTACGTAAATTGCCCGACATAGAGAGGGGACTCTGTAGCATTTAT 1860
                    ************************************************************

hMSH3wildtype       CACAAAAAATGTTCTACCCAAGAGTTCTTCTTGATTGTCAAAACTTTATATCACCTAAAG 1920
hMSH3K893A          CACAAAAAATGTTCTACCCAAGAGTTCTTCTTGATTGTCAAAACTTTATATCACCTAAAG 1920
                    ************************************************************

hMSH3wildtype       TCAGAATTTCAAGCAATAATACCTGCTGTTAATTCCCACATTCAGTCAGACTTGCTCCGG 1980
hMSH3K893A          TCAGAATTTCAAGCAATAATACCTGCTGTTAATTCCCACATTCAGTCAGACTTGCTCCGG 1980
                    ************************************************************

hMSH3wildtype       ACCGTTATTTTAGAAATTCCTGAACTCCTCAGTCCAGTGGAGCATTACTTAAAGATACTC 2040
hMSH3K893A          ACCGTTATTTTAGAAATTCCTGAACTCCTCAGTCCAGTGGAGCATTACTTAAAGATACTC 2040
                    ************************************************************

hMSH3wildtype       AATGAACAAGCTGCCAAAGTTGGGGATAAAACTGAATTATTTAAAGACCTTTCTGACTTC 2100
hMSH3K893A          AATGAACAAGCTGCCAAAGTTGGGGATAAAACTGAATTATTTAAAGACCTTTCTGACTTC 2100
                    ************************************************************

hMSH3wildtype       CCTTTAATAAAAAAGAGGAAGGATGAAATTCAAGGTGTTATTGACGAGATCCGAATGCAT 2160
hMSH3K893A          CCTTTAATAAAAAAGAGGAAGGATGAAATTCAAGGTGTTATTGACGAGATCCGAATGCAT 2160
                    ************************************************************

hMSH3wildtype       TTGCAAGAAATACGAAAAATACTAAAAAATCCTTCTGCACAATATGTGACAGTATCAGGA 2220
hMSH3K893A          TTGCAAGAAATACGAAAAATACTAAAAAATCCTTCTGCACAATATGTGACAGTATCAGGA 2220
                    ************************************************************

hMSH3wildtype       CAGGAGTTTATGATAGAAATAAAGAACTCTGCTGTATCTTGTATACCAACTGATTGGGTA 2280
hMSH3K893A          CAGGAGTTTATGATAGAAATAAAGAACTCTGCTGTATCTTGTATACCAACTGATTGGGTA 2280
                    ************************************************************
```


hMSH3wildtype AAGGTTGGAAGCACAAAAGCTGTGAGCCGCTTTCACTCTCCTTTTATTGTAGAAAATTAC 2340
hMSH3K893A AAGGTTGGAAGCACAAAAGCTGTGAGCCGCTTTCACTCTCCTTTTATTGTAGAAAATTAC 2340
hMSH3wildtype AGACATCTGAATCAGCTCCGGGAGCAGCTAGTCCTTGACTGCAGTGCTGAATGGCTTGAT 2400
hMSH3K893A AGACATCTGAATCAGCTCCGGGAGCAGCTAGTCCTTGACTGCAGTGCTGAATGGCTTGAT 2400
hMSH3wildtype TTTCTAGAGAAATTCAGTGAACATTATCACTCCTTGTGTAAAGCAGTGCATCACCTAGCA 2460
hMSH3K893A TTTCTAGAGAAATTCAGTGAACATTATCACTCCTTGTGTAAAGCAGTGCATCACCTAGCA 2460
hMSH3wildtype ACTGTTGACTGCATTTTCTCCCTGGCCAAGGTCGCTAAGCAAGGAGATTACTGCAGACCA 2520
hMSH3K893A ACTGTTGACTGCATTTTCTCCCTGGCCAAGGTCGCTAAGCAAGGAGATTACTGCAGACCA 2520
hMSH3wildtype ACTGTACAAGAAGAAAGAAAAATTGTAATAAAAAATGGAAGGCACCCTGTGATTGATGTG 2580
hMSH3K893A ACTGTACAAGAAGAAAGAAAAATTGTAATAAAAAATGGAAGGCACCCTGTGATTGATGTG 2580
hMSH3wildtype TTGCTGGGAGAACAGGATCAATATGTCCCAAATAATACAGATTTATCAGAGGACTCAGAG 2640
hMSH3K893A TTGCTGGGAGAACAGGATCAATATGTCCCAAATAATACAGATTTATCAGAGGACTCAGAG 2640
AAG to GCG
hMSH3wildtype AGAGTAATGATAATTACCGGACCAAACATGGGTGGAAAGAGCTCCTACATAAAACAAGTT 2700
hMSH3K893A AGAGTAATGATAATTACCGGACCAAACATGGGTGGAGCGAGCTCCTACATAAAACAAGTT 2700
hMSH3wildtype GCATTGATTACCATCATGGCTCAGATTGGCTCCTATGTTCCTGCAGAAGAAGCGACAATT 2760
hMSH3K893A GCATTGATTACCATCATGGCTCAGATTGGCTCCTATGTTCCTGCAGAAGAAGCGACAATT 2760
hMSH3wildtype GGGATTGTGGATGGCATTTTCACAAGGATGGGTGCTGCAGACAATATATATAAAGGACGG 2820
hMSH3K893A GGGATTGTGGATGGCATTTTCACAAGGATGGGTGCTGCAGACAATATATATAAAGGACGG 2820
hMSH3wildtype AGTACATTTATGGAAGAACTGACTGACACAGCAGAAATAATCAGAAAAGCAACATCACAG 2880
hMSH3K893A AGTACATTTATGGAAGAACTGACTGACACAGCAGAAATAATCAGAAAAGCAACATCACAG 2880
hMSH3wildtype TCCTTGGTTATCTTGGATGAACTAGGAAGAGGGACGAGCACTCATGATGGAATTGCCATT 2940
hMSH3K893A TCCTTGGTTATCTTGGATGAACTAGGAAGAGGGACGAGCACTCATGATGGAATTGCCATT 2940
hMSH3wildtype GCCTATGCTACACTTGAGTATTTCATCAGAGATGTGAAATCCTTAACCCTGTTTGTCACC 3000
hMSH3K893A GCCTATGCTACACTTGAGTATTTCATCAGAGATGTGAAATCCTTAACCCTGTTTGTCACC 3000
hMSH3wildtype CATTATCCGCCAGTTTGTGAACTAGAAAAAAATTACTCACACCAGGTGGGGAATTACCAC 3060
hMSH3K893A CATTATCCGCCAGTTTGTGAACTAGAAAAAAATTACTCACACCAGGTGGGGAATTACCAC 3060
hMSH3wildtype ATGGGATTCTTGGTCAGTGAGGATGAAAGCAAACTGGATCCAGGCGCAGCAGAACAAGTC 3120
hMSH3K893A ATGGGATTCTTGGTCAGTGAGGATGAAAGCAAACTGGATCCAGGCGCAGCAGAACAAGTC 3120
hMSH3wildtype CCTGATTTTGTCACCTTCCTTTACCAAATAACTAGAGGAATTGCAGCAAGGAGTTATGGA 3180
hMSH3K893A CCTGATTTTGTCACCTTCCTTTACCAAATAACTAGAGGAATTGCAGCAAGGAGTTATGGA 3180
hMSH3wildtype TTAAATGTGGCTAAACTAGCAGATGTTCCTGGAGAAATTTTGAAGAAAGCAGCTCACAAG 3240
hMSH3K893A TTAAATGTGGCTAAACTAGCAGATGTTCCTGGAGAAATTTTGAAGAAAGCAGCTCACAAG 3240
hMSH3wildtype TCAAAAGAGCTGGAAGGATTAATAAATACGAAAAGAAAGAGACTCAAGTATTTTGCAAAG 3300
hMSH3K893A TCAAAAGAGCTGGAAGGATTAATAAATACGAAAAGAAAGAGACTCAAGTATTTTGCAAAG 3300
hMSH3wildtype TTATGGACGATGCATAATGCACAAGACCTGCAGAAGTGGACAGAGGAGTTCAACATGGAA 3360
hMSH3K893A TTATGGACGATGCATAATGCACAAGACCTGCAGAAGTGGACAGAGGAGTTCAACATGGAA 3360
hMSH3wildtype GAAACACAGACTTCTCTTCTTCATTAA 3387
hMSH3K893A GAAACACAGACTTCTCTTCTTCATTAA 3387

```
CLUSTAL 2.0.11 multiple sequence alignment

hMSH3wildtype       ATGTCTCGCCGGAAGCCTGCGTCGGGCGGCCTCGCTGCCTCCAGCTCAGCCCCTGCGAGG 60
hMSH3K893R          ATGTCTCGCCGGAAGCCTGCGTCGGGCGGCCTCGCTGCCTCCAGCTCAGCCCCTGCGAGG 60
                    ************************************************************

hMSH3wildtype       CAAGCGGTTTTGAGCCGATTCTTCCAGTCTACGGGAAGCCTGAAATCCACCTCCTCCTCC 120
hMSH3K893R          CAAGCGGTTTTGAGCCGATTCTTCCAGTCTACGGGAAGCCTGAAATCCACCTCCTCCTCC 120
                    ************************************************************

hMSH3wildtype       ACAGGTGCAGCCGACCAGGTGGACCCTGGCGCTGCAGCGGCCGCAGCGCCCCCAGCGCCC 180
hMSH3K893R          ACAGGTGCAGCCGACCAGGTGGACCCTGGCGCTGCAGCGGCCGCAGCGCCCCCAGCGCCC 180
                    ************************************************************

hMSH3wildtype       GCCTTCCCGCCCCAGCTGCCGCCGCACGTAGCTACAGAAATTGACAGAAGAAAGAAGAGA 240
hMSH3K893R          GCCTTCCCGCCCCAGCTGCCGCCGCACGTAGCTACAGAAATTGACAGAAGAAAGAAGAGA 240
                    ************************************************************

hMSH3wildtype       CCATTGGAAAATGATGGGCCTGTTAAAAAGAAAGTAAAGAAAGTCCAACAAAAGGAAGGA 300
hMSH3K893R          CCATTGGAAAATGATGGGCCTGTTAAAAAGAAAGTAAAGAAAGTCCAACAAAAGGAAGGA 300
                    ************************************************************

hMSH3wildtype       GGAAGTGATCTGGGAATGTCTGGCAACTCTGAGCCAAAGAAATGTCTGAGGACCAGGAAT 360
hMSH3K893R          GGAAGTGATCTGGGAATGTCTGGCAACTCTGAGCCAAAGAAATGTCTGAGGACCAGGAAT 360
                    ************************************************************

hMSH3wildtype       GTTTCAAAGTCTCTGGAAAAATTGAAAGAATTCTGCTGCGATTCTGCCCTTCCTCAAAGT 420
hMSH3K893R          GTTTCAAAGTCTCTGGAAAAATTGAAAGAATTCTGCTGCGATTCTGCCCTTCCTCAAAGT 420
                    ************************************************************

hMSH3wildtype       AGAGTCCAGACAGAATCTCTGCAGGAGAGATTTGCAGTTCTGCCAAAATGTACTGATTTT 480
hMSH3K893R          AGAGTCCAGACAGAATCTCTGCAGGAGAGATTTGCAGTTCTGCCAAAATGTACTGATTTT 480
                    ************************************************************

hMSH3wildtype       GATGATATCAGTCTTCTACACGCAAAGAATGCAGTTTCTTCTGAAGATTCGAAACGTCAA 540
hMSH3K893R          GATGATATCAGTCTTCTACACGCAAAGAATGCAGTTTCTTCTGAAGATTCGAAACGTCAA 540
                    ************************************************************

hMSH3wildtype       ATTAATCAAAAGGACACAACACTTTTTGATCTCAGTCAGTTTGGATCATCAAATACAAGT 600
hMSH3K893R          ATTAATCAAAAGGACACAACACTTTTTGATCTCAGTCAGTTTGGATCATCAAATACAAGT 600
                    ************************************************************

hMSH3wildtype       CATGAAAATTTACAGAAAACTGCTTCCAAATCAGCTAACAAACGGTCCAAAAGCATCTAT 660
hMSH3K893R          CATGAAAATTTACAGAAAACTGCTTCCAAATCAGCTAACAAACGGTCCAAAAGCATCTAT 660
                    ************************************************************

hMSH3wildtype       ACGCCGCTAGAATTACAATACATAGAAATGAAGCAGCAGCACAAAGATGCAGTTTTGTGT 720
hMSH3K893R          ACGCCGCTAGAATTACAATACATAGAAATGAAGCAGCAGCACAAAGATGCAGTTTTGTGT 720
                    ************************************************************

hMSH3wildtype       GTGGAATGTGGATATAAGTATAGATTCTTTGGGGAAGATGCAGAGATTGCAGCCCGAGAG 780
hMSH3K893R          GTGGAATGTGGATATAAGTATAGATTCTTTGGGGAAGATGCAGAGATTGCAGCCCGAGAG 780
                    ************************************************************

hMSH3wildtype       CTCAATATTTATTGCCATTTAGATCACAACTTTATGACAGCAAGTATACCTACTCACAGA 840
hMSH3K893R          CTCAATATTTATTGCCATTTAGATCACAACTTTATGACAGCAAGTATACCTACTCACAGA 840
                    ************************************************************

hMSH3wildtype       CTGTTTGTTCATGTACGCCGCCTGGTGGCAAAAGGATATAAGGTGGGAGTTGTGAAGCAA 900
hMSH3K893R          CTGTTTGTTCATGTACGCCGCCTGGTGGCAAAAGGATATAAGGTGGGAGTTGTGAAGCAA 900
                    ************************************************************

hMSH3wildtype       ACTGAAACTGCAGCATTAAAGGCCATTGGAGACAACAGAAGTTCACTCTTTTCCCGGAAA 960
hMSH3K893R          ACTGAAACTGCAGCATTAAAGGCCATTGGAGACAACAGAAGTTCACTCTTTTCCCGGAAA 960
                    ************************************************************

hMSH3wildtype       TTGACTGCCCTTTATACAAAATCTACACTTATTGGAGAAGATGTGAATCCCCTAATCAAG 1020
hMSH3K893R          TTGACTGCCCTTTATACAAAATCTACACTTATTGGAGAAGATGTGAATCCCCTAATCAAG 1020
                    ************************************************************

hMSH3wildtype       CTGGATGATGCTGTAAATGTTGATGAGATAATGACTGATACTTCTACCAGCTATCTTCTG 1080
hMSH3K893R          CTGGATGATGCTGTAAATGTTGATGAGATAATGACTGATACTTCTACCAGCTATCTTCTG 1080
                    ************************************************************

hMSH3wildtype       TGCATCTCTGAAAATAAGGAAAATGTTAGGGACAAAAAAAAGGGCAACATTTTTATTGGC 1140
hMSH3K893R          TGCATCTCTGAAAATAAGGAAAATGTTAGGGACAAAAAAAAGGGCAACATTTTTATTGGC 1140
                    ************************************************************
```

```
hMSH3wildtype       ATTGTGGGAGTGCAGCCTGCCACAGGCGAGGTTGTGTTTGATAGTTTCCAGGACTCTGCT 1200
hMSH3K893R          ATTGTGGGAGTGCAGCCTGCCACAGGCGAGGTTGTGTTTGATAGTTTCCAGGACTCTGCT 1200
                    ************************************************************

hMSH3wildtype       TCTCGTTCAGAGCTAGAAACCCGGATGTCAAGCCTGCAGCCAGTAGAGCTGCTGCTTCCT 1260
hMSH3K893R          TCTCGTTCAGAGCTAGAAACCCGGATGTCAAGCCTGCAGCCAGTAGAGCTGCTGCTTCCT 1260
                    ************************************************************

hMSH3wildtype       TCGGCCTTGTCCGAGCAAACAGAGGCGCTCATCCACAGAGCCACATCTGTTAGTGTGCAG 1320
hMSH3K893R          TCGGCCTTGTCCGAGCAAACAGAGGCGCTCATCCACAGAGCCACATCTGTTAGTGTGCAG 1320
                    ************************************************************

hMSH3wildtype       GATGACAGAATTCGAGTCGAAAGGATGGATAACATTTATTTTGAATACAGCCATGCTTTC 1380
hMSH3K893R          GATGACAGAATTCGAGTCGAAAGGATGGATAACATTTATTTTGAATACAGCCATGCTTTC 1380
                    ************************************************************

hMSH3wildtype       CAGGCAGTTACAGAGTTTTATGCAAAAGATACAGTTGACATCAAAGGTTCTCAAATTATT 1440
hMSH3K893R          CAGGCAGTTACAGAGTTTTATGCAAAAGATACAGTTGACATCAAAGGTTCTCAAATTATT 1440
                    ************************************************************

hMSH3wildtype       TCTGGCATTGTTAACTTAGAGAAGCCTGTGATTTGCTCTTTGGCTGCCATCATAAAATAC 1500
hMSH3K893R          TCTGGCATTGTTAACTTAGAGAAGCCTGTGATTTGCTCTTTGGCTGCCATCATAAAATAC 1500
                    ************************************************************

hMSH3wildtype       CTCAAAGAATTCAACTTGGAAAAGATGCTCTCCAAACCTGAGAATTTTAAACAGCTATCA 1560
hMSH3K893R          CTCAAAGAATTCAACTTGGAAAAGATGCTCTCCAAACCTGAGAATTTTAAACAGCTATCA 1560
                    ************************************************************

hMSH3wildtype       AGTAAAATGGAATTTATGACAATTAATGGAACAACATTAAGGAATCTGGAAATCCTACAG 1620
hMSH3K893R          AGTAAAATGGAATTTATGACAATTAATGGAACAACATTAAGGAATCTGGAAATCCTACAG 1620
                    ************************************************************

hMSH3wildtype       AATCAGACTGATATGAAAACCAAAGGAAGTTTGCTGTGGGTTTTAGACCACACTAAAACT 1680
hMSH3K893R          AATCAGACTGATATGAAAACCAAAGGAAGTTTGCTGTGGGTTTTAGACCACACTAAAACT 1680
                    ************************************************************

hMSH3wildtype       TCATTTGGGAGACGGAAGTTAAAGAAGTGGGTGACCCAGCCACTCCTTAAATTAAGGGAA 1740
hMSH3K893R          TCATTTGGGAGACGGAAGTTAAAGAAGTGGGTGACCCAGCCACTCCTTAAATTAAGGGAA 1740
                    ************************************************************

hMSH3wildtype       ATAAATGCCCGGCTTGATGCTGTATCGGAAGTTCTCCATTCAGAATCTAGTGTGTTTGGT 1800
hMSH3K893R          ATAAATGCCCGGCTTGATGCTGTATCGGAAGTTCTCCATTCAGAATCTAGTGTGTTTGGT 1800
                    ************************************************************

hMSH3wildtype       CAGATAGAAAATCATCTACGTAAATTGCCCGACATAGAGAGGGGACTCTGTAGCATTTAT 1860
hMSH3K893R          CAGATAGAAAATCATCTACGTAAATTGCCCGACATAGAGAGGGGACTCTGTAGCATTTAT 1860
                    ************************************************************

hMSH3wildtype       CACAAAAAATGTTCTACCCAAGAGTTCTTCTTGATTGTCAAAACTTTATATCACCTAAAG 1920
hMSH3K893R          CACAAAAAATGTTCTACCCAAGAGTTCTTCTTGATTGTCAAAACTTTATATCACCTAAAG 1920
                    ************************************************************

hMSH3wildtype       TCAGAATTTCAAGCAATAATACCTGCTGTTAATTCCCACATTCAGTCAGACTTGCTCCGG 1980
hMSH3K893R          TCAGAATTTCAAGCAATAATACCTGCTGTTAATTCCCACATTCAGTCAGACTTGCTCCGG 1980
                    ************************************************************

hMSH3wildtype       ACCGTTATTTTAGAAATTCCTGAACTCCTCAGTCCAGTGGAGCATTACTTAAAGATACTC 2040
hMSH3K893R          ACCGTTATTTTAGAAATTCCTGAACTCCTCAGTCCAGTGGAGCATTACTTAAAGATACTC 2040
                    ************************************************************

hMSH3wildtype       AATGAACAAGCTGCCAAAGTTGGGGATAAAACTGAATTATTTAAAGACCTTTCTGACTTC 2100
hMSH3K893R          AATGAACAAGCTGCCAAAGTTGGGGATAAAACTGAATTATTTAAAGACCTTTCTGACTTC 2100
                    ************************************************************

hMSH3wildtype       CCTTTAATAAAAAAGAGGAAGGATGAAATTCAAGGTGTTATTGACGAGATCCGAATGCAT 2160
hMSH3K893R          CCTTTAATAAAAAAGAGGAAGGATGAAATTCAAGGTGTTATTGACGAGATCCGAATGCAT 2160
                    ************************************************************

hMSH3wildtype       TTGCAAGAAATACGAAAAATACTAAAAAATCCTTCTGCACAATATGTGACAGTATCAGGA 2220
hMSH3K893R          TTGCAAGAAATACGAAAAATACTAAAAAATCCTTCTGCACAATATGTGACAGTATCAGGA 2220
                    ************************************************************

hMSH3wildtype       CAGGAGTTTATGATAGAAATAAAGAACTCTGCTGTATCTTGTATACCAACTGATTGGGTA 2280
hMSH3K893R          CAGGAGTTTATGATAGAAATAAAGAACTCTGCTGTATCTTGTATACCAACTGATTGGGTA 2280
                    ************************************************************
```


hMSH3wildtype AAGGTTGGAAGCACAAAAGCTGTGAGCCGCTTTCACTCTCCTTTTATTGTAGAAAATTAC 2340
hMSH3K893R AAGGTTGGAAGCACAAAAGCTGTGAGCCGCTTTCACTCTCCTTTTATTGTAGAAAATTAC 2340
hMSH3wildtype AGACATCTGAATCAGCTCCGGGAGCAGCTAGTCCTTGACTGCAGTGCTGAATGGCTTGAT 2400
hMSH3K893R AGACATCTGAATCAGCTCCGGGAGCAGCTAGTCCTTGACTGCAGTGCTGAATGGCTTGAT 2400
hMSH3wildtype TTTCTAGAGAAATTCAGTGAACATTATCACTCCTTGTGTAAAGCAGTGCATCACCTAGCA 2460
hMSH3K893R TTTCTAGAGAAATTCAGTGAACATTATCACTCCTTGTGTAAAGCAGTGCATCACCTAGCA 2460
hMSH3wildtype ACTGTTGACTGCATTTTCTCCCTGGCCAAGGTCGCTAAGCAAGGAGATTACTGCAGACCA 2520
hMSH3K893R ACTGTTGACTGCATTTTCTCCCTGGCCAAGGTCGCTAAGCAAGGAGATTACTGCAGACCA 2520
hMSH3wildtype ACTGTACAAGAAGAAAGAAAAATTGTAATAAAAAATGGAAGGCACCCTGTGATTGATGTG 2580
hMSH3K893R ACTGTACAAGAAGAAAGAAAAATTGTAATAAAAAATGGAAGGCACCCTGTGATTGATGTG 2580
hMSH3wildtype TTGCTGGGAGAACAGGATCAATATGTCCCAAATAATACAGATTTATCAGAGGACTCAGAG 2640
hMSH3K893R TTGCTGGGAGAACAGGATCAATATGTCCCAAATAATACAGATTTATCAGAGGACTCAGAG 2640
AAG to AGA
hMSH3wildtype AGAGTAATGATAATTACCGGACCAAACATGGGTGGAAAGAGCTCCTACATAAAACAAGTT 2700
hMSH3K893R AGAGTAATGATAATTACCGGACCAAACATGGGTGGAAGAAGCTCCTACATAAAACAAGTT 2700
hMSH3wildtype GCATTGATTACCATCATGGCTCAGATTGGCTCCTATGTTCCTGCAGAAGAAGCGACAATT 2760
hMSH3K893R GCATTGATTACCATCATGGCTCAGATTGGCTCCTATGTTCCTGCAGAAGAAGCGACAATT 2760
hMSH3wildtype GGGATTGTGGATGGCATTTTCACAAGGATGGGTGCTGCAGACAATATATATAAAGGACGG 2820
hMSH3K893R GGGATTGTGGATGGCATTTTCACAAGGATGGGTGCTGCAGACAATATATATAAAGGACGG 2820
hMSH3wildtype AGTACATTTATGGAAGAACTGACTGACACAGCAGAAATAATCAGAAAAGCAACATCACAG 2880
hMSH3K893R AGTACATTTATGGAAGAACTGACTGACACAGCAGAAATAATCAGAAAAGCAACATCACAG 2880
hMSH3wildtype TCCTTGGTTATCTTGGATGAACTAGGAAGAGGGACGAGCACTCATGATGGAATTGCCATT 2940
hMSH3K893R TCCTTGGTTATCTTGGATGAACTAGGAAGAGGGACGAGCACTCATGATGGAATTGCCATT 2940
hMSH3wildtype GCCTATGCTACACTTGAGTATTTCATCAGAGATGTGAAATCCTTAACCCTGTTTGTCACC 3000
hMSH3K893R GCCTATGCTACACTTGAGTATTTCATCAGAGATGTGAAATCCTTAACCCTGTTTGTCACC 3000
hMSH3wildtype CATTATCCGCCAGTTTGTGAACTAGAAAAAAATTACTCACACCAGGTGGGGAATTACCAC 3060
hMSH3K893R CATTATCCGCCAGTTTGTGAACTAGAAAAAAATTACTCACACCAGGTGGGGAATTACCAC 3060
hMSH3wildtype ATGGGATTCTTGGTCAGTGAGGATGAAAGCAAACTGGATCCAGGCGCAGCAGAACAAGTC 3120
hMSH3K893R ATGGGATTCTTGGTCAGTGAGGATGAAAGCAAACTGGATCCAGGCGCAGCAGAACAAGTC 3120
hMSH3wildtype CCTGATTTTGTCACCTTCCTTTACCAAATAACTAGAGGAATTGCAGCAAGGAGTTATGGA 3180
hMSH3K893R CCTGATTTTGTCACCTTCCTTTACCAAATAACTAGAGGAATTGCAGCAAGGAGTTATGGA 3180
hMSH3wildtype TTAAATGTGGCTAAACTAGCAGATGTTCCTGGAGAAATTTTGAAGAAAGCAGCTCACAAG 3240
hMSH3K893R TTAAATGTGGCTAAACTAGCAGATGTTCCTGGAGAAATTTTGAAGAAAGCAGCTCACAAG 3240
hMSH3wildtype TCAAAAGAGCTGGAAGGATTAATAAATACGAAAAGAAAGAGACTCAAGTATTTTGCAAAG 3300
hMSH3K893R TCAAAAGAGCTGGAAGGATTAATAAATACGAAAAGAAAGAGACTCAAGTATTTTGCAAAG 3300
hMSH3wildtype TTATGGACGATGCATAATGCACAAGACCTGCAGAAGTGGACAGAGGAGTTCAACATGGAA 3360
hMSH3K893R TTATGGACGATGCATAATGCACAAGACCTGCAGAAGTGGACAGAGGAGTTCAACATGGAA 3360
hMSH3wildtype GAAACACAGACTTCTCTTCTTCATTAA 3387
hMSH3K893R GAAACACAGACTTCTCTTCTTCATTAA 3387

Figure 9. **Sequences of mutant MSH3 (K893R) and MSH3 (K893A) showing codon mutations.** The codon for lysine 893 in MSH3 is AAG; it was mutated to AGA in MSH3(K893R) and to GCG in MSH (K893A). Sequences were aligned against wild-type MSH3 to scan for random mutations using ClustalW2 software.

CFP-MSH3 hybrid and Conformational Transitions

The discovery and purification of the fluorescent jellyfish proteins has allowed biochemists to look into nanometer-scale conformational transitions associated with protein-DNA interaction, along with a host of other applications. Mutations in the original GFP (green fluorescent protein) have provided a variation of fluorescent proteins with variable excitation and emission spectra[37,38]. These physical characteristics of the fluorescent proteins have been utilized using fluorescence resonance energy transfer (FRET). Resonance energy transfer occurs whenever the emission spectrum of a fluorophore, called the donor, overlaps with the absorption spectrum of another molecule, called the acceptor[39]. The extent of the spectral overlap between the donor and acceptor, as well as the distance between them determines the extent of energy transfer. The distance between the donor and acceptor pair can be calculated directly from the measured transfer efficiency of the pair, providing an *in vitro* "spectroscopic ruler" with which to measure distances between sites on proteins tagged with fluorophores[40].

CFP (cyan fluorescent proteins) and YFP (yellow fluorescent protein) are ideal candidates to study molecular interactions due to their high photostability, and the fact that their adsorption and emission spectra are at favorably long wavelengths[41]. The donor and acceptor need to be within 15-60 Å for FRET to occur, which is comparable to the size of biological macromolecules. MSH2-MSH3 is an ideal candidate to study to how the protein changes conformations upon binding to IDLs as well as the trinucleotide

repeats found in HD. The sensitivity of FRET is such that we will be able to determine how these conformational changes may influence other biochemical characteristics of the protein such as that seen when MSH2-MSH3 binds to CAG hairpin loops[29]. To characterize and understand physical changes that occur upon lesion recognition, a CFP-MSH3 hybrid was constructed, and will be coexpressed and purified with a YFP-MSH2 hybrid constructed previously by Sarah Javaid.

To construct this hybrid, a pECFP-C1 vector containing the CFP gene was purchased from Clontech. PCR primers were designed and purchased from IDT. The forward primer (5'-CCC GCA TGC CGC CACC ATG CAC CAC CAC CAC CAC CAC ATG GTG AGC AAG GGC-3') introduces an SphI restriction site, as well as a Kozak consensus translation initiation site for increased translational efficiency[42]. Also incorporated is a His_6 tag for purification on a nickel column, all on the N-terminus of the CFP gene. The reverse primer (5'-CCC CAT ATG ACC ACC CTT GTA CAG CTC GTC CAT GCC) will bind to the last 18 base pairs of the CFP gene, remove the stop codon, and place two glycine residues to increase protein mobility and degrees of freedom.

A PCR reaction was set up utilizing these primers and the pECFP-C1 vector as template. The resulting 759 bp band was gel extracted and purified, digested with both SphI and NdeI (New England Biolabs) and gel purified again. Simultaneously, a pET29a vector harboring the *MSH3* gene cloned between *Sph*I and *Nde*I was digested with the same enzymes, and gel purified. The resulting insert and vector were ligated overnight at 16 ºC and transformed into XL-1 blue competent *E. coli* cells (Stratagene) and plated on agar plates containing kanamycin (60 μg/mL). Plasmid DNA was isolated from selected

colonies and analyzed for the CFP insert by linearization with *Sph*I and observing a 10 kb band when subjected to gel electrophoresis. A pFastBac1 vector and the Pet29a-CFP-MSH3 vector were both digested with *Sph*I and *Hind*III. The appropriate bands were gel purified and ligated to create the CFP-MSH3 hybrid in pFastBac1. DNA sequencing of the subsequent plasmid was done to verify the integrity of the construct (Fig. 9).

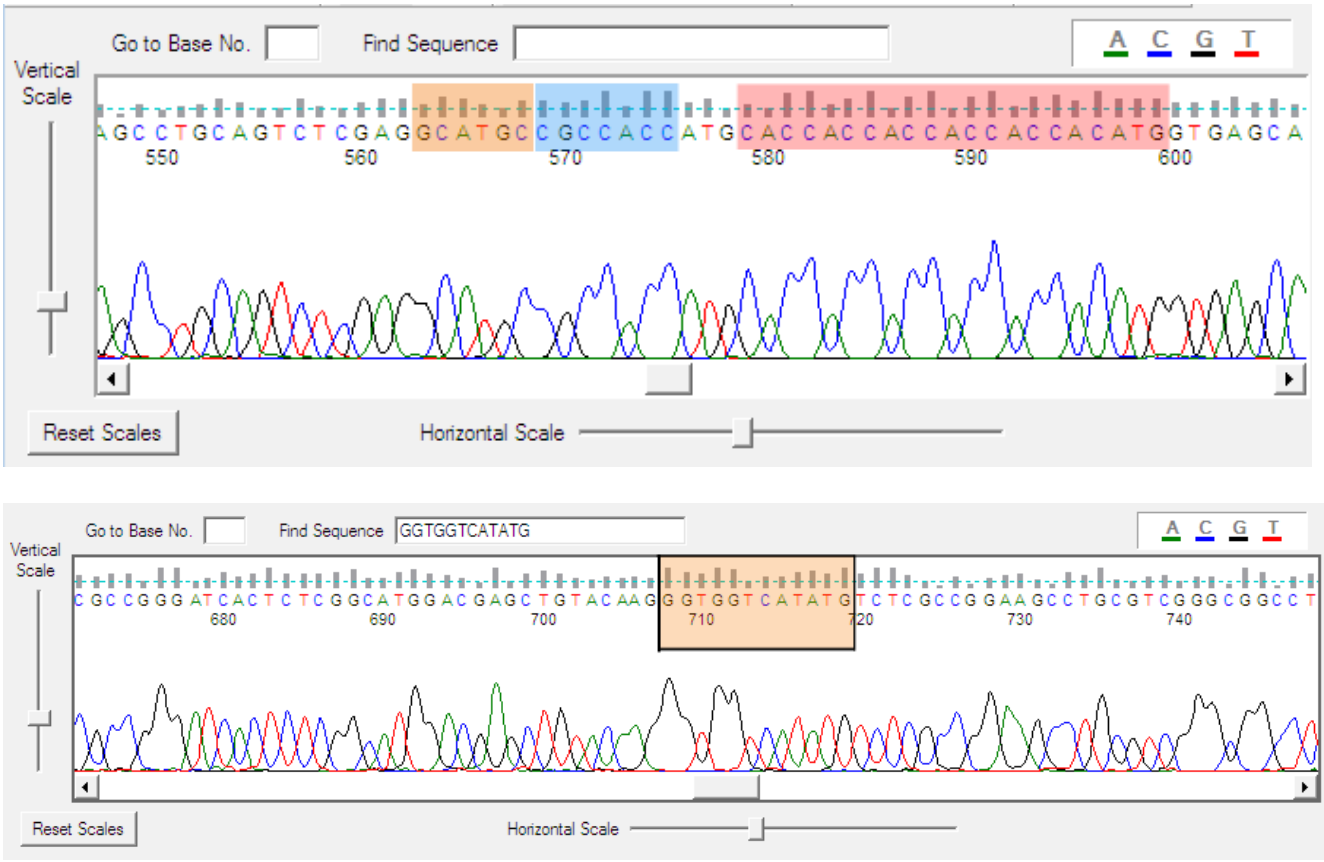


Figure 9. **Sequence of the CFP-MSH3 Hybrid.** The top chromatogram shows the N-terminus of CFP including the Sph I restriction site used for cloning (orange box), the Kozak translation initiation sequence (blue box), and the 6 x Histidine tag (red box). The bottom chromatogram shows the C-terminus of CFP connected to the N-terminus of MSH3 which includes a two residue glycine linker and the Nde I site used in cloning (orange box).

Overexpression and Purification of MSH2-MSH3

The original MSH2 and MSH3 clones in pFastBac1 were a gift from Nidhi Punja. MSH2 and MSH3 were co-overexpressed in Sf9 insect cells from recombinant bacmid and frozen at -80 ºC as previously described (chapter 3). All purification steps were carried out at 4 ºC. Cells in freeze buffer (300 mM NaCl, 25 mM HEPES-NaOH (pH 8.1), 20 mM imidazole, 10% glycerol, and protease cocktail) were thawed on ice and lysed by repeated passage through a 25 gauge needle. After centrifugation at 40,000 x *g* for 1 hr, the supernatant was loaded on to a nickel nitrilotriacetic acid Superflow column (Qiagen) and equilibrated with 10 % buffer B-1 (25 mM HEPES-NaOH (pH8.1), 300 mM NaCl, 10% glycerol, 0.24 mM PMSF, 0.384 µg /mL pepstatin, 0.384 µg /mL leupeptin, 200 mM imidazole). The protein was eluted with a linear gradient of imidazole from 20 mM to 200 mM. The peak fractions containing MSH2 and MSH3 were eluted at approximately 75 mM imidazole and were then loaded onto a PBE 94 (Sigma) column in tandem with a heparin-Sepharose column (GE Healthcare). The column was equilibrated with 30% buffer B-2 (25 mM HEPES-NaOH (pH8.1), 1 mM DTT, 0.1 mM EDTA, 10% glycerol, 0.24 mM PMSF, 0.384 µg /mL pepstatin, 0.384 µg /mL leupeptin, 1 M NaCl) and the protein was eluted with a linear gradient of salt from 300 mM to 1 M. MSH2-MSH3 eluted at approximately 450 mM NaCl. Peak fractions were diluted to 100 mM NaCl with buffer A (25 mM HEPES-NaOH (pH8.1), 1 mM DTT, 0.1 mM EDTA, 10% glycerol, 0.24 mM PMSF, 0.384 µg /mL pepstatin, 0.384 µg /mL leupeptin) and loaded onto an S-Sepharose Fastflow column (GE Healthcare) equilibrated with 10% buffer B-2. MSH2-MSH3 was eluted with a linear gradient of salt from 100 mM to 1 M, with peak fractions eluting at approximately 300 mM. Peak

fractions were then diluted to 100 mM NaCl with buffer A and loaded onto a Mono-S column (Pharmacia Biotech) equilibrated with 10% buffer B-2 and eluted with a linear gradient of salt from 100 mM to 1 M NaCl. The protein complex eluted at approximately 425 mM NaCl. Peak fractions were collected and diluted to 100 mM NaCl in buffer A and directly loaded onto a Mono-Q column (Pharmacia Biotech). The column was equilibrated with 10% buffer B-2 and eluted with a linear gradient of salt from 100 mM to 500 mM. Peak fractions eluted at approximately 200 mM NaCl and were dialyzed against 25 mM HEPES-NaOH (pH 8.1), 150 mM NaCl, 1 mM DTT, 0.1 mM EDTA and 20% glycerol (v/v). Aliquots were frozen in liquid nitrogen and stored at -80 ºC. The purity of MSH2-MSH3 was analyzed on an 8% SDS-PAGE gel and was found to be more than 95% pure by Coomassie staining (Fig.4). Protein concentration was determined by Bradford assay.

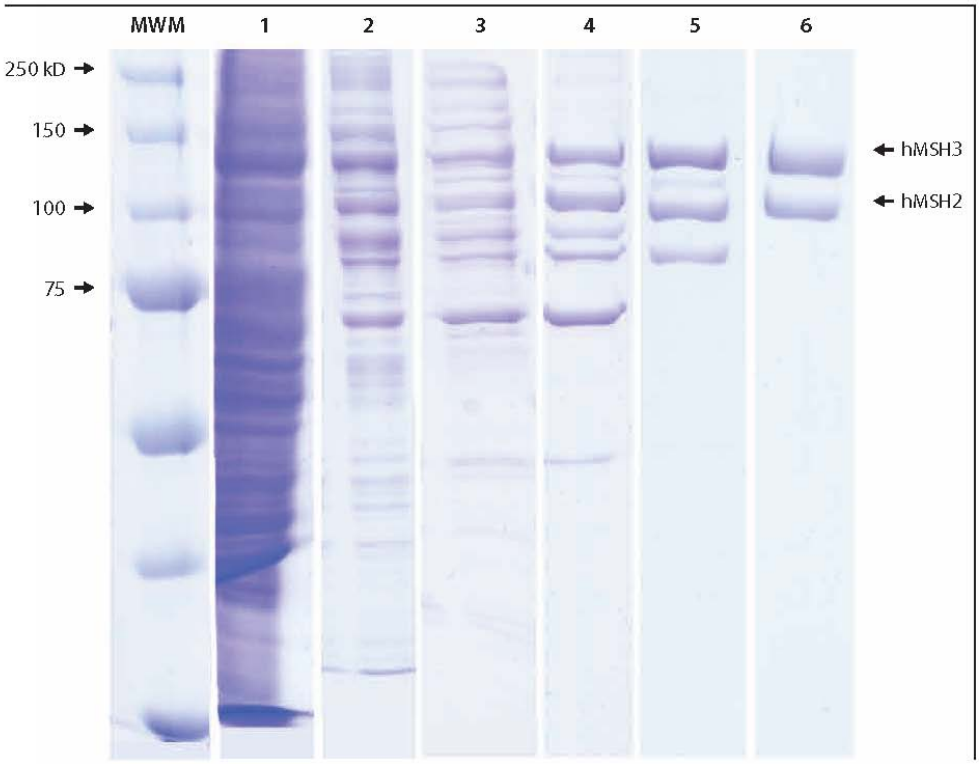


Figure 4. **Purification of wild-type hMSH2-hMSH3**. An 8% Coomassie stained gel following a five step purification procedure. Molecular weight markers (Biorad) are shown. (Lane 1) Crude extract from insect cells infected with expression virus containing MSH2 and MSH3; (lane 2) peak fractions from a nickel affinity column eluted with imidazole; (lane 3) peak fractions from a PBE 94 column in tandem with a heparin-Sepharose column and eluted with a NaCl gradient; (lane 4) peak fractions from an S-sepharose column; (lane 5) peak fractions from a mono-S column; (lane 6) pure MSH2-MSH3 after elution with NaCl from a mono-Q column. Arrows indicate hMSH2 (104.7 kD) and hMSH3 (126.75 kDa).

Steady-State ATPase activity of MSH2-MSH3

ATP binding and hydrolysis and mismatch recognition are the key conserved functions carried out by all MutS homologs. Upon lesion recognition, MSH2-MSH3 with ADP bound in the MSH2 subunit binds IDL DNA and undergoes ADP→ATP exchange[13,35]. In this experiment with a 41 bp IDL oligomer, the protein slides off the DNA, hydrolyzes ATP and repeats another round of DNA binding forming a cycle (fig.5).

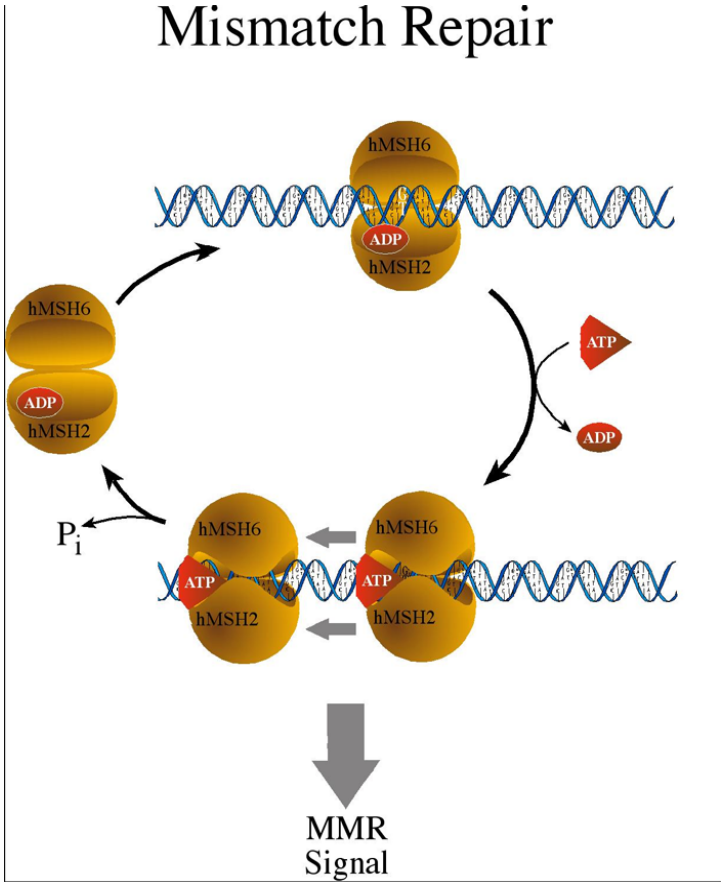


Figure 5. **MSH2-MSH3 ATP hydrolysis cycle, based on MSH2-MSH6.** ADP bound MSH2-MSH3 binds IDL DNA which stimulates ADP→ ATP exchange and translocation along the DNA backbone. ATP hydrolysis revives the competent lesion recognition complex (figure reproduced from R. Fishel, 1998. Mismatch repair, molecular switches, and signal transduction. *Genes Dev*. **12**: 2096-2101)

Figure 6 demonstrates how the inherent ATP hydrolysis (ATPase) activity of MSH2-MSH3 is stimulated by DNA containing an eight base pair IDL, and that homoduplex DNA fails to effectively trigger ATPase activity.

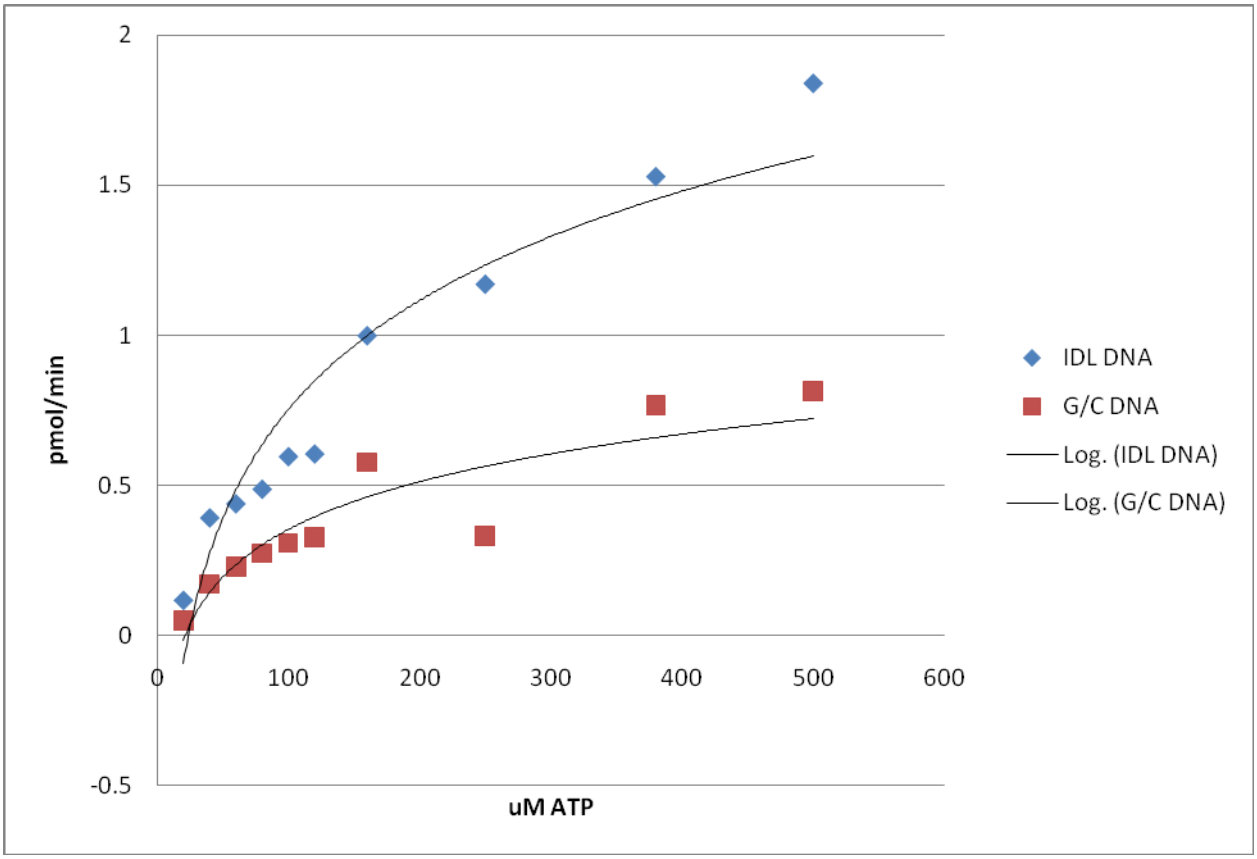


Figure 6. **ATPase activity of hMSH2-hMSH3.** IDL stimulates the inherent ATPase activity of the MSH2-MSH3 heterodimer. ATPase activity in the prescence of homoduplex DNA is attenuated.

DNA Substrates

The sequences of the DNA substrates are as follows: G/C DNA, 5'-CCG CTG AAT TGC ACC GAG CTC GAT CCT CGA TGA TCC TAA GC-3'; IDL DNA, 5'-CCG CTG AAT TGC ACC GAG CTC CA CA CA CA GAT CCT CGA TGA TCC TAA GC-3'. Complimentary strand for the G/C and IDL oligomer, 5' GCT TAG GAT CAT CGA

GGA TCG AGC TCG GTG CAA TTC AGC GG-3'. Oligomers were ordered from Integrated DNA Technologies and annealed over night at 55 ºC in annealing buffer (100 mM NaCl, 10 mM Tris (pH 8.0), and 2 mM EDTA), and purified by HPLC (by Thomas Haver). Steady-state ATPase assays were performed in reaction buffer (130 mM NaCl, 25 mM HEPES-NaOH (pH 7.8), 1 mM DTT, 200 ug/mL acetylated BSA (Promega), 10 mM $MgCl_2$, 15 nM [g-32P]-ATP) and varying amounts of unlabelled ATP as described. Reactions were performed in the presence of either IDL DNA or G/C DNA at 200 nM. Reactions were initiated by addition of 10, 30, and 50 nM MSH2-MSH3 depending on ATP concentration. Reactions were incubated for 30 min at 37 ºC and stopped by the addition of 400 uL of charcoal solution (10% activated charcoal (Sigma), 10 mM EDTA). Samples were centrifuged at 14,000 RPM for 15 minutes to pellet the charcoal and 100 uL supernatant was removed and counted by liquid scintillation to measure the released phosphate. At different ATP concentrations, the velocity of the reaction was measured as pmol ATP/min by normalizing for the amount of protein in the sample and plotted vs. ATP concentration.

Real time binding-dissociation of hMSH2-hMSH3 using surface plasmon resonance

Surface plasmon resonance (SPR) is a phenomenon used to study biomolecular interactions that occurs at an interface between media with different refractive indices. This technology was exploited using a Biacore 3000 to follow binding interactions of the hMSH2-hMSH3 heterodimer in real time. In these experiments, the sensor chip utilized consists of a carboxymethylated dextran matrix immobilized with streptavidin. The streptavidin is used specifically to bind to molecules that are tagged with biotin, due to

the extremely high affinity these two molecules have for each other ($K_d \approx 10^{-15}$ M)[43]. The DNA oligomers used in this study were identical to those used in the ATPase experiment with the exception that they were biotinylated at the 3' end for binding to the sensor chip.

The integral component essential for an SPR experiment is the optical system which is focused on the sensor chip containing a layer of gold on glass, which provides the necessary components for detection[44]. The other side of the gold contains streptavidin covalently linked to the sensor chips which permits binding of the biotinylated DNA and houses the flow channel through which protein and buffer flow (fig. 7). Binding of protein to the immobilized DNA causes a change in the refractive index, and thus a change in the SPR angle. The SPR angle change is reported as resonance units (RU), and a 1000 RU response equals a surface concentration change of approximately 1 ng/mm^2 (45, 46).

The basic experimental reaction is visualized in Figure 7. A baseline is first established with buffer only as a reference. With the DNA immobilized on the sensor chip, the protein solution is injected and binding is monitored in real time by a change in SPR angle as the solution flows over the surface. Buffer can then be passed through the system and dissociation of the protein from the DNA can be measured. The flow cell is regenerated by passing through a solution of 1 M NaCl which removes any remaining bound protein from the DNA. A baseline is then reestablished and the experiment can be repeated with alteration of experimental variables (protein concentration as an example).

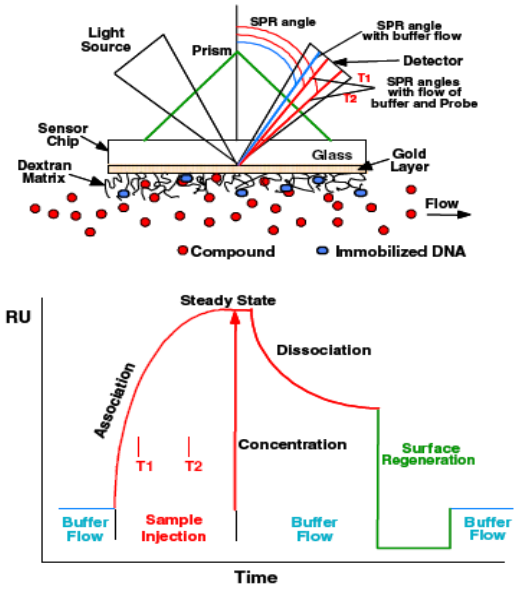


Figure 7. **Schematic and Experimental Outline of SPR.** The top panel illustrates the components of the SPR experiment including the sensor chip containing a dextran matrix of streptavidin molecules bound by biotinylated DNA (blue ovals) and the compound, in this case MSH2-MSH3. The bottom panel shows the steps in a standard SPR experiment; the change in SPR angle is converted to an RU versus time response (figure reproduced from **Nguyen, B., Tanious, F. A., Wilson, W. D.** 2006.)

In this experiment, a sensor chip SA (Biacore) was used to immobilize both IDL DNA and homo duplex DNA at a concentration of 1 ng/uL. After attaching DNA to the chip, binding buffer (0.005% Surfactant, 25 mM HEPES (pH 7.8), 130 mM NaCl, 10% glycerol, 1 mM DTT, 10 mM $MgCl_2$, 150 ug/mL BSA) was passed through to establish a baseline concentration with no protein. After stabilization of the baseline, protein was passed through the system and binding was observed until saturation was reached at which point binding buffer was again added to remove any unbound protein. Next an ATP buffer (0.005% Surfactant, 25 mM HEPES (pH 7.8), 130 mM NaCl, 10% glycerol, 1 mM DTT, 10 mM $MgCl_2$, 150 ug/mL BSA, 1 mM ATP) was passed through the system, followed by a 1 M salt buffer (0.005% Surfactant, 25 mM HEPES (pH 7.8), 130

mM NaCl, 10% glycerol, 1 mM DTT, 10 mM $MgCl_2$, 150 ug/mL BSA, 1 M NaCl) to remove any bound protein from the DNA and regenerate the chip. Binding buffer was again passed through to reestablish the base line and repeat the experiment at higher concentrations of protein.

ATP induces rapid dissociation of MSH2-MSH3

Previous studies have shown that the addition of ATP to IDL-bound MSH2-MSH3 results is dissociation[13]. This is in contrast to studies with MSH2-MSH3 in yeast which show that the heterodimer remains stably bound to small IDL DNA in the presence of ATP[47]. In the experiment illustrated in Figure 8, increasing titrations of MSH2-MSH3 through the system show increased binding to IDL DNA to the point of saturation, at which point buffer is passed through to remove unbound protein. After stabilization, an injection of a 1mM ATP buffer resulted in immediate dissociation of almost all protein from the DNA. Identical experiments were done simultaneously with homoduplex DNA. MSH2-MSH3 did not readily bind non-mismatch DNA at low concentrations. Increasing the concentration of protein in the solution forced binding that was either non-specific or was induced by the free 5' end of the DNA which has been shown to induce binding by MSH2-MSH6[36].

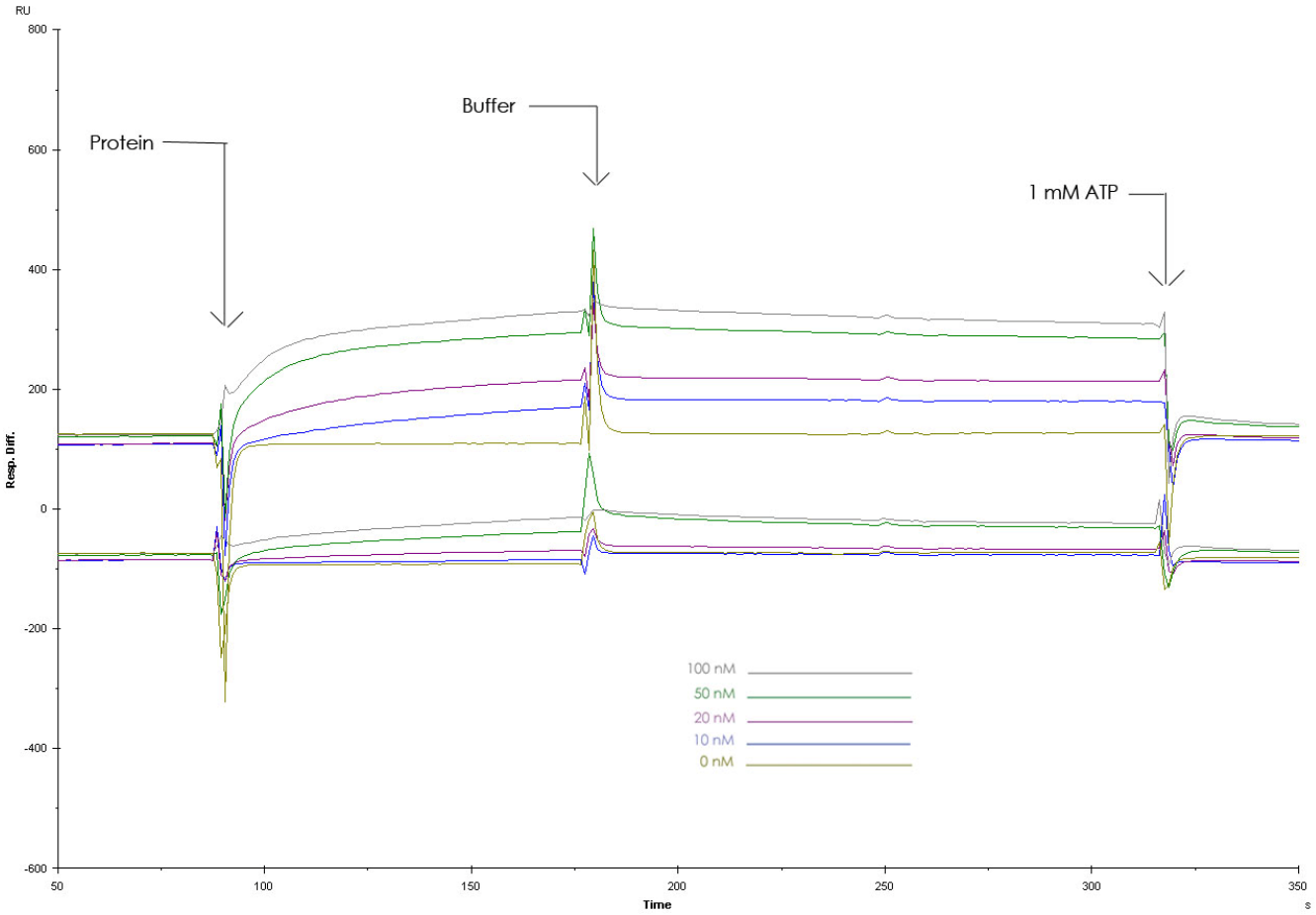


Figure 8. **A representative sensogram revealing real-time binding of MSH2-MSH3 to IDL and homoduplex DNA, and rapid dissociation by the addition of ATP.** Buffer is injected into the system to establish a baseline response after which varying concentrations of protein are added (first injection peak). After binding is complete more buffer is passed through the system to clear any unbound protein after which a buffer containing 1 mM ATP is added which induces immediate dissociation (*Top section*: IDL DNA *Bottom section*: homoduplex DNA)

Discussion

There are currently two representative models for mismatch repair. One model suggests that MutS and its homologs bind mismatch DNA followed by association with MutL or its homologs[1]. This protein DNA complex is then hypothesized to translocate on the DNA bi-directionally in an ATP-dependent manner to the site of the incision prior to the excision/repair reaction[48]. The results presented in part here and in previous publications for the MutS homologs MSH2-MSH3 and MSH2-MSH6 provide an alternative model, mainly that the MSH proteins operate as an adenosine nucleotide-regulated molecular switch[13,31,36]. The molecular switch model suggests that ADP → ATP exchange provoked by mismatch DNA results in an ATP-bound sliding clamp that diffuses along DNA to signal downstream mismatch repair factors, as opposed to relying on ATP hydrolysis to drive motion of the protein along the DNA. ATP then induces release of the protein from DNA which can recycle the mismatch repair proteins through another round of hydrolysis[36,49,50]. The recycling of the mismatch repair complex in the ATPase assay as well the real-time binding-dissociation study on MSH2-MSH3 in the presence of ATP have reinforced this model and stabilized the notion that not only MSH2-MSH6 acts in this manner but also MSH2-MSH3.

The next step in understanding how MSH2-MSH3 functions will be to understand how the structural transformations upon lesion binding alter inherent ATP binding and signaling properties. The use of the fluorescently labled MSH2-MSH3 proteins will allow us to visualize these changes in real time. Protein footprinting by partial proteolysis has been used in the past to show that the MSH2-MSH3 complex undergoes

adenosine nucleotide-regulated conformational changes, but these experiments leave us in the dark with regards to the structure and nature of these changes, and how specifically they affect other functions of the protein[13]. These experiments will be of great significance in trying to determine the role of MSH2-MSH3 in trinucleotide repeat expansion, as well as the mismatch repair pathway in general.

It is well known that bacterial MutS demonstrates ATPase activity that has been conserved in the yeast and human homologs[51,52]. Most of the literature has focused on the MSH2-MSH6 heterodimer, and, as a result, much remains to be studied regarding the ATPase activity of MSH2-MSH3 and how it processes adenosine nucleotides upon lesion recognition. Although recent literature has provided some insight, little work has been done to understand how each subunit interacts in processing ATP and ADP. Here, the mutants that were created will allow examination of these issues. Mutations in the Walker A box of both MSH2 and MSH3 will allow us to see how the protein uses adenylate to bind the DNA, and will also permit us to study how it processes dissociation from the DNA backbone after lesion binding. The mutants will allow us to see which subunit of the dimer specifically is responsible in processing adenosine nucleotide during these processes.

This and future work will be important to the field of mismatch repair; to understand the mechanisms behind initiation, lesion recognition, and also downstream signaling of mismatch repair machinery. How these events and defects in mismatch repair lead HNPCC and trinucleotide repeat expansion is still an area of active research and study. Understanding and working out the underlying fundamental mechanisms and

models of these processes will open up new areas of research and expand the tools we have at our disposal to combat disease.

Acknowledgments – I thank R. Fishel, N. Punja, J. Perry, S. Javaid, and T. Haver for their mentorship and advice, L. Kerepesie and M. McIlhatton for editing and for helpful discussions, and A. Cook for assistance with figures.

References:

1. **Modrich, P., and R. Lahue.** 1996. Mismatch repair in replication fidelity , genetic recombination, and cancer biology. *Annu. Rev. Biochem*. **65**:101-133.
2. **Parker, B. O., and Marinus, M. G.** 1992. *Proc. Natl. Acad. Sci. U. S. A*. **89**: 1730-1734
3. **Su, S.-S., and P. Modrich.** 1986. *Escherichia coli* mutS-encoded protein binds to mismatched DNA base pairs. *Proc. Natl. Acad. Sci. U. S. A.* **83**: 5057-5061
4. **Acharya, S., Foster, P. L., Brooks, P., and Fishel, R.** 2003. The Coordinated functions of the *E.Coli* MutS and MutL proteins in mismatch repair. *Mol Cell* **12**: 233-246
5. **Surtees, J. A., Argueso, J. L., & Alani, E.** 2004. Mismatch repair proteins: key regulators of genetic recombination. *Cytogen. Genome Res.* **107**: 146-159
6. **Yamaguchi, M., Dao, V., and Modrich, P.** 1998. MutS and MutL activate DNA helicase II in a mismatch-dependent manner. *J. Biol. Chem.* **273**: 9197-9201.
7. **Viswanathan, M., and Lovett, S. T.**, 1998. Single-strand DNA-specific exonucleases in *Escherichia coli* – roles in repair and mutation avoidance. *Genetics* **149**: 7-16

8. **Fishel, R., M. K. Lescoe, M. R. Rao, N. G. Copeland, N.A. Jenkins, J. Garber, M. Kane, and R. Kolodner.** 1993. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* **75**: 1027-1038

9. **Dragileva, E., Hendricks, A., Teed, A., Gillis, T., Lopez, E., Friedberg, E. C., Kucherlapati, R., Edelmann, W., Lunetta, K. L., MacDonald, M. E., and Wheeler, V. C.** 2009. Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. *Neurobiology of Disease* **3**3: 37-47

10. **Acharya, S., Wilson, T., Gradia, S., Kane, M. F., Guerrette, S., Marsischky, G. T., Kolodner, R., Fishel, R.**, 1996. hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. *Proc. Natl. Acad. Sci. U.S.A.* **93**: 13629-13634

11. **Marsischky, G.T., Filosi, N., Kane, M. F., and Kolodner, R.** 1996. Redundancy of *Saccharomyces Cerevisiae* MSH3 and MSH6 in MSH2-dependent mismatch repair. *Genes & Dev*. **10**: 407-420

12. **Harrington, J. M., and Kolodner, R**. 2007. *Saccharomyces cerevisiae* MSH2-MSH3 acts in repair of base-base mispairs. *Molecular and Cellular Biology*. **27**: 6546-6554

13. **Wilson, T., Guerrette, S., and Fishel, R**. 1999. Dissociation of mismatch recognition and ATPase activity by hMSH2-hMSH3. *J. Biol. Chem*. **274**: 21659-21664

14. **Muller, A. & Fishel, R**. 2002. Mismatch repair and the hereditary non-polyposis colorectal cancer syndrome (HNPCC). *Cancer Invest.* **20**: 102-109

15. **Leach, F. S. *et al.*** 1993. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* **75**:1215-1225

16. **Nicolaides, N. C. *et al.*** 1994. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature* **371**: 75-80

17. **Papadopoulos, N. *et al.*** 1994. Mutation of a mutL homolog in hereditary colon cancer. *Science* **263:** 1625-1629

18. **Fishel, R. *et al.*** *1993.* The human mutator gene homolog *MSH2* and its association with hereditary nonpolyposis colon cancer. *Cell* **75**: 1027-1038

19. **Kolodner, R. D. *et al.*** 1999. Germ-line msh6 mutations in colorectal cancer families. *Cancer Res.* **59**: 5068-5074

20. **Modrich, P. & LaHue, R. S.** 1996. Mismatch repair in replication fidelity, genetic recombination and cancer biology. *Annu. Rev. Biochem.* **65**: 101-113

21. **Vasen, H.F., Wijnen, I. T., Meiiko, F. H., Kleibeuker, J.H., Taal, B.G., Grifiioen, G., Nagengast, F.M., Meijers Heijboer, E.H., Bertario, L., Varesco, L., Bisgaard, M. L., Mohr, J., Fodde, R., Khan, P. M.** 1996. Cancer risk in families with hereditary nonpolyposis colorectal cancer diagnosed by mutation analysis. *Gastroenterology* **110:** 1020-1027.

22. **Thibodeau, S. N., Bren, G., Schaid, D.** 1993. Microsatellite instability in caner of the proximal colon. *Science* **260**: 816-819

23. **Strand, M., Prolla, T., Liskay, R. M. & Petes, T. D.** 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**: 274-276

24. **Van den Broek, W. J. et al.** 2002. Somatic expansion behavior of the $(CTG)_n$

repeat in myotonic dystrophy knock-in mice is differentially affected by MSH3 and MSH6 mismatch-repair proteins. *Hum. Mol. Genet.* **11**: 191-199

25. **Cummings, C. J. & Zoghbi, H. Y.** 2000. Fourteen and counting; unraveling trinucleotide repeat diseases. *Hum. Mol. Genet.* **9**: 909-916

26. **Lahue, R.S. & Slater, D. L.** 2003. DNA repair and trinucleotide repeat instability. *Front. Biosci.* **8**: s653-s666

27. **McMurray, C.T.** 2008. Hijacking of the mismatch repair system to cause CAG expansion and cell death in neurodegenerative disease. *DNA Repair* **7**: 1121-1134

28. **Gacy, A. M., Goellner, G., Juranic, N., Macura, S. & McMurray, C. T.** 1995. Trinucleotide repeats that expand in human disease from hairpin structures in vitro. *Cell* **81**: 533-540

29. **Owen, B. A. L., Yang, Z., Lai, M., Gajek, M., Badger II, J., Hayes, J., Edelmann, W., Kucherlapati, R., Wilson, T., McMurray, C. T.** 2005. $(CAG)_n$-hairpin DNA binds to MSH2-MSH3 and changes properties of mismatch recognition. *Nat. Struct. & Mol. Bio.* **12**: 663-670

30. **Moore, H., Greewell, P.W., Liu, C.P., Arnheim, N., Petes, T.D.** 1999. Triplet repeats form secondary structures that escape DNA repair in yeast. *Proc. Natl. Acad. Sci. U.S.A.* **96**: 1504-1509

31. **Fishel, R.** 1998. Mismatch repair, molecular switches, and signal transduction. *Genes Dev*. **12**: 2096-2101

32. **Walker, J.E., M. Saraste, M.J. Runswick, and N.G. Gay.** 1982. Distantly related sequences in the alpha-and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1:** 945–

951

33. **Warren, J. J. *et al.*** 2007. Structure of the human MutS DNA lesion recognition complex. *Mol. Cell* **26**: 579-592

34. **Lamers, M.H.** et al. 2000. The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. *Nature* **407**: 711-717

35. **Owen, B.A.L., Lang, W. H., & McMurray, C.T.** 2009. The nucleotide binding dynamics of human MSH2-MSH3 are lesion dependent. *Nat. Struct. & Mol. Bio*. **16**: 550-557

36. **Gradia, S., Subramanian, D., Wilson, T., Acharya, S., Makhov, A., Griffith, J., Fishel, R.** 1999. hMSH2-hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA. *Mol. Cell* **3:** 255-261

37. **C. W. Cody, D. C. Prasher, W. M. Westler, F. G. Prendergast, and W. W. Ward**, 1993. Chemical structure of the hexapeptide chromophore of the Aequorea green-fluorescent protein. *Biochemistry* **32**: 1212–1218

38. **R. Heim, D. C. Prasher, and R. Y. Tsien,** 1994. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.* **91:** 12501–12504

39. **F öster, Th**., 1948. Intermolecular energy migration and fluorescence, Ann. Phys. **2:** 55-75

40. **Stryer, L.**, 1978. Fluorescence energy transfer as a spectroscopic ruler. *Annu. Rev. Biochem*. **47**: 819-846

41. **Lackowicz, Joseph R**. Principles of Fluorescence Spectroscopy. New York: Kluwer Academic/Plenum Publishers, 1999.

42. **Kozak, M.** 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15:** 8125-8148

43. **Chilkoti, A., Stayton, P.S.**, 1995. Molecular orgins of the slow streptavidin-biotin dissociation kinetics. *J. Am. Chem. Soc*. **117**: 10622-10628

44. **Nguyen, B., Tanious, F. A., Wilson, W. D.** 2006. Biosensor-surface plasmon resonance: Quantitative analysis of small molecule-nucleic acid interactions. *Methods* **42**: 150-161

45. **Davis, T. M., Wilson, W. D.**, 2000. *Anal. Biochem.* **284**: 348-353

46. **Davis, T. M., Wilson, W. D**., 2001. Methods Enzymol. **340**: 22-51

47**. Habraken, Y., Sung, P., Prakash, L., and Prakash, S**. 1996. Binding of insertion/deletion DNA mismatches by the heterodimer of yeast mismatch repair proteins MSH2 and MSH3. *Curr. Biol.* **6**: 1186-1187

48. **Allen, D. J., Makhov, A., Grilley, M., Taylor, J., Thresher, R., Modrich, P., and Griffith, J. D**. 1997. MutS mediates heteroduplex loop formation by a translocation mechanism. *EMBO J.* **16**: 4467-4476

49. **Gradia, S., Acharya, S., Fishel, R.** 2000. The Role of Mismatched Nucleotides in Activating the hMSH2-hMSH6 Molecular Switch. *J. Biol. Chem*. **275**: 3922-3930.

50. **Fishel, R., Acharya, S., Berardini, M., Bocker, T., Charbonmneau, N., Cranston, A., Gradia, S., Guerrette, S., Heinen, C. D., Mazurek, A., Snowden, T., Schmutte, C., Shim, K.-S., Tombline, G., Wilson, T.** 2000. Signaling Mismatch Repair: The mechanics of an adenosine-nucleotide molecular switch. *Cold Spring Harbor Symposia on Quantitative Biology*. Vol. **LXV**. 217-224

51. **Gradia, S., Acharya, S., and Fishel, R**., 1997. The human mismatch recognition

compex hMSH2-hMSH6 functions as a novel molecular switch**.** *Cell* **91:** 995-1005

52. **Alani, E., Sokolsky, T., Studamire, B., Miret, J. J., and Lahue, R. S.** 1997. Genetic and biochemical analysis of Msh2p-Msh6p: role of ATP hydrolysis and Msh2p-Msh6p subunit interactions in mismatch base pair recognition. *Mol. Cell. Biol*. **17**: 2436-2447

53. **Kolodner, R., Mendillo, M. L., Putnam, C. D.,** 2007. Coupling distant sites in DNA during DNA mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.* **104:** 12953-12954