ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

INVESTIGATION OF STRUCTURAL DIFFERENCES BETWEEN WILD-TYPE AND MUTANT FORMS OF MutSα BY MOLECULAR DYNAMICS SIMULATIONS

M.Sc. THESIS

Clara Xazal BURAN

Department of Molecular Biology-Genetics and Biotechnology

Molecular Biology-Genetics and Biotechnology Programme

AUGUST 2022

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

INVESTIGATION OF STRUCTURAL DIFFERENCES BETWEEN WILD-TYPE AND MUTANT FORMS OF MutSα BY MOLECULAR DYNAMICS SIMULATIONS

M.Sc. THESIS

Clara Xazal BURAN
(521201107)

Department of Molecular Biology-Genetics and Biotechnology

Molecular Biology-Genetics and Biotechnology Programme

Thesis Advisor: Assoc. Prof. Mert GUR

AUGUST 2022

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

MutSα HETERODİMERİNİN YABANIL TİP VE MUTANT FORMLARI ARASINDAKİ YAPISAL FARKLILIKLARININ MOLEKÜLER DİNAMİK SİMÜLASYONLARI İLE İNCELENMESİ

YÜKSEK LİSANS TEZİ

Clara Xazal BURAN
(521201107)

Moleküler Biyoloji-Genetik ve Biyoteknoloji Ana Bilim Dalı

Moleküler Biyoloji-Genetik ve Biyoteknoloji Ana Bilim Dalı Programı

Tez Danışmanı: Doç. Dr. Mert GÜR

AĞUSTOS 2022

Clara Xazal Buran, a M.Sc. student of İTU Graduate School student ID 521201107 successfully defended the thesis/dissertation entitled "INVESTIGATION OF STRUCTURAL DIFFERENCES BETWEEN WILD-TYPE AND MUTANT FORMS OF MutSα BY MOLECULAR DYNAMICS SIMULATIONS", which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Assoc. Prof. Mert GÜR** İstanbul Technical University	.....
Jury Members :	**Prof. Dr. Gizem DİNLER DOĞANAY** İstanbul Technical University	.....
	Prof. Dr. Batu ERMAN Boğaziçi University	.....

Date of Submission : 18 July 2022
Date of Defense : 03 August 2022

To my family,

FOREWORD

Since I have decided to get a master's degree, many people supported me on this journey. First of all, I would like to express my gratitude to my advisor, Assoc. Prof. Mert Gür, who encourages each of his students to perform top level science and widens his students' horizons with his in-depth and extensive knowledge. Secondly, I would like to express my deep thanks to "GURLAB" colleagues Mert Gölcük, Elhan Taka, Ayten Dilara Gürsel, Sema Zeynep Yılmaz, and Derman Baştürk for their friendship, help, and all the time they contribute their joy of living.

I am also grateful to my dear friends, Büşra Nakipoğlu, Nuray Çınkılıç, Aydan Ercan, Berfin Gülaçan, Ömer Özdilek, Furkan Şen, Selen Akpınar and also Emre Çağatay Bilgin for support me during my hard times.

Last but not least, I would like to honor my lovely father, mother, and sister for their love, and they believed in me.

I would also like to acknowledge the financial support of TUBITAK for supporting my thesis with the project no. 318S129.

August 2022 Clara Xazal BURAN

TABLE OF CONTENTS

Page

ABBREVIATIONS

ATPase : Adenosine Triphosphatase
C_α : Alpha carbon
kDa : KiloDalton
MD : Molecular Dynamics
MMR : Mismatch Repair
MSH2 : MutS Homolog 2
MSH3 : MutS Homolog 3
MSH6 : MutS Homolog 6
MutS : Mutator protein
NAMD : Nanoscale Molecular Dynamics
VMD : Visual Molecular Dynamics
PCA : Principal Component Analysis
PME : Particle-Mesh Ewald
RMSD : Root-Mean-Square Deviation
RMSF : Root-Mean-Square Fluctuation

SYMBOLS

$\mathbf{C_p}$	: Heat Capacity
H	: Enthalpy
$\boldsymbol{k_B}$	: Boltzmann coefficient
KCl	: Potassium chloride
N	: Total number of atoms
NaCl	: Sodium chloride
P	: Pressure
$\boldsymbol{R}$	: Universal gas constant
R	: Coordinate
T	: Temperature
V	: Volume

LIST OF FIGURES

Page

INVESTIGATION OF STRUCTURAL DIFFERENCES BETWEEN WILD-TYPE AND MUTANT FORMS OF MutSα BY MOLECULAR DYNAMICS SIMULATIONS.

SUMMARY

Cancer is the most common cause of death after cardiovascular diseases. Among the cancer types, colon cancer, which is very common in our country and the world, endangers human health. The aim of the studies in this thesis is to determine the changes in the structure and dynamics of the MutSα protein due to mutations in the protein complex associated with colon cancer. Molecular Dynamics (MD) Simulations provide important information about protein dynamics by providing the opportunity to observe changes in the structures and functions of proteins. Computational methods have been a remarkable field in recent years because they provide energetic, structural and dynamic information at the atomic level that cannot be obtained using experimental techniques. In this study, the effects of mutations in MutSα on protein structure and dynamics were analyzed using computational and theoretical methods.

MutSα complex, formed by the MSH2-MSH6 heterodimer is responsible for the recognition and repair of mismatches. Some mutations in MutSα have been associated with cancer studies. Certain mutations that occur on protein complex cause the accumulation of mutations in the cell and eventually causing cancer. Only some of the mutations in protein complex have been associated with cancer, but the effect of some mutations, defined as pathogenic variation, on protein activity is not yet known. The MSH2 and MSH6 proteins that make up the MutSα complex are divided into five major regions: the mismatch binding domain, the connector domain, the lever domain, the clamp region, and ATPase domain. The mismatch binding domain recognizes mismatched bases and binds to these bases. The connector domain is responsible for allosteric signaling, while the lever region is thought to be responsible for signaling between the mismatch binding domain and the ATPase domain. The clamp domain stabilizes the DNA by interacting and bonding with the DNA backbone. The ATPase domain is the region where the energy required for the protein to move along the DNA is obtained by hydrolysis of ATP. MutSα complex detects mismatches by scanning the DNA and binds to mismatch region to start repair mechanisms.

In a healthy cell, DNA repair mechanisms work smoothly, DNA damages are repaired, and apoptosis occurs in case of irreparable damage. However, mutations in the proteins in the DNA repair mechanism prevent the correct functioning of the DNA repair mechanism, causing the accumulation of damaged/mutant DNA in the cell and causing tumor cells. Studies have revealed that specific mutations in MutSα, one of the DNA mismatch repair proteins, can be hereditary and cause cancer. Some mutations obtained as a result of genetic screening tests by Prof. Dr. Levent Doğanay and his research group, with samples taken from individuals with a family history of

colon cancer have been reported to our study group. Ala733Thr, Arg577Cys and Ser127Asn mutations were detected, and the effects of these mutations on MutSα without DNA were investigated by performing all-atom MD simulations. In order to look at the changes that these mutations may cause in protein function, first of all, systems were prepared for the MD simulations of the MutSα protein using structures with PDB ID: 2O8B. Two replica systems were created for each construct to compare the simulation results for wild-type and mutant proteins. While simulating the proteins, the required ion concentration was provided by considering the cell's physiological conditions. After the structures were dissolved in water, short minimization and equilibrium simulations were performed. The equilibrated structures were simulated for 200ns. Two replicate simulations were carried out for wild-type and three mutant proteins, and totaling 1600ns MD simulations were performed.

Root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), heat capacity, interaction analysis and principal components analysis (PCA) were performed to investigate the effect of the mutations on the MutSα complex. The RMSD analysis gives information about whether the simulation has reached equilibrium and whether there has been a conformational change. The average RMSD values of mutant proteins were higher than the wild-type protein. Therefore, these mutations resulted in changes in protein conformation and dynamics. RMSF analysis is used to provide insight about structural flexibility, mobility and thermal stability. In this thesis, fluctuations were investigated for five domains separately and it was observed that the mutant proteins had similar or higher fluctuations than the wild-type protein in general. For the wild-type, the region with the highest fluctuations was observed to be the clamp region. Heat capacity can provide information about the thermodynamic properties of the protein. The findings obtained as a result of the analysis showed higher heat capacities for all mutants compared to that of the WT protein and show that the thermodynamic properties of the protein may have been differentiated. As a result of the interaction analysis, it was observed that the mutants had changes in hydrogen bond and salt bridge interactions compared to the WT protein. With PCA analysis, the two most dominant movements in the protein were obtained. Conformations were projected onto principal components and based on the conformational space sampled by the proteins during the MD simulations heat map surfaces were constructed. In this study, the effects of mutations in the DNA-free MutSα protein structure and dynamics were investigated and it was concluded that mutant structures showed different thermodynamic and structural properties compared to wild-type.

MutSα HETERODİMERİNİN YABANIL TİP VE MUTANT FORMLARI ARASINDAKİ YAPISAL FARKLILIKLARININ MOLEKÜLER DİNAMİK SİMÜLASYONLARI İLE İNCELENMESİ

ÖZET

Kanser, günümüzde kardiyovasküler rahatsızlıklardan sonra en çok ölüme neden olan hastalık türü olmuştur. Kanser türleri içerisinde ise ülkemizde ve dünyada oldukça sık görülen kolon kanseri insan sağlığını tehlikeye atmaktadır. Herediter Non-Polipozis Kolorektal Kanser (HNPCC), kalıtsal bir kolon kanseri türüdür ve tüm klorektal kanserlerin %3-5'ini oluşturmaktadır. HNPCC ile ilişkilendirilen birkaç protein bulunmaktadır. Bu proteinler, DNA onarım mekanizmalarında görev alan ve DNA'da oluşan hataların giderilmesinde görev almaktadır. Yapılan çalışmalar, DNA'daki lezyonların onarımından sorumlu olan proteinlerde meydana gelen mutasyonların, proteinlerin doğru çalışmasını engelleyerek hücrede mutasyonların birikmesine neden olduğunu ve hücrelerin kanser hücresine dönüşmesine yol açtığını göstermiştir. Bu çalışmaların amacı, kolon kanseri ile ilişkilendirilen MutSα protein kompleksinde meydana gelen bazı mutasyonların proteinin yapısında ve dinamiğinde meydana getirdiği değişimleri tespit etmektir. Moleküler Dinamik (MD) Simülasyonları, proteinlerin yapıları ve işlevlerindeki değişimleri gözlemleyebilme olanağı sağlayarak protein dinamiği ile ilgili önemli bilgiler vermektedir. Hesaplamalı yöntemler, deneysel teknikler kullanılarak elde edilemeyen atomik seviyede enerjetik, yapısal ve dinamik bilgiler sunması nedeniyle son yıllarda oldukça dikkat çeken bir alan olmuştur. Bu çalışmada, hesaplamalı ve teorik yöntemler kullanılarak MutSα protein kompleksinde meydana gelen mutasyonların protein yapı ve dinamiğine etkileri analiz edilmiştir.

Çeşitli eksojen ve endojen etkenler genoma zarar vererek hasarlı DNA'ya neden olmaktadırlar. Genomik stabilizasyon sonucunda meydana gelebilecek DNA hasarları, kodladıkları proteinlerin hatalı olmasına ve hücrenin işleyişinin aksamasına neden olarak kanser hücrelerine ya da hücre ölümüne neden olabilmektedir. Hücreler, hasarlı DNA'yı düzeltebilmek ve hücrenin temel işleyişindeki aksamaları engellemek amacıyla çeşitli DNA tamir sistemleri geliştirmişerdir. Bu tamir sistemleri, meydana gelen DNA lezyonlarının çeşitlerine ve büyüklüklerine göre hasarlı DNA'ya müdahale ederek hasarın giderilmesini sağlamaktadırlar. Yanlış eşleşme tamiri (İngilizcesi Mismatch Repair, MMR), DNA tamir sistemlerinden biridir. Bu sistem, replikasyon sırasında polimerazların hata düzeltme (İngilizcesi proofreading) etkinliğinden kaçan yanlış baz eşleşmelerinin düzeltilmesinde ve polimeraz kayması nedeniyle tekrarlayan DNA dizilerinin replikasyonunda ortaya çıkabilen ekleme/silme döngülerinin (İngilizcesi insertion/deletion loop, IDL) onarımında görevlidir. MMR, MutS ve MutL protein kompleksleri tarafından gerçekleştirilmektedir. Uyumsuzluk tanıma adımıyla başlayan MMR sisteminde iki heterodimer protein kompleksi görev almaktadır. MutSα heterodimeri, baz-baz uyumsuzluklarının ve küçük IDL'lerin tanınmasını sağlarken MutSβ heterodimeri, daha büyük IDL'lerin tanınmasında görev alır.

MutSα, MSH2 ve MSH6 genlerinin kodladığı aynı isimli proteinlerin dimer olarak bulunduğu kompleks yapıya verilen isimdir. MutSα, DNA üzerinde hareket ederek yanlış eşleşmiş bazları tanır ve bu bölgeye bağlanarak DNA tamirini başlatır. MutSα'da meydana gelen bazı mutasyonlar kanser çalışmaları ile ilişkilendirilmiştir. Bu kompleks proteinin üzerinde meydana gelen belirli mutasyonlar, proteinin işlevinin gerçekleşmesini engelleyerek hücrede mutasyonların birikmesine ve sonucunda kansere neden olmaktadır. Bu proteinde meydana gelen mutasyonların sadece bir kısmı kanser ile ilişkilendirilmiş ancak patojenik varyasyon olarak tanımlanan bazı mutasyonların protein etkinliği üzerindeki etkisi henüz kesin bilinmemektedir. MutSα kompleksini oluşturan MSH2 ve MSH6 proteinleri beş ana bölgeye ayrılmıştır: yanlış eşleşme bağlanma bölgesi, bağlayıcı bölge, kol bölgesi, kıskaç bölgesi ve ATPaz bölgesi. Bu bölgeler her iki monomerde de bulunan ve benzer görevleri gerçekleştiren yapılardır. Yanlış eşleşme bağlanma bölgesi, DNA'da bulunan yanlış eşleşen bazları tanıyan ve bu bazlara bağlanmanın gerçekleştiği bölgedir. Bağlayıcı bölge allosterik sinyalleşmeden sorumluyken kol bölgesinin ise yanlış eşleşme bölgesi ile ATPaz bölgesi arasında sinyal iletiminden sorumlu olduğu düşünülmektedir. Kıskaç bölgesi DNA omurgası ile etkileşerek ve bağ yaparak DNA'yı stabilize eder. ATPaz bölgesi, ATP hidrolizi yaparak proteinin DNA boyunca hareket etmesi için gerekli enerjinin elde edildiği bölgedir. Bu domainleri içeren MSH2 ve MSH6 proteinlerinin bir araya gelmesiyle oluşan MutSα kompleksi DNA üzerinde taramalar yaparak yanlış eşleşmeleri tanıyıp bu bölgelere bağlanır, DNA tamirinden sorumlu olan diğer proteinlerin de bu bölgeye gelerek hatanın boyutuna bağlı olarak hasar tamiri ya da programlı hücre ölümünün (apoptozis) yolunu açar.

Sağlıklı bir hücrede DNA tamiri mekanizmaları sorunsuz çalışarak DNA hasarları tamir edilir, tamir edilemeyecek boyuttaki hasar olması durumda apoptozis gerçekleşir. Ancak DNA tamir mekanizması içerisindeki proteinlerde meydana gelen mutasyonlar DNA tamir mekanizmasının doğru çalışmasını engelleyerek hücrede hasarlı/mutant DNA'ların birikmesine ve tümör hücrelerine ve erken yaşlanmayla sonuçlanabilecek metabolik süreçlere neden olmaktadır.

Yapılan çalışmalar DNA yanlış eşleşme tamiri proteinlerinden olan MutSα'da meydana gelen spesifik mutasyonların kalıtsal olabildiğini ve kansere neden olduğunu ortaya koymuştur. Kalıtsal kolon kanserine neden olan varyasyonların patojenitesine bakılması amacıyla yapılan çoklu gen testleri sonucunda tespit edilen varyantların patojenisitesi ile ilgili olarak sonuçlara varılmaya çalışılmaktadır. Bazı varyantların patojenisitesi tam olarak tayin edilemediğinden bu varyantlar klinik önemi belirsiz varyant (İngilizcesi variant of uncertain significance, VUS) olarak isimlendirilmektedir. Prof. Dr. Levent Doğanay ve grubunun yapmış olduğu çalışmada ailesinde kolon kanseri öyküsü bulunan bireylerden alınan örnekler ile yapılan genetik tarama testleri sonucunda elde edilen Ala733Thr, Arg577Cys ve Ser1279Asn VUS'ları çalışma grubumuza bildirilmiştir. Bu mutasyonların DNA'sız MutSα'ya olan etkileri tüm atom Moleküler Dinamik (MD) simülasyonları yapılarak incelenmiştir. Bu mutasyonların protein işlevinde meydana getirebileceği değişimlere bakmak üzere öncelikle, MutSα proteininin MD simülasyonları için PDB ID: 2O8B numaralı yapılar kullanılarak DNA'ya bağlı olmayan sistemler hazırlanmıştır. Yabanıl tip ve mutant proteinler için simülasyon sonuçlarını karşılaştırabilmek üzere her yapı için iki replika sistem oluşturulmuştur. Proteinler için sistem hazırlanırken hücrede bulundukları fizyolojik ortam ve koşullar dikkate

alınarak gerekli iyon konsantrasyonu sağlanmıştır. Yapılar, su kutusunda çözdürüldükten sonra kısa minimizasyon ve denge simülasyonları gerçekleştirilmiştir. Dengeye gelen yapılar her replika için 200 ns boyunca simüle edilmiştir. Yabanıl tip ve üç adet mutant protein için ikişer replika simülasyon yürütülmüş ve toplamda 1600 ns uzunluğunda MD simülasyonu gerçekleştirilmiştir.

Simülasyon sonuçları ile kök ortalama kare sapması (RMSD), kök ortalama kare dalgalanması (RMSF), ısı kapasitesi, etkileşim analizi ve temel bileşenler analizi (PCA) gerçekleştirilmiştir. RMSD analizi, simülasyonun dengeye gelip gelmediği ve konformasyon değişikliği olup olmadığı hakkında bilgi vermektedir. RMSD ortalamalarına bakıldığında mutant proteinlerin RMSD değerlerinin yabanıl tip proteine göre daha yüksek olduğu bu nedenle mutasyonların protein konformasyonunda bir miktar değişime neden olabileceği düşünülmüştür. Ancak replikalar arasındaki RMSD değerlerine bakıldığında bazı yapıların henüz dengeye gelmediği görülmüştür. RMSF analizi, yapısal esneklik, hareketlilik ve termal stabilite ile ilgili çıkarımlar yapabilmek için kullanılmaktadır. Çalışmada beş bölge için ayrı ayrı dalgalanmalara bakılmıştır. Yabanıl tip protein ile kıyaslandığında mutant proteinlerde genel olarak daha yüksek dalgalanma olduğu görülmüştür. Mutasyonların bulunduğu bölge itibariyle doğrudan etkilemeyeceği bölgelerde meydana gelen dalgalanma hareketi, mutasyonların protein dinamiğinde meydana getirdiği değişimlerin farklı bölgelerde ortaya çıktığını da göstermektedir. Yabanıl tipe göre en çok dalgalanmanın görüldüğü bölge kıskaç bölgesi olmuştur. Bu bölge için simülasyon başında ve sonundaki değişime bakıldığında mutasyonun gerçekleştirği proteinlerin kıskaç bölgererinin mutasyondan etkilenmiş olduğu sonucuna varılmıştır. Isı kapasitesi, proteinin termodinamik özellikleri ile ilgili bilgi verebilmektedir. Analiz sonucu elde edilen bulgular, tüm mutant yapıların ısı kapasitesinin yabanıl tip proteine göre daha yüksek olduğunu ortaya koymakta ve proteinin termodinamik özelliklerinin farklılaşmış olabileceğini göstermektedir. Bu sonuçlar, mutasyonların proteinin açılması yönünde etkisi olabileceğini düşündürmektedir. Etkileşim analizi sonucunda mutant yapıların yabanıl tip proteine kıyasla hidrojen bağı ve tuz köprüsü etkileşimlerinde değişimler olduğu gözlemlenmiştir. Yabanıl tipte sıklıkla gözlemlenen etkileşimlerin bazıları kaybolurken yabanıl tipte gözlemlenmeyen yeni etkileşimler kurulduğu da görülmüştür. PCA analizi ile protein içideki en baskın iki hareket elde edilmiş ve bunların proteinin hangi bölgelerinde etkili olduğu gözlemlenmiştir. Proteindeki en baskın hareket MSH2 proteini üzerinde görülmüş ve protein üzerinde döndürme etkisi olabileceği düşünülmüştür. MSH6 proteini üzerinde ise en baskın ikinci hareket etkilidir. Bu hareket çoğunlukla kıskaç bölgelerinde etkili olmuş ve kıskacın dışarı yönünlü hareketini etkilediği düşünülmüştür. Ayrıca serbest enerji yüzeyinde proteinlerin MD simülasyonları boyunca uzayda örnekledikleri bölgeler gösterilmiştir.

Bu çalışmada, DNA'ya bağlı olmayan MutSα proteininde meydana gelen mutasyonların protein yapı ve dinamiğine etkileri incelenmiş ve mutant yapıların yabanıl tipe göre daha farklı termodinamik ve yapısal özellikler gösterdiği sonucuna varılmıştır.

1. INTRODUCTION

1.1 Purpose of Thesis

The main purpose of the thesis is to investigate the structural differences between mutant and WT MutSα proteins with molecular dynamics (MD) simulations. MutSα, one of the DNA mismatch repair (MMR) proteins, enables the recognition and repair of damaged DNA. Mismatches/damages in MMR proteins can prevent the repair of damaged DNA, accumulate mutations in the cell, and cause cancer. Mutations in some MMR proteins cause pathogenic protein variants. Some pathogenic variants are known to cause hereditary colon cancer. More than half of the mutations that cause hereditary colon cancers are in the monomers MSH2 and MSH6, which form the MutSα complex. Samples taken from patients with a family history of cancer were examined with genetic screening tests by Prof. Dr. Levent Doğanay and his research group, and as a result, the investigation of Ala733Thr (A733T), Arg577Cys (R577C) and Ser1279Asn (S1279N) mutations were deemed appropriate for further molecular testing. Within the scope of the TÜBİTAK 1003 project, grant numbers 318S127 and 318S129 these mutations were shared with Assoc. Prof. Mert Gur and his research group for molecular dynamics investigation. In this thesis, MutSα protein systems with these mutations were prepared and MD simulations were performed to compare molecular structure and dynamics and also thermodynamic properties of mutants with those of the wild-type (WT) protein.

1.2 DNA Repair Mechanisms and Mismatch Repair Proteins in Eukaryotes

1.2.1 DNA Repair Mechanisms

Cells in the human body are exposed to DNA-damaging endogenous or exogenous agents thousands of times a day (Lindahl and Barnes, 2000; Schärer, 2003). These factors, which cause genomic instability and structural damage in DNA (Tubbs and Nussenzweig, 2017), can cause lesions such as base and sugar modifications, strand breaks, DNA-protein cross-links, and insertion/deletion, (Cadet et al., 1997; Friedberg,

2003; Dexheimer, 2013) leading to disruptions in fundamental cellular processes such as replication and transcription (Hakem, 2008). In addition, specific DNA damages that may occur can cause aging, cancer and some diseases (Hoeijmakers, 2009). Cells have many DNA repair mechanisms such as mismatch repair, double-strand break repair, base excision repair, nucleotide excision repair for repairing DNA damages (Onur et al., 2009; Dexheimer, 2013).

Base excision repair (BER) is a DNA damage repair mechanism that is responsible for repairing damaged DNA bases to protect against cancer and aging (Krokan et al., 2000; Fromme and Verdine, 2004). DNA glycosylases, which are lesion-specific and known to have at least 12 types, are the repair initiation enzyme responsible for the recognition and removal of damaged bases (Dexheimer, 2013; Krokan and Bjørås, 2013). The base is removed by cleavage of the glycosidic bond between the damaged base and deoxyribose, leaving behind the abasic site or AP site (Dizdaroglu, 2005). In this region, AP-endonuclease (APE1) hydrolyzes the 5' phosphodiester backbone, allowing the synthesis/ligation step to begin. In the synthesis/ligation step, the gap is filled with DNA polymerase, and finally, DNA ligases establish the establishment of the last bonds and complete the damage repair by ensuring that the gaps are completely closed (Verdine and Bruner, 1997; Fromme and Verdine, 2004). Repair occurs in nuclei and mitochondria (Krokan and Bjørås, 2013).

Nucleotide Excision Repair (NER) is responsible for recognizing voluminous lesions caused by UV rays and some chemotherapeutic agents and repairing these lesions by removing them from DNA (Schärer, 2003; Kusakabe et al., 2019). Although similar to BER, NER has a more complex mechanism using 30 different proteins (Dexheimer, 2013; Schärer, 2013). In this mechanism, which can start in two ways: global genome NER (GG-NER) or transcription-coupled NER (TC-NER), GG-NER provides repair of lesions anywhere on the genome (Petruseva et al., 2014; Puumalainen et al., 2016), while TC-NER repairs lesions formed during transcription (Hanawalt and Spivak, 2008). Disturbances in the NER mechanism have been shown to result in genetic diseases (Kraemer et al., 2007) related to sun sensitivity such as Cockayne syndrome and trichothiodystrophy (Gillet and Schärer, 2006), or results in cancer and premature aging (Dexheimer, 2013). After the lesions are recognized, common reactions occur for the mechanism that starts in both ways. As a result of the multi-step cutting and patching process in which many different proteins are used, the excision of the region

containing the lesion of approximately 30 nucleotides takes place (Dexheimer, 2013; Schärer, 2013). In the final step, DNA polymerase d or e correctly synthesizes the removed region and the repaired region is completely closed with the help of DNA ligase (Dexheimer, 2013).

Double-strand breaks (DSBs) are one of the most dangerous DNA damages and if not corrected (Trenner and Sartori, 2019), they cause chromosome breakage and cell death (Dexheimer, 2013), while in the case of faulty repair, they cause cancer and genomic instability (Aparicio et al., 2014). Most of the damage to the DNA backbone is caused by reactive oxygen species and ionizing radiation (Vilenchik and Knudson, 2003; Vítor et al., 2020). Unlike other types of lesions, DBSs are remarkable because both DNA sequences are damaged (Dudáš and Chovanec, 2004; Schärer, 2013) . Since there is no template available for repair, homologous recombination (HR) and non-homologous end-joining (NHEJ) mechanisms have evolved for this condition (Pâques and Haber, 1999; van Gent et al., 2001). While HR uses the sister chromatid as the template, NHEJ repairs the DBSs by connecting the broken ends (Vilenchik and Knudson, 2003; Dexheimer, 2013). Many proteins are involved in the repair of DSBs in HR. The first stage in HR is the 5′ to 3′ resection (Dudáš and Chovanec, 2004; Dexheimer, 2013). At this stage, DNA ends with DSB are cut to form 3' single-stranded DNA (ssDNA). The next step is DNA sequence homology. Single-stranded DNA interacts with the sister chromatid, which is used as a template, and initiates invasion. DNA synthesis is initiated from the 3' end of single-stranded DNA by DNA polymerase h. Finally, an faultless repair is achieved by ligation with DNA ligase I (Li and Heyer, 2008; Dexheimer, 2013). HR is a crucial mechanism, and mutations in genes encoding BRCA1 and BRCA2 proteins, which are among the proteins involved in this mechanism, are associated with breast and ovarian cancers (O'Donovan and Livingston, 2010; Trenner and Sartori, 2019).

Mismatch repair is responsible for correcting incorrect base pairings that escape the error-correcting activity of polymerases during replication (Manhart and Alani, 2017), and repairing insertion/deletion loops (IDL) that can occur in the replication of repetitive DNA sequences due to polymerase slippage (Modrich and Lahue, 1996; Li, 2008). MMR is carried out by the MutS and MutL protein complexes (Modrich, 2006; Fukui, 2010). Two heterodimer protein complexes are involved in the MMR system, which starts with the mismatch recognition step (Kunkel and Erie, 2005). The MutSα

heterodimer is involved in recognition of base-base mismatches and small IDLs (Drummond et al., 1995; Marsischky et al., 1996), while the MutSβ heterodimer is involved in recognition of larger IDLs (Palombo et al., 1995; Acharya et al., 1996) . In the second step, the defective DNA strand is opened and excision of the lesion site takes place (Dexheimer, 2013). In the last step, repair synthesis is performed after removing the faulty DNA base, in which the gap is filled with the correct bases (Li, 2008; Dexheimer, 2013).

1.2.2 Human Mismatch Repair (MMR) Proteins

It is known that humans have many MMR proteins such as MutS, MutL, and MutH heterodimers formed by MSH2, MSH6, MSH3, MLH1, MLH3, PMS2, and PMS1 (Wang et al., 2016; Bateman, 2021). These heterodimer proteins, nuclease, and polymerase work together to repair defective DNA (Edelbrock et al., 2013). It is known that the MutSα complex, formed by the MSH2-MSH6 heterodimer is responsible for the recognition and repair of base substitution mismatches and insertion/deletion mismatches (Warren et al., 2007; Edelbrock et al., 2013), while the MSH2-MSH3 heterodimer, also known as the MutSβ complex, is involved in the repair of insertion/deletion mismatches (Kolodner and Marsischky, 1999; Peltomäki, 2001).

MMR system induce apoptosis by initiating a cellular response at specific DNA damages. Missing or defective MMR proteins prevent the cell from undergoing apoptosis despite specific DNA errors (Iyer et al., 2006; Negureanu and Salsbury, 2012). It has been observed that mutations in genes encoding DNA mismatch repair proteins may cause the inactivation of tumor suppressor genes, and as a result, increased mutations in DNA may cause colon cancer (Fishel et al., 1993; Leach et al., 1993). Spontaneous mutation rates in cells with defective MMR pathways increase, resulting in cancer-prone cells (Woods et al., 2007; Hsieh and Yamane, 2008). In a previous study, two knock-in mouse strains harboring the Msh2 G674A or Msh6 T1217D mutations were produced and the mice developed cancer as a result of the experiment. In addition, microsatellite instability (Velho et al., 2014) and production of embryonic fibroblasts with damaged MMR were observed in these mice (Geng et al., 2012). Another study was conducted with Saccharomyces cerevisiae. Yeast MutSα proteins carrying the G693A mutation in MSH2 and the G1067D mutation only in

MSH6 cause the dominant mutator phenotype (Drotschmann et al., 1999; Das Gupta and Kolodner, 2000). Another consequence of the mutation in MSH2 was that the protein retained its defective binding to MSH6 or mismatch binding activity, while causing failure in mismatch recognition and hydrolysis (Loconte et al., 2014). Therefore, a damaged or missing MMR pathway has been associated with cancer (Martin et al., 2010; Negureanu and Salsbury, 2012). It was reported that almost half of the mutations that cause HNPCC are in the genes encoding the MutSα complex (Fishel and Kolodner, 1995; Peltomäki, 2003). Hereditary non-polyposis colon cancer (HNPCC), which constitutes 1-5% of all colon cancers (Jacob and Praz, 2002), is the most common type of cancer due to the inactivation of MMR proteins (Carethers and Stoffel, 2015). Families with a history of HNPCC have a risk of colorectal carcinoma at a rate of 70-85% (Leach et al., 1993; Peltomäki, 2001).

1.3 MutS Proteins and Human MutSα Protein Structure

1.3.1 Prokaryotic and Eukaryotic MutS Proteins

The discovery of the MutS gene dates back to the mid-1950s, and it was introduced as a mutator gene when it was seen that mutations in this gene increased spontaneous mutation in bacteria (Treffers et al., 1954). By the mid-1970s, a group of mutator genes, including the MutS gene, were found to be involved in MMR (Tiraby and Fox, 1973; Fishel and Lee, 2016). MutS homologues in E.coli are found in nearly all prokaryotic and eukaryotic organisms (Jiricny, 1998; Fishel and Lee, 2016). Studies in E.coli revealed that there is only one gene encoding MutS proteins (Su and Modrich, 1986; Allen et al., 1997), but with human genetic studies since the early 1990s, it has been observed that human MutS proteins are encoded by multiple mutS homologous (MSH) genes (Culligan et al., 2000; Obmolova et al., 2000). It is also known that all members of the MutS family have conserved ATPase activity (Jiricny, 1998). Compared to human MutSα proteins, E.coli MutS proteins have very low homology and are approximately 600 amino acids shorter than MutSα. For prokaryotic MutS that share limited sequence identity, the identity with the conserved regions of MSH2 and MSH6 of human MutSα was 21% and 24%, respectively. Therefore, the use of prokaryotic MutS in cancer research has been limited. There are also studies showing that the human MutSα protein is involved in apoptosis and immune enhancement, and in this respect it differs from the prokaryotic MutS (Warren et al., 2007).

1.3.2 Human MutSα heterodimer structure

In humans, newly synthesized defective DNAs are recognized by the MutSα and MutSβ, MutS homologs proteins (Jacob and Praz, 2002). The MutSα complex is 260 kDa and consists of MSH2 and MSH6 polypeptide subunits. In the human MutSα structure obtained by Warren et al, a full-length structure was obtained for MSH2. Since the protease-resistant fragment at the N-terminus of MSH6 could not be obtained, the first 340 amino acids in the MSH6 structure are missing. However, it is known that MSH6 has a disordered N-terminal which differs from MSH2 (Edelbrock et al., 2013).

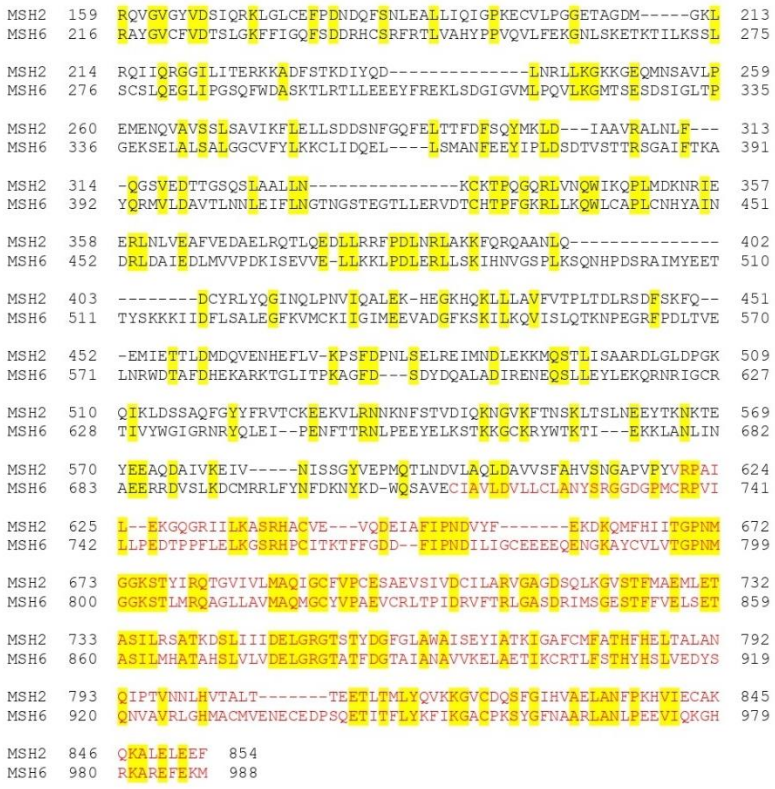


Figure 1.3.2.1. Sequence alignment of the human MutSα protein complex subunits MSH2 and MSH6. Residues identical in both MSH2 and MSH6 are highlighted according to BLOSUM62. Sequences in red are the ATPase domain sequence found in both proteins.

In the Warren et al. study which the structure was obtained, the human MSH2 and MSH6 proteins and the yeast MSH2 and MSH6 proteins were structurally aligned, and the sequence similarity of the proteins was low. The human MSH2 and MSH6 proteins have similar secondary structures and domains that perform similar functions (Warren et al., 2007). The structural alignment of MSH2 and MSH6 proteins, which form the

MutSα heterodimeric structure, is given in Figure 1.3.1. The most sequence similarity was observed in the ATPase domain, which is known to be shown the most conservation in all MutS homologs (Sachadyn, 2010).

MSH2 and MSH6 subunits can be structurally divided into five domains, each consisting of at least 800 residues. These domains are named lever domain (residue 300 to 456 and 554 to 619 in MSH2, residue 718 to 934 and 1009 to 1075 in MSH6), clamp domain (residue 457 to 553 in MSH2, residue 935 to 1008 in MSH6), mismatch binding domain (residue 1 to 124 in MSH2, residue 326 to 518 in MSH6), connector domain (residue 125 to 297 in MSH2, residue 519 to 717 in MSH6), and ATPase domain (residue 620 to 855 in MSH2, residue 1076 to 1355 in MSH6) (Warren et al., 2007; Mukherjee and Feig, 2009; Negureanu and Salsbury, 2012).

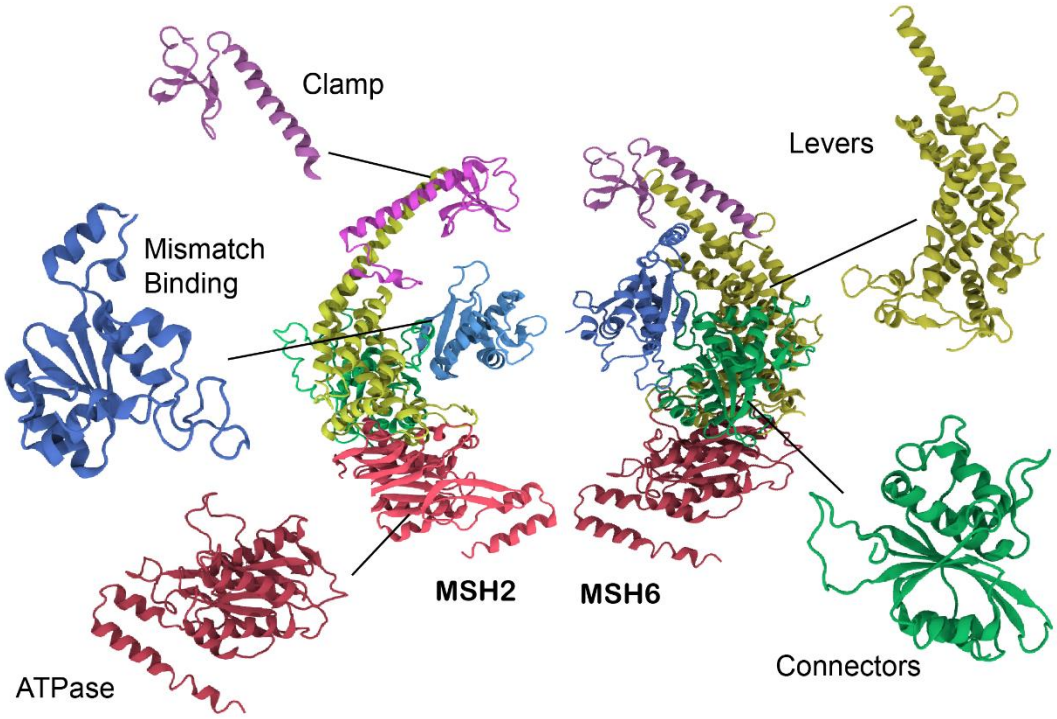


Figure 1.3.2.2. Domains of the MutSα protein. Mismatch binding domain, connector domain, levers domain, clamp domain and ATPase domain are found in both MSH2 and MSH6 monomers colored as blue, green, yellow, magenta, and red, respectively.

The mismatch binding domain on MSH6 is a conservative structure, it recognizes the mismatch region and establishes specific contacts. The connector domain also called the interdomain, is where allosteric signals occur. The lever domain is responsible for signaling between ATPase and DNA binding domains, while the clamp domain

stabilizes DNA by non-specific binding to the DNA backbone (Mukherjee and Feig, 2009).

1.4 Domains of Human MutSα Protein Complex

The mismatch binding domain (Figure 1.3.2) is the domain that binds to the mismatches on the DNA. DNA binding domains have α/β structure and their structures and functions differ in MSH2 and MSH6 subunits (Iaccarino et al., 1998). It is known that the 14 amino acids in the N-terminal of MSH2 block the DNA binding face and establish only one connection with the DNA backbone to stabilize DNA.stabilizing DNA. The domain in MSH6 stabilizes MutSα on substrates such as C-C mispair by establishing non-specific interactions with DNA (Warren et al., 2007). Studies show that the mutations in the mismatch binding domain can cause the blocking mismatch binding to DNA and affect the apoptosis process (Dufner et al., 2000; Lin et al., 2004).

The connector domains (Figure 1.3.2) are located between the levers and ATPase domains, making allosteric signaling between them. The MSH2 and MSH6 connector domains are located in different positions. It is thought that domains containing three surface loops can regulate the protein-protein interaction (Warren et al., 2007) and interacts with MutL homologs (Mendillo et al., 2009).

Lever domains (Figure 1.3.2) with long a-helical structures interact with the ATPase and the clamp domains. Although this domain is a less conserved structure among other homologous proteins, the presence of a highly conserved loop region within the domain suggests that this loop provides signal transduction between ATPase and mismatch binding domain (Warren et al., 2007).

The clamp domain (Figure 1.3.2) in MSH6, which consists of a short α-helical and largely β-strand structure, makes contact with approximately six base pairs of B form DNA, while domains in MSH2 make contact with bases on either side of the mismatch (Warren et al., 2007). Clamp and lever domains show conformational changes during DNA binding, allowing the gaps in the protein to be opened or closed. When MutSα recognize a mismatch on DNA, it formed clamp around the mismatch (Groothuizen et al., 2015). In one of the studies on DNA-free MutSα, when the DNA in the MutSα structure was removed and the behavior of the structure was examined with MD simulations, it was seen that after 40 ns, the clamp and lever domains moved in the

opposite direction and opened the cavity where the DNA should have been (Mukherjee and Feig, 2009). Therefore, it was concluded that binding the clamp domain to DNA stabilizes the protein. Else, the protein showed other conformations (Warren et al., 2007; Mukherjee and Feig, 2009).

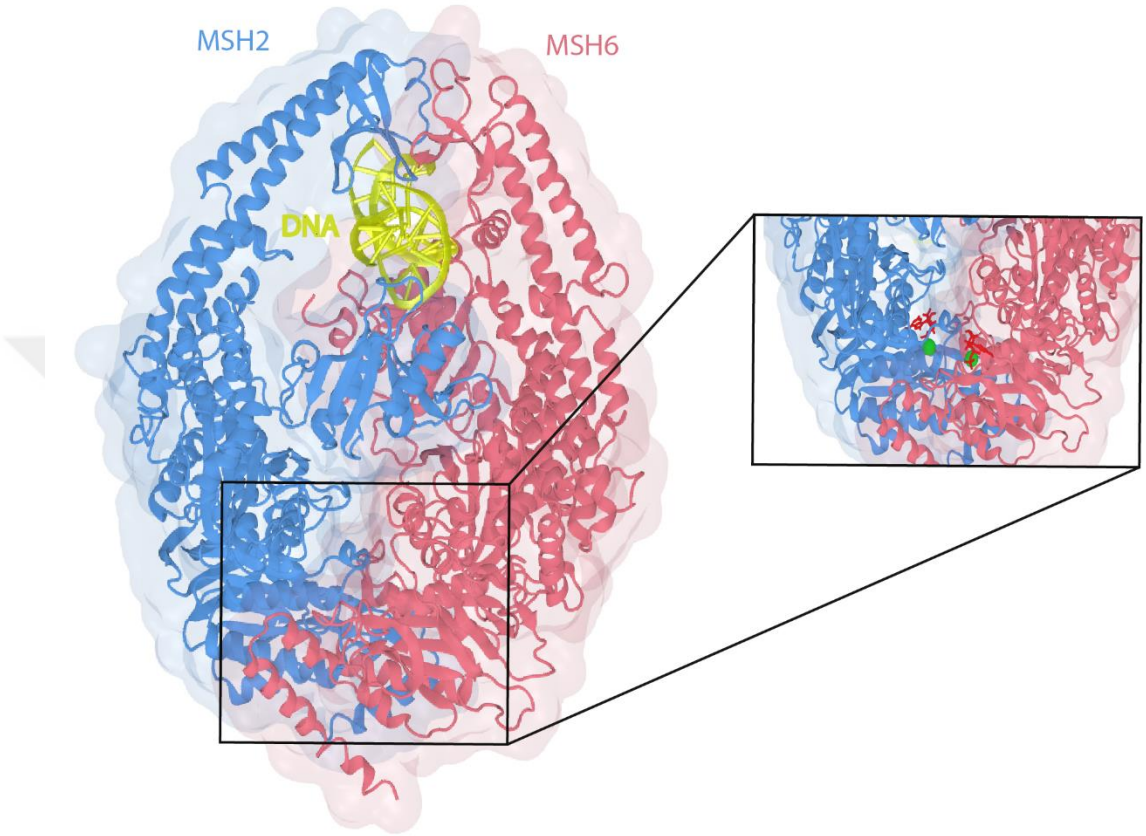


Figure 1.4.1. MutSα heterodimer with DNA and ADP. ADP (red) and Mg^{2+} ions (green spheres), and DNA (yellow ribbon) are represented. The ATPase domains of MSH2 and MSH6 monomers interact, and ADP and Mg^{2+} ions are found in the interface of ATPase domains, shown in the extended panel.

The last domain, the ATPase domains (Figure 1.4.1), are a region located at the C-Terminal of MSH2 and MSH6 and members of the ATP binding cassette (ABC)-transporter superfamily (Gorbalenya and Koonin, 1990; Lamers et al., 2003). Both MSH2 and MSH6's ATPase domains have highly conserved sequences and interact with each other. These interactions are necessary to stabilize the ATPase interface in the absence of ATP (Warren et al., 2007; Mendillo et al., 2009).

1.5 Human MutSα Substrate Recognition and Reaction Cycle

MutSα should detect the smallest anomaly in long DNA strands (Heinen, 2016). Studies with prokaryotic MutS reveal that correctly synthesised and defective DNA binding to MutSα is different. The interactions of prokaryotic MutS with defect-free DNA do not cause bending in DNA (Schofield et al., 2001; Warren et al., 2007). However, it has been concluded that the protein interacting with the faulty DNA causes sharp bends on the DNA, thus can make extra interactions with the defective DNA. Studies show that MutSα binding changes according to the errors that occur in DNA. For example, in MutSα and G●T mispaired DNA complex, Glu434 residue in the DNA binding domain of MSH6 makes a hydrogen bond (H-bond) with mispaired thymine. Following the hydrogen bonding of Glu and thymine, the backbone carbon of Val429 forms a hydrogen bond with Guanine. As a result of hydrogen bonds and other non-specific interactions, the mispaired thymine area is bent, resulting in the widening of the minor groove (Warren et al., 2007). However, MSH2 does not establish H-bond or non-specific interactions with DNA (Schofield et al., 2001; Warren et al., 2007). The fact that MSH6 makes interactions that can provide DNA bending alone confirms that MSH2 does not have similar interactions. On the other hand, Glu434 cannot form hydrogen bonds in cases such as C-C mispair. Therefore, non-specific interactions seem sufficient to perform DNA binding and bending functions (Bowers et al., 1999; Warren et al., 2007).

In the MMR system, signalling must inform other repair proteins while preserving the information about the location of a lesion that has occurred in the DNA (Stojic et al., 2004). Three models have been proposed for how DNA excision repair works. The first model is that MutSα induces DNA bending to denote the mismatch site and bring other necessary repair factors into close proximity (Blackwell et al., 1998; Blackwell et al., 1998). In the second model, MutSα bound to the mismatched substrate induces ATP-dependent conformational change (Gradia et al., 1997; Gradia et al., 2000). It moves on the DNA to encounter other repair proteins such as endonucleases, exonucleases and helicases. In the last model, it is thought that mismatch binding and ATP binding occur simultaneously (Junop et al., 2001; Alani et al., 2003). In this way, starting repair is prevented before connecting to mispair bases (Alani et al., 2003; Warren et al., 2007).

1.6 Molecular Dynamics (MD) Simulation

Proteins are stable dynamical systems in which internal movements determine their function. In recent studies, the dynamics of structure-function relationships of proteins have been investigated and concluded that movements in the protein cause net atomic displacements (McCammon, 1984; Negureanu and Salsbury, 2012). Movement in protein dynamics has been associated with critical properties in proteins, such as binding and signal transduction. While it is challenging to investigate the dynamics of proteins experimentally at atomic resoltion, MD simulations provide a crucial atomic insight to understand the effect protein dynamicson protein function. MD simulations correlate the movements occurring in the protein on the scale of space and time, allowing the movements in the protein to be sampled (Debrunner and Frauenfelder, 1982; Negureanu and Salsbury, 2012).

Most of the studies on mismatch repair have been performed with prokaryotic MutS protein (Negureanu and Salsbury, 2012). The increase in studies with eukaryotic MutSα is an essential step toward understanding the function of this protein and the communication between domains in the protein. The increase in MutSα studies will provide insight into the detection of cancer-causing mutations in MMR proteins, the effects of these mutations and the changes in the conformation of the protein (Mukherjee and Feig, 2009; Negureanu and Salsbury, 2012).

Here, the atomic details of the structural differences between WT and mutant forms of MutSα are investigated, based on MD simulations of MSH2 (1-854)- MSH6 (362-1335) with ADP and without DNA. Simulations were initiated from the crystal structure PDB ID: 2O8B (Warren et al., 2007) in the absence without the DNA found in the crystal structure.

2. METHODS

2.1 Starting Structure and System Preparation

The all atom MD simulations can be used to examine the structure and dynamics of proteins at all atom resolution. They mimic the physiological conditions in which proteins are found, and allow the behavior of proteins to be studied in the environment. The systems were prepared for MD simulations using the crystal structure with PDB ID: 2O8B (Warren et al., 2007). In this structure, the MutSα complex is bound to ADP and DNA with G:T mismatch. Bound DNA structures were removed, and all simulations were carried out with the protein bound to an ADP. The DNA-free structure was dissolved in the TIP3P solvent box with 12 Å padding in each direction. The system was ionized with 20 mM sodium chloride (NaCl) and 210 mM potassium chloride (KCl) because of the cell nucleus's physiological conditions (Nagy et al., 1981; Bischof et al., 2017). The final dimensions of the solvated and ionized systems were 137 × 136 × 103 Å3 with 211,803 atoms (Figure 2.1.1).

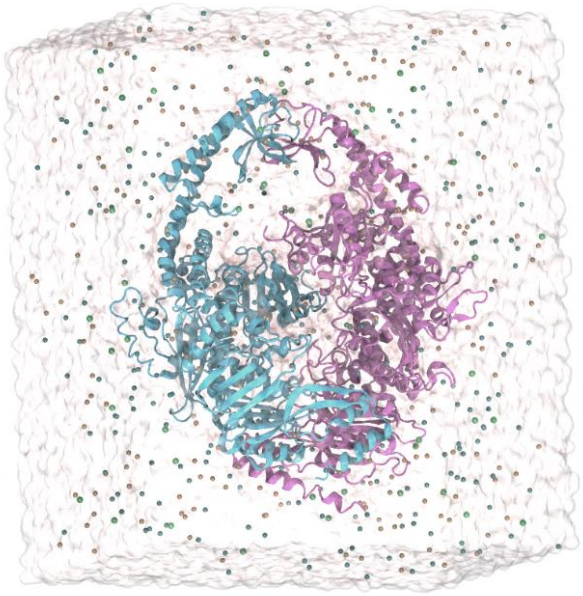


Figure 2.1.1. MutSα protein shown as blue (MSH2) and purple color (MSH6) in water box which have 210 mM KCl and 20mM NaCl ion concentration.

Mutations of MutSα protein A733T, R577C and S1279N (Figure 2.1.2) were performed by using the VMD Mutator plugin (Humphrey et al., 1996).

2.2 Molecular Dynamics Simulations

CHARMM36 all-atom force-field (Best et al., 2012) was used to describe molecular interactions and simulations were run with NAMD 3.0alpha8 software (Phillips et al., 2005; Kim et al., 2020). SHAKE algorithm was used for all MD simulations with a 2 fs time step. The Particle Mesh Ewald (PME) technique was used to calculate long-range electrostatic interactions, and also cut-off of 12 Å was applied for Van Der Waals (vdW) interactions. All simulations were carried out in the NPT ensemble and keeping the pressure 1 atm using the Langevin Nosé–Hoover method with an oscillation period of 100 fs and damping time scale of 50 fs. The temperature was kept constant at 310 K using a 1 ps damping coefficient for Langevin dynamics to maintain isothermal conditions.

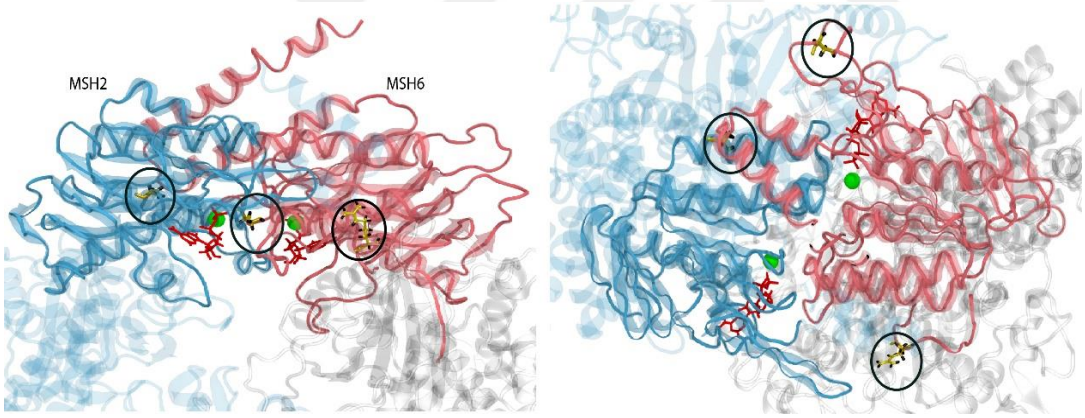


Figure 2.1.2. Front (left) and top (right) view of the MutSα protein ATPase domain. Two of the performed mutations are in the ATPase domain and the other is close to this domain. Mutated residues are shown in yellow. ADPs are represented with red color, and the green spheres represent Mg^{2+} atoms.

The systems were firstly minimized and equilibrated by harmonic constraints on the protein's C_α, ADP and Mg^{2+} atoms allowing only the water molecules to move freely. For this purpose, systems were minimized for 10,000 steps without any constraints. Then heavy atoms of protein, ADP and Mg^{2+} atoms were fixed, and water molecules were equilibrated for 2 ns. The second minimization was performed with 10,000 steps without any constraints. After all atoms in the systems were minimized, systems were equilibrated for 20 ns with 2 $kcal \cdot mol^{-1} Å^{-2}$ harmonic constraints on the protein's C_α, ADP and Mg^{2+} atoms. Subsequently all constraints were removed and a final

equilibration simualtion of 5ns lengt were performed. Then, production runs were initiated each being of 200ns length.

2.3 RMSD Analysis

The structural distance at certain coordinates can be calculated by Root-Mean-Square Deviation (RMSD) analysis. Using the RMSD, the distance between the backbone atoms of the protein can be measured, and conclusions can be drawn about its structural change and stability (Phillips et al., 2005). RMSD between two structures is calculated according to equation 2.1.

$$\boldsymbol{RMSD\alpha(tj)} = \sqrt{\frac{\sum_{\alpha=1}^{N\alpha}(\vec{r}\alpha(tj) - \langle\vec{r}\alpha\rangle)^2}{N\alpha}} \qquad (2.1)$$

The number of atoms by which the change between positions is measured is *Nα*. $\vec{\mathbf{r}}\boldsymbol{\alpha}(\mathbf{t}\boldsymbol{j})$ mean position of $\boldsymbol{\alpha}$ atom's at time *tj* and $\langle\vec{\mathbf{r}}\boldsymbol{\alpha}\rangle$ mean average value of $\boldsymbol{\alpha}$ atom positions, which is compared to $\vec{\boldsymbol{r}}\boldsymbol{\alpha}(\boldsymbol{tj})$.

2.4 RMSF Analysis

Root Mean Square Fluctuation (RMSF) measures a residue's elasticity, or deviation, from its mean position throughout the simulation. With the RMSF calculation, the most fluctuating structural regions can be identified. RMSF values are calculated according to equation 2.2.

$$\langle\Delta\boldsymbol{R_i^2}\rangle^{1/2} = \langle(\boldsymbol{R_i} - \langle\boldsymbol{R_i}\rangle)^2\rangle^{1/2} \qquad (2.2)$$

The average of the coordinates of the ith C_α atoms during the simulation is $\langle R_i\rangle$ and R_i is the ith C_α atoms instant coordinate.

2.5 Heat Capacity Analysis

The heat capacity is important to investigate the effects of temperature on the phase and reaction equilibrium, and order-disorder phase transition of protein folding (Yeh et al., 2008). As previously mentioned under the heading "Molecular Dynamics Simulations", all MD simulations were performed under NPT conditions (constant

pressure (P), temperature (T), and particle (N)). The heat capacity under constant pressure was calculated according to equation 2.3.

$$\boldsymbol{C_p = (\partial H / \partial T)} \tag{2.3}$$

The heat capacity can be calculated as the Mean Squared Fluctuation of energy values in equation 2.4 (Jorgensen, 1982; Prabhu and Sharp, 2005).

$$\boldsymbol{C_p = \frac{\langle H^2 \rangle - \langle H \rangle^2}{NkT^2}} \tag{2.4}$$

Enthalpy values are extracted from MD simulations by assuming that the potential energy of the force fields applied in the MD simulations is the same as the enthalpy (van der Kamp et al., 2018). Thus, the necessary information to calculate the variance of the enthalpy values is available in the MD simulations. Potential energies for each time instant along the MD simulations were obtained using the NAMDEnergy plugin of the NAMD software (Phillips et al., 2005).

2.6 Principle Component Analysis

The MutSα conformations sampled during MD simulations were structurally aligned with the C_α atoms of the helical and beta sheet regions.. Thus, removing translational and rotational movements of the protein while keeping the internal movements of the protein. For Principle Component Aanalysis (PCA), the covariance matrix C was constructed using the 3x1147 configuration vector with the instantaneous C_α atom of the helical and beta sheet regions of the protein calculated according to equation 2.5.

$$\boldsymbol{C = \langle (R - \langle R \rangle)(R - \langle R \rangle)^T \rangle} \tag{2.5}$$

3N dimensional configuration vector is represented as R, and trajectory average of configuration vector is represented as $\langle \boldsymbol{R} \rangle$. The eigenvalue decomposition of C for PCA is calculated according to equation 2.6.

$$\boldsymbol{C = \sum_{i=1}^{3N} \sigma_i p_i p_i^T} \tag{2.6}$$

Each pi specifies a PC corresponding to the σi variance. PC values are sorted in descending order of σi values. Therefore, p1 (first PC) has the largest variance and represents the most significant move. The conformations that the MutSα proteins sample in MD simulations are used to create the free energy surface. PC1 shows the most dominant movement and PC2 shows the second most dominant movement.

3. RESULTS AND DISCUSSION

3.1 RMSD Analysis

All conformations sampled in the MD simulations were aligned via their alpha helix and beta sheet C_α atoms to their starting structures, and subsequently the RMSD to their starting structures were evaluated. Two replica of MD simulations were performed for each of the A733T, R577C, S1279N mutants, and WT. The evolution of RMSD values along their trajectories are shown in Figure 3.1.1. They all converged within 200 ns of simulation lengths.

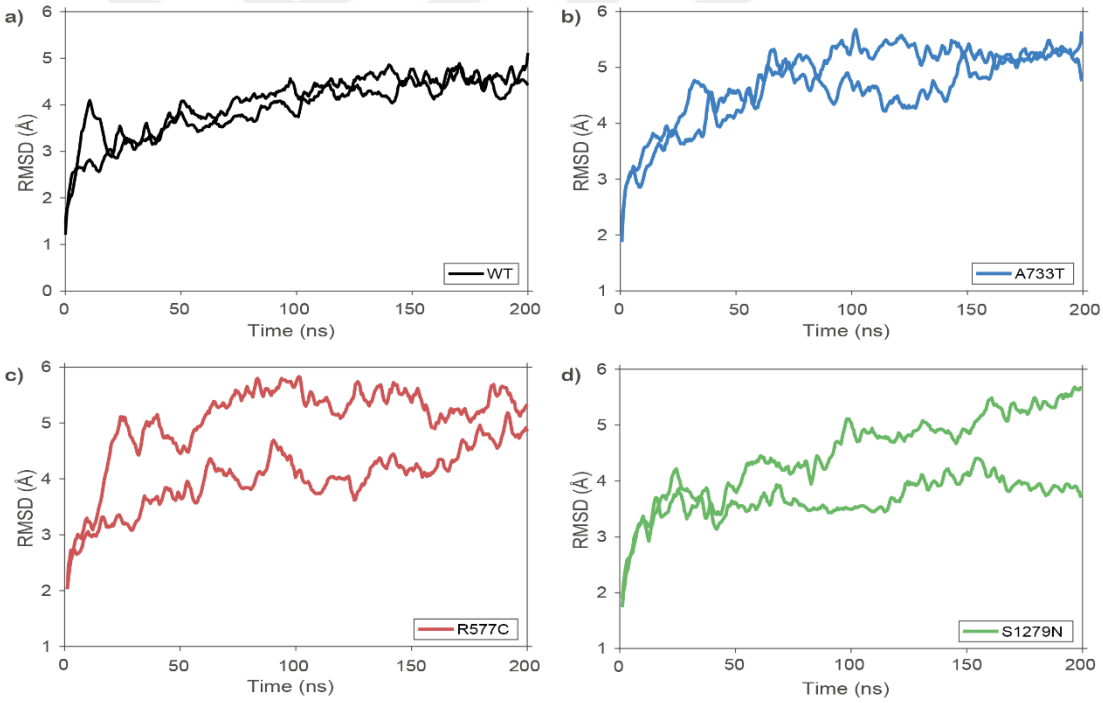


Figure 3.1.1. MutSα protein RMSD results for two replicas shown as a) WT, b) A733T mutant, c) R577C mutant, and d) S1279N mutant.

The RMSD curves for the WT replicas showed similar characteristics and converged towards each other. Similarly, the RMSD curves of A733T converged toward each other. However, for R577C and S1279N, the RMSD curves of the replicas for each mutant on the other showed different characteristics and converged to different values.

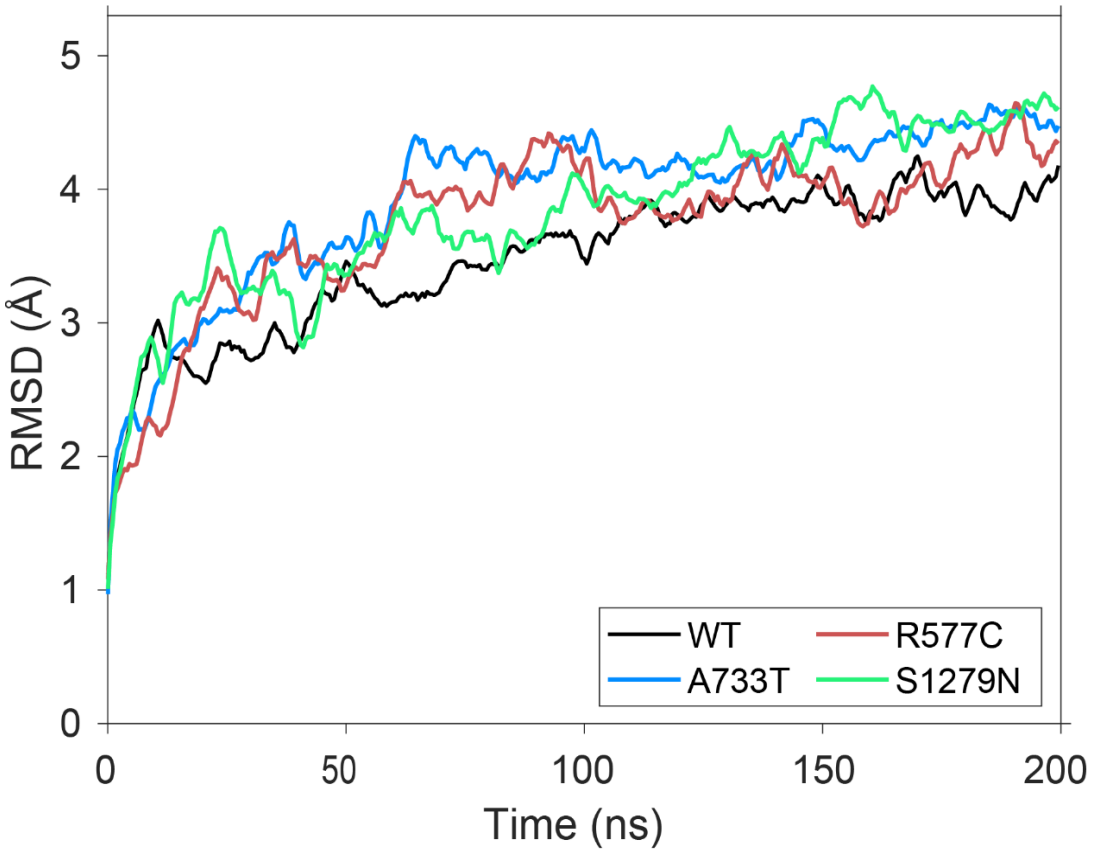


Figure 3.1.2. Average RMSD value of WT and mutant MutSα proteins based on the replica of MD simulations. Average RMSD values of WT (black), A733T mutant (blue), R577Cys mutant (red), and S1279N mutant (green) are shown.

The average of the curves shown in Figure 3.1.1 are shown in Figure 3.1.2.Based on the replica averages it is seen that the average RMSD values of all mutant proteins were slightly higher than the WT protein. Thus, mutations caused changes in the dynamics and structure the protein.

3.2 RMSF Analysis

RMSF calculation was performed to determine the structural flexibility and changes in residue fluctuations. Since MutSα is a large protein with many domains, RMSF values of each domain were investigated separately. The RMSF analysis results for the mismatch binding domain are shown in Figure 3.2.1. RMSF values were evaluated by combining the two replicas of trajectories into a single trajectory. To calculate MSH6 RMSF values all conformations were aligned with their starting conformations via the MSH2 alpha helical and beta sheet C_α atoms. Similarly, for MSH2 RMSF calculations conformations were aligned via MSH6. This was performed to highlight the relative residue fluctuations of MSH2 and MSH6.

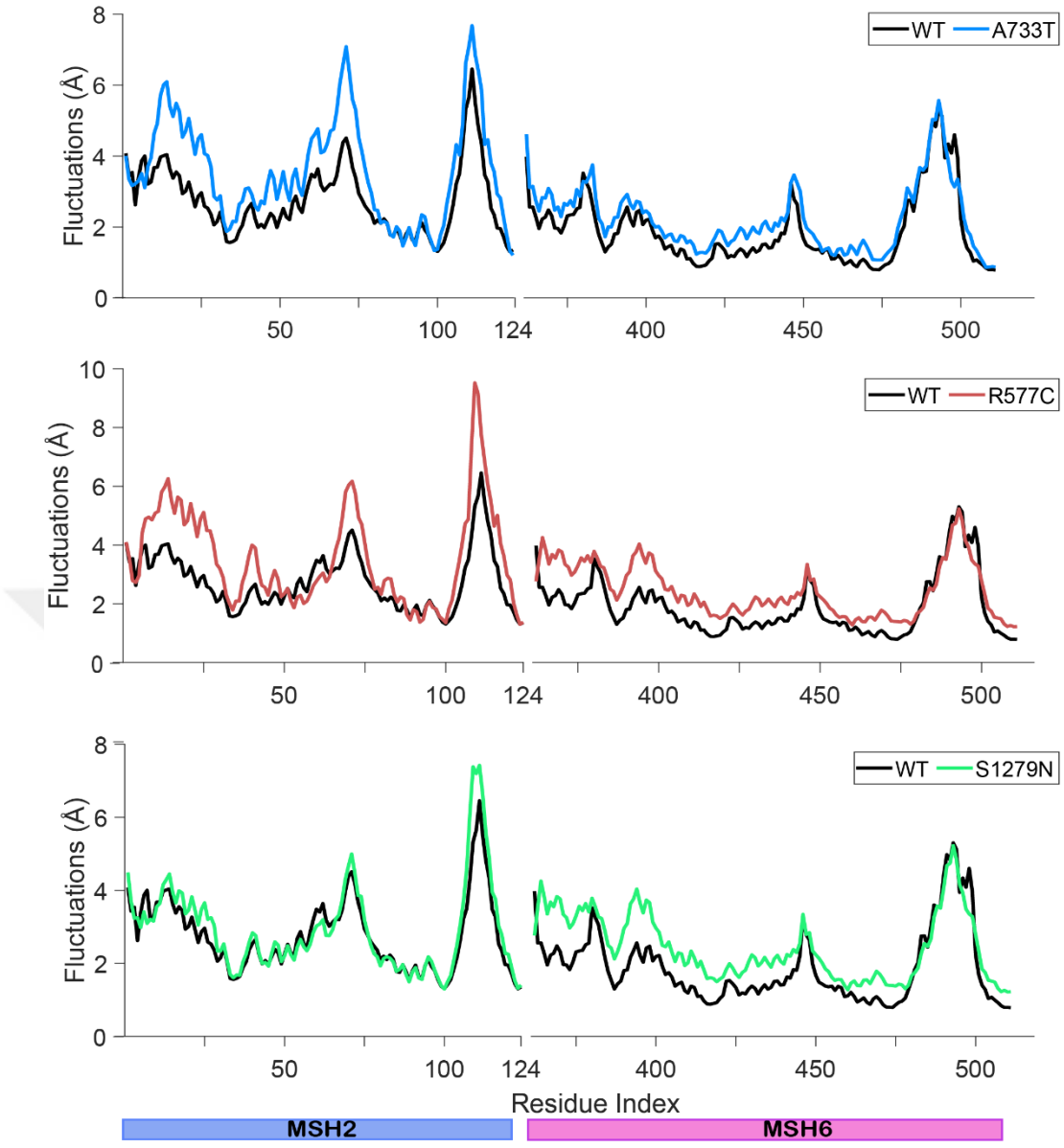


Figure 3.2.1. RMSF values of WT and mutant MutSα protein's mismatch binding domains are shown. WT, A733T, R577C and S1279N mutants are colored black, blue, red, and green, respectively. The blue bar indicates the location of the mismatch binding domain on MSH2, and the second bar indicates the MSH6's domain.

Overall, RMSF values of mismatch binding domains in MSH2 and MSH6 are higher in mutant proteins compared with the WT. Upon visual inspection of the graphs, it may be concluded that among the mutants the R577C mutation in MSH6 resulted in the largest increase in MSH2 fluctuations compared to WT. S1279N, on the other hand, resulted in the lowest change in in MSH2 fluctuations compared to WT.

The A733T and R577C mutations resulted in high fluctuations in the first 30 residues and residues between 112 and 114 in the MSH2 mismatch binding domain. Since the first 30 amino acids are located in the N-terminal and have loop structure, such larger fluctuations can be expected.

Each structure has two replicas, and all mutations give higher RMSF values than WT. Thus, mutations may cause more fluctuation of the MSH2 mismatch binding domain than WT. Based on our simulations especially R577C and A733T mutations appear to cause increased level of flexibility and stability of MSH2.

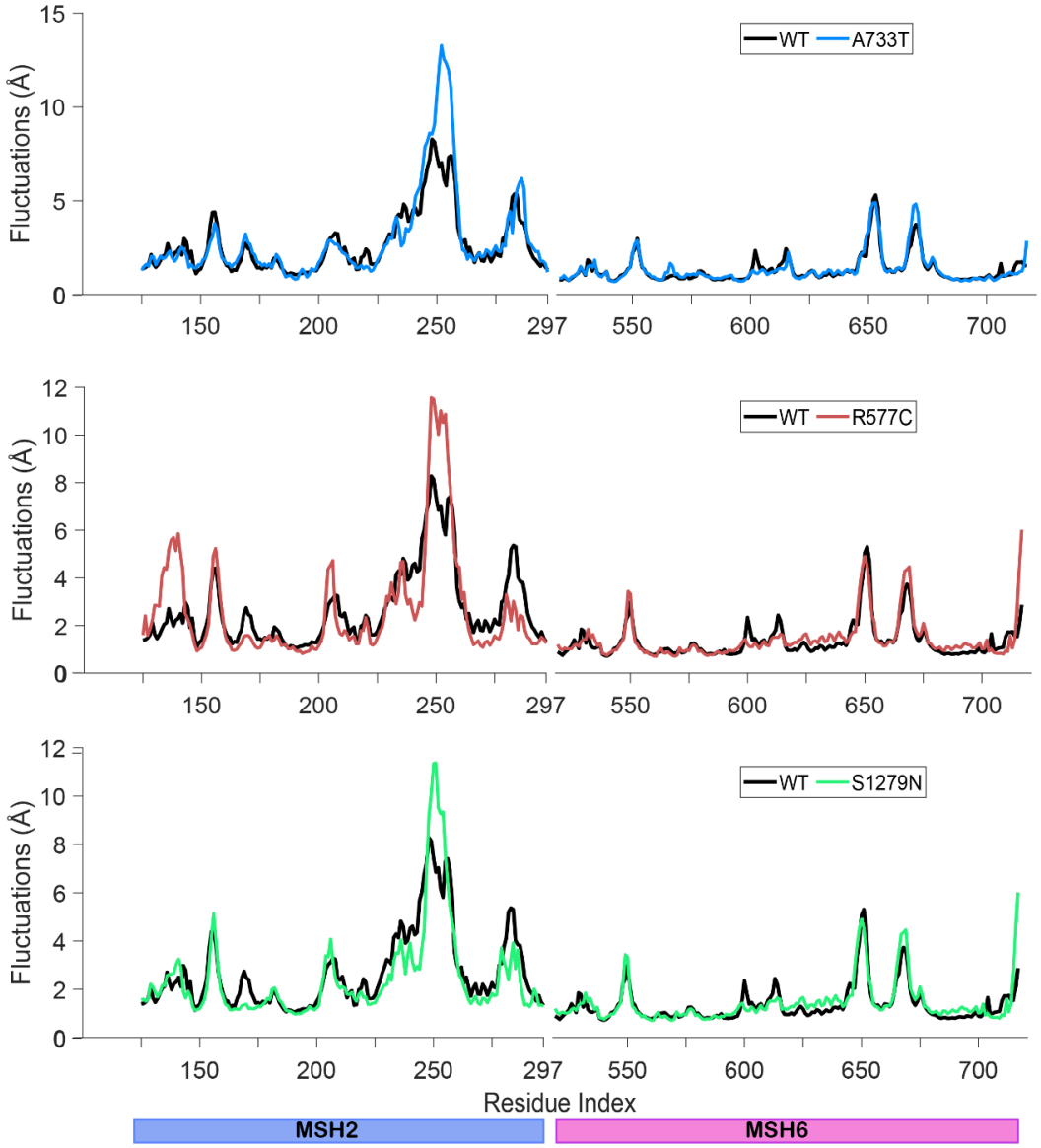


Figure 3.2.2. RMSF values of WT and mutant MutSα protein's connector domains are shown. WT, mutant A733T, mutant R577C and mutant S1279N are colored black, blue, red and green, respectively. The blue bar indicates the connector domain on MSH2, while the purple bar (after the line) showas MSH6 connectors**.**

Among the tested mutations, A733T mutation on MSH2 affected the mismatch binding domain structure and dynamics in MSH6 the least. Interestingly, fluctuations in the C-Terminal region of the domain were reduced compared to WT. It was observed that mutations on MSH6 did not affect the fluctuation in the domain as much as the MSH2 mismatch binding domain.

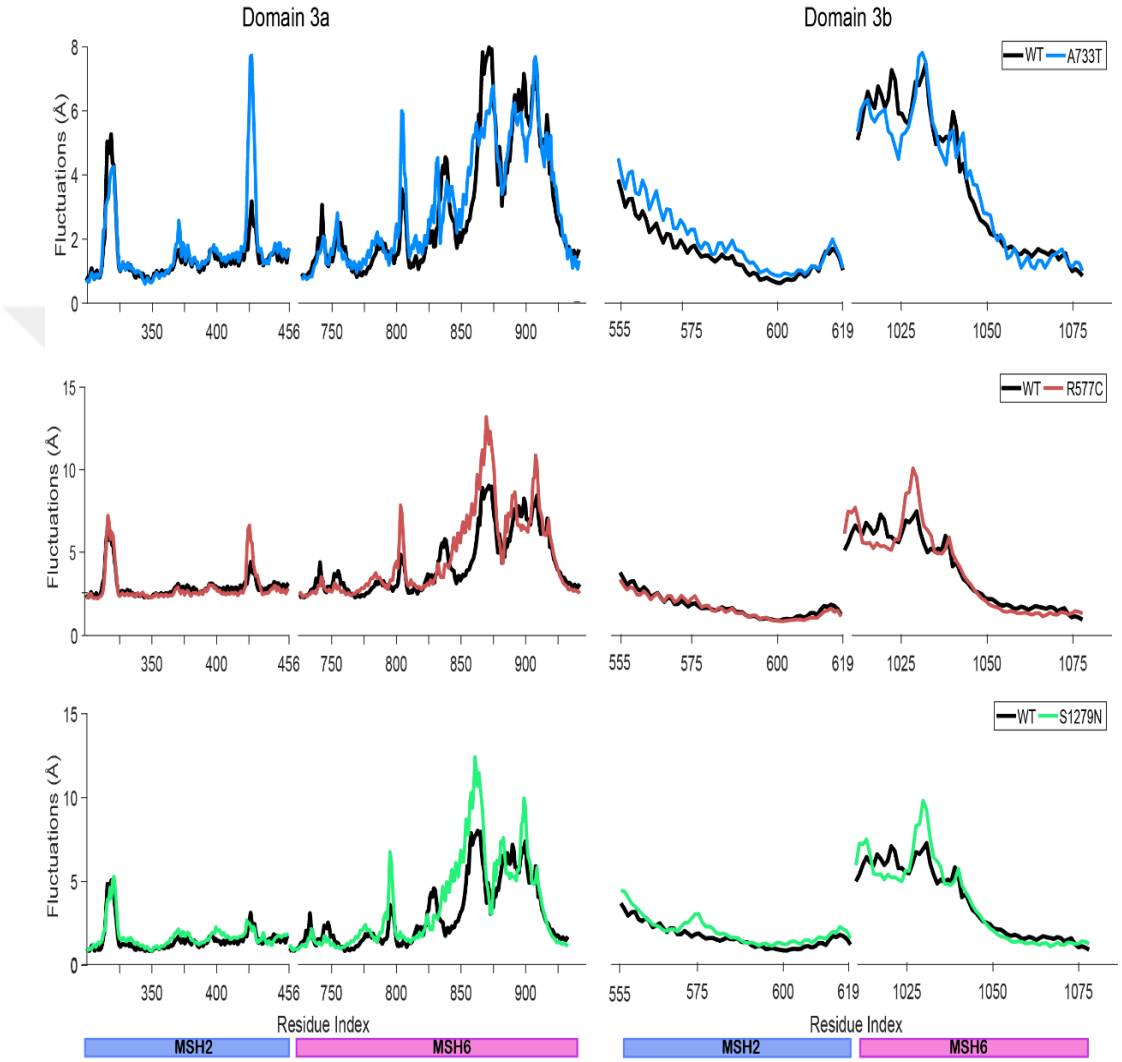


Figure 3.2.3. Average RMSF values of their two replicas of MutSα protein's long α-helical lever domains are shown. WT, A733T, R577C, and S1279N mutants are colored black, blue, red and green, respectively. Doman consists of two sequences for each MSH proteins. The blue bar indicates the Domain 3a the lever domain on MSH2, while the purple bar represents MSH6 levers. Domain 3b part of the graphics also consists of MSH2 and MSH6 parts**.**

RMSF values of the second domain, the connector domain, are represented in Figure 3.2.2. According to Figure 3.2.2, it is seen that mutations do not cause a considerable overall change in fluctuation of the connector domain. The effect of mutations is more focused. The residue between 245 to 263 in the loop structure fluctuates more in the

mutant structures. Although the R577C mutation is located in the connector domain of MSH6, it did not cause a considerable change in the overall fluctuation in MSH connector domains compared to WT.

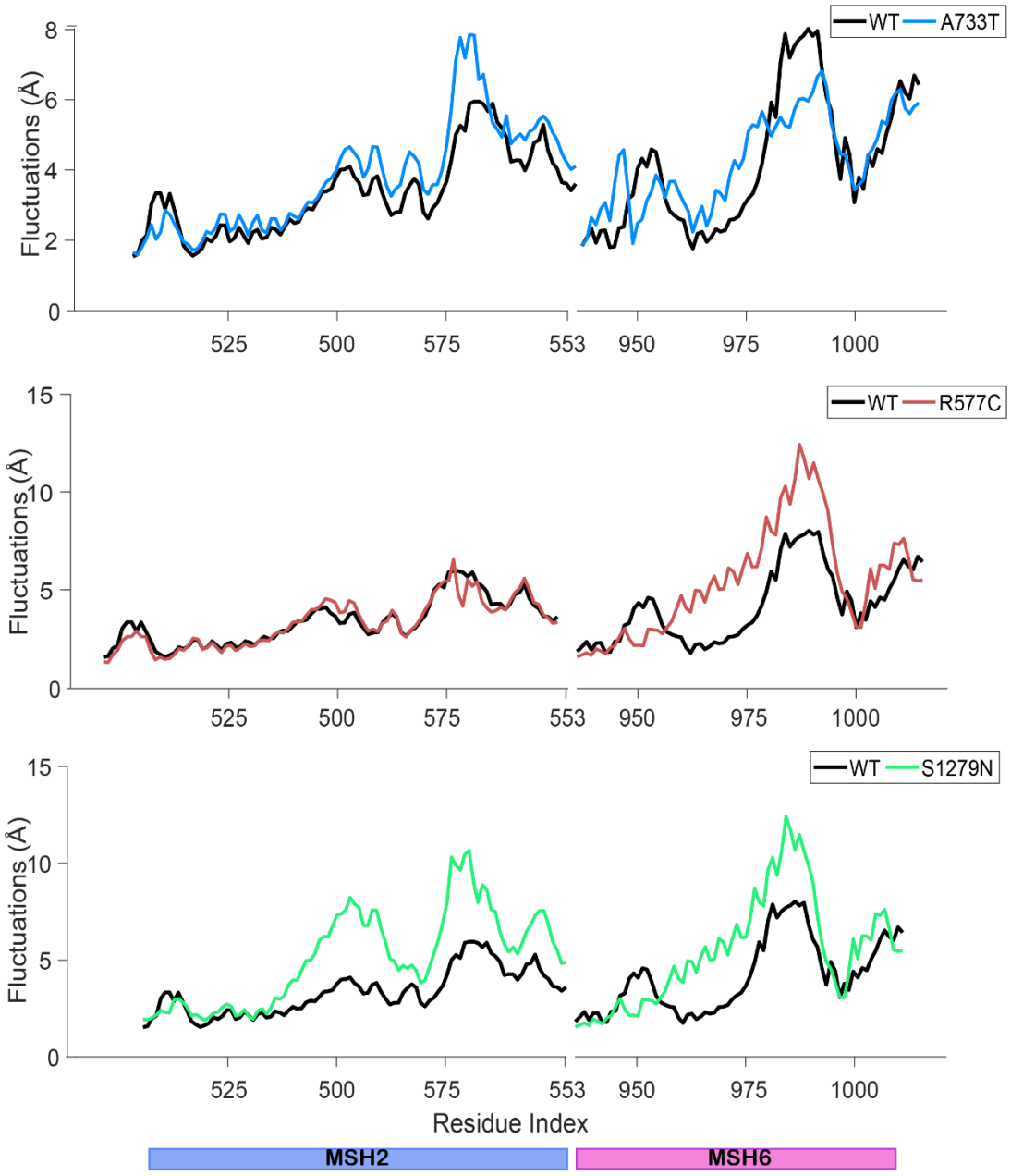


Figure 3.2.4. RMSF values of WT and mutant MutSα protein's clamp domains are shown. WT, A733T, R577C, and S1279N mutants are colored black, blue, red and green, respectively. The blue bar indicates the clamp domain on MSH2, while the purple bar showas the MSH6 clamp domain.

The lever domain consists of two sequence ranges (300 to 456 and 554 to 619 in MSH2; 718 to 934 and 1009 to 1075 in MSH6), and the fluctuations in the domains are shown in Figure 3.2.3. The R577C and S1279N mutations on MSH6 did not result

in extensive changes in the fluctuation profile in the lever domains of either MSH protein. Most of the fluctuations profiles of the mutant proteins followed a very similar fluctuation pattern to the WT protein, while local deviations from WT were observed. The region where mutant MSH6s fluctuated more in the lever domain (Figure 3.2.3, Domain 3a) is a long α-helix structure. Both mutations cause the same region of the level domain to fluctuate more on average than WT. The A733T mutation caused comparable average fluctuation magnitudes with the WT MSH2 lever domain, but lower average fluctuations according to WT MSH6 domains.

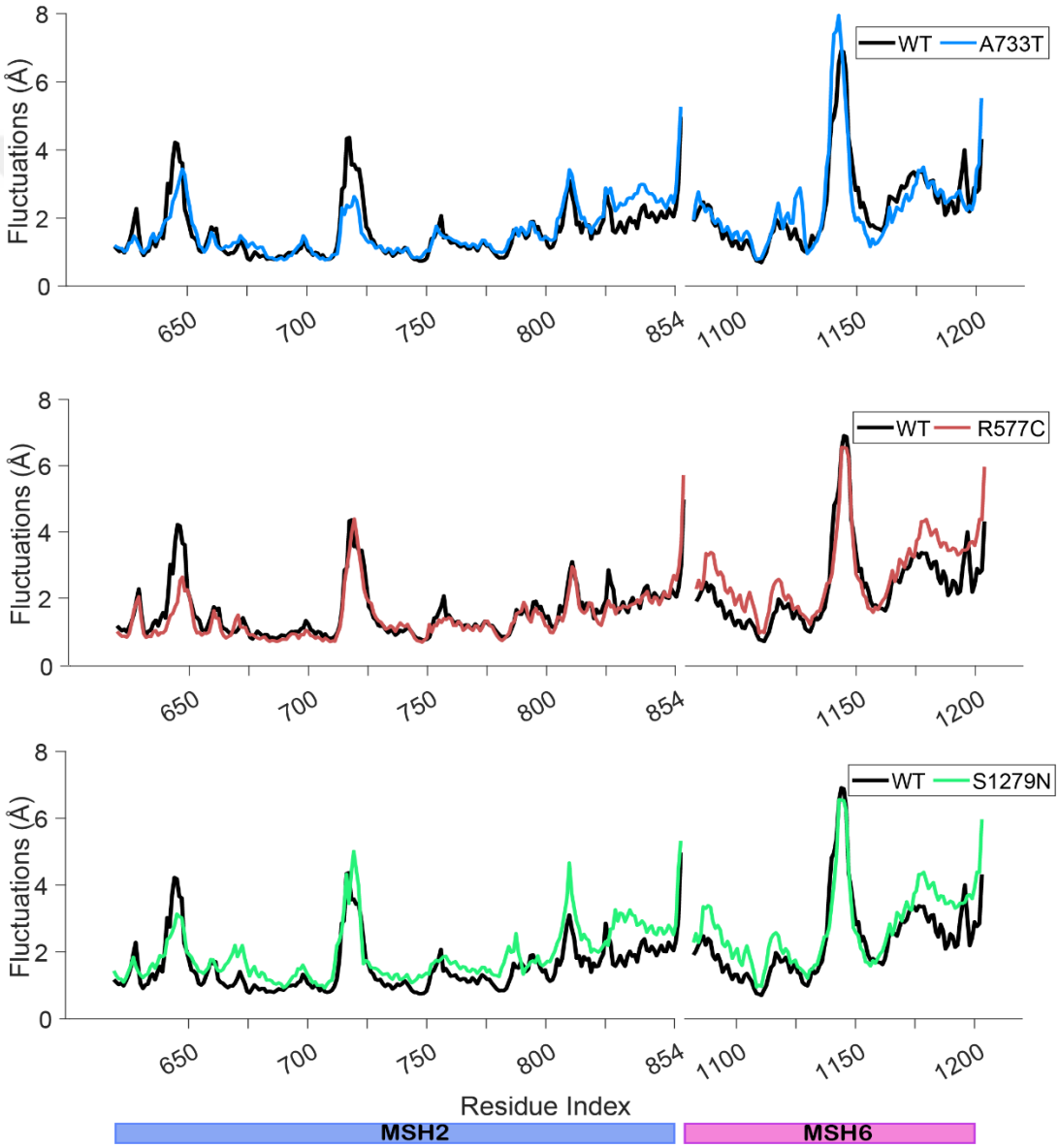


Figure 3.2.5. RMSF values of WT and mutant MutSα protein's ATPase domains are shown. WT, A733T, R577C, and S1279N mutants are colored black, blue, red and green, respectively. The blue bar indicates the ATPase domain on MSH2, while the purple bar showas the MSH6 ATPase domain.

A733T mutation causes especially increased fluctuation in the MSH2 clamp alpha helixes(Figure 3.2.4). In MSH6, on the other hand, the mutation appears to cause less fluctuation compared to WT. The R577C mutation does not appear to considerably affect MSH2 clamp domain fluctuation, but has a larger effect on fluctuations in MSH6. The S1279N mutation appears to cause the highest change in fluctuation of the clamp domain among all mutants.

RMSF values of the ATPase domain for the WT and mutants are shown in Figure 3.2.5. The A733T mutation is located in the MSH2 ATPase domain, and there was no considerable fluctuation difference in both the MSH2 domain and the MSH6 ATPase domain. The other two mutations show mostly similar fluctuations with the WT. However, the R577C and S1279N mutations showed considerable higher fluctuation in the C-Terminal of MSH6. WT and mutant MSH2 and MSH6 domain conformation at the end of MD simulations are shown in Figure 3.2.6.

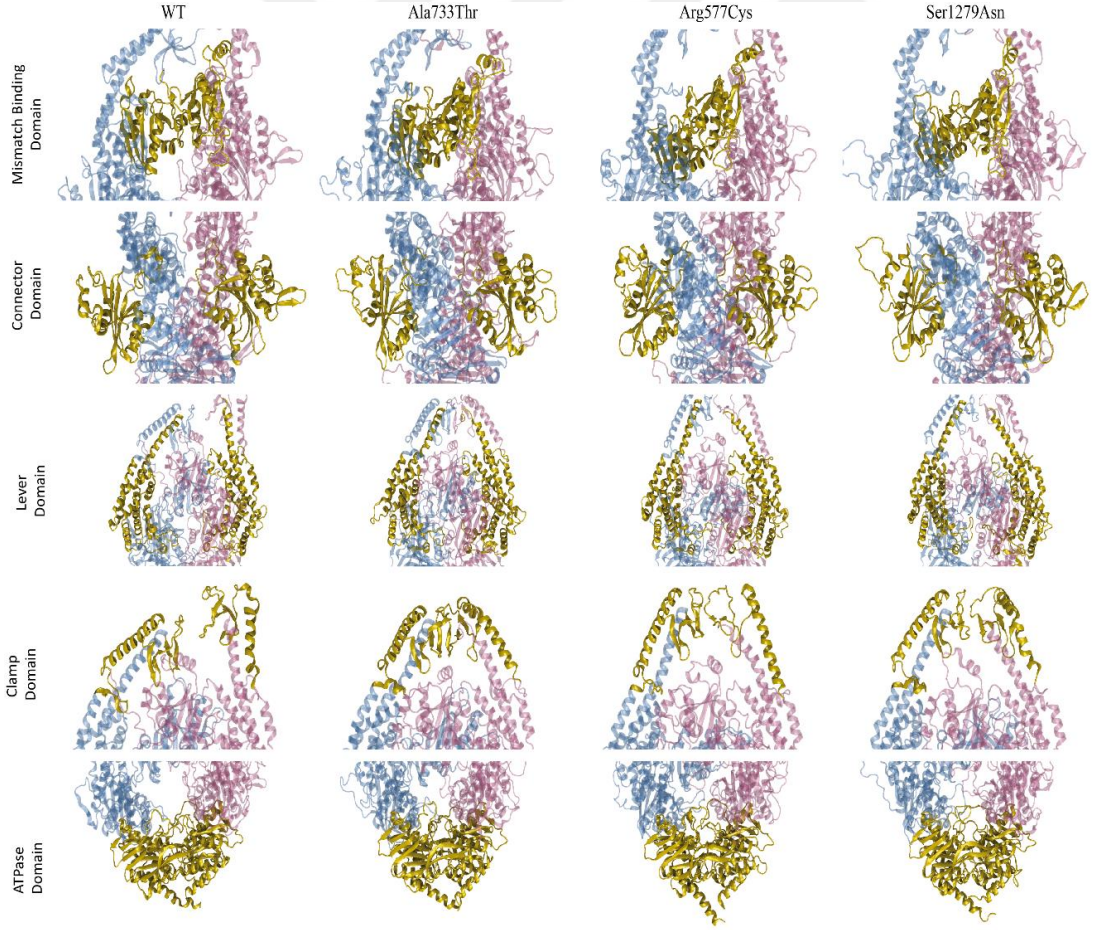


Figure 3.2.6. MutSα protein's WT, A733T, R577C, and S1279N mutants specific residue interval structures at the end of the MD simulations according to RMSF differences.

3.3 Heat Capacity Analysis

The heat capacity (Cp) calculation is related to the enthalpy values of the structures sampled in the conformation space. Changes in heat capacity provide information about the thermodynamic properties of proteins (Cooper, 2005; Prabhu and Sharp, 2005). The heat capacity changes of two replicas of all structures are given in Figure 3.3.1. Since the starting conformations are different in each production MD simulation, it is expected that there will be differences between replicas at the beginning of the simulations since they sample different regions at the beginning. However, it can be thought that these samples will converge towards each other when the MD simulation are performed long enough. For all systems the heat capacities obtained from the replicas started to converge towards each other.

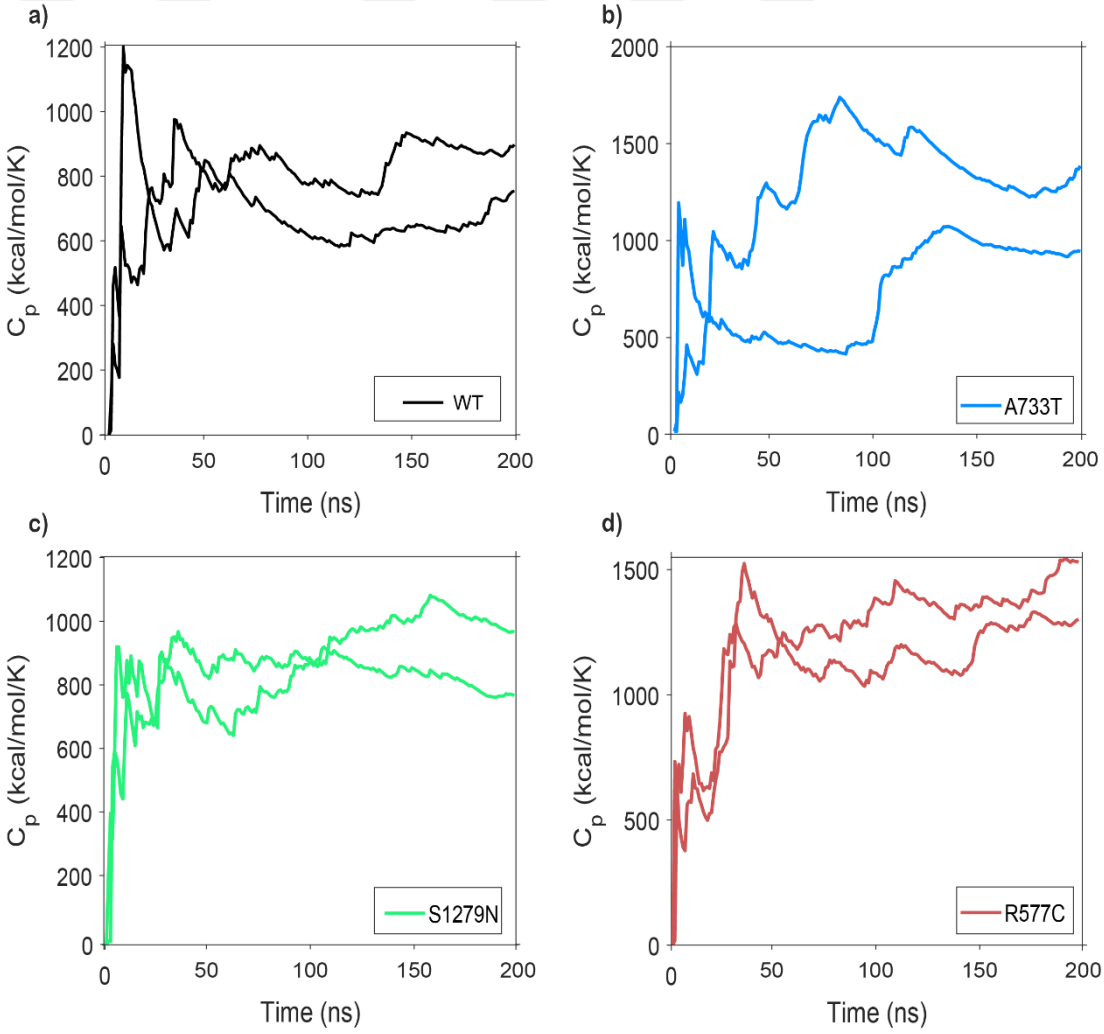


Figure 3.3.1. Change of heat capacities for WT and mutant MutSα proteins for two replicas. a) WT, b) mutant A733T, c) mutant R577C and d) mutant S1279N are shown.

The average heat capacity curves for the WT and mutants are given in Figure 3.3.2. Compared to the WT structure, the Cp values of the mutant proteins were higher. Since higher heat capacity may be associated with protein unfolding (Cooper, 2005), the R577C mutation may be the most susceptible mutation to protein unfolding among the tested mutations. The highest Cp value was calculated as 1386.53 kcal/mol.K for the R577C mutant. The other two mutations gave similar but higher Cp values compared to WT. It can be said that all mutations make the protein structure more susceptible to unfolding compared to WT. However, the heat capacity difference with WT was low for S1279N.

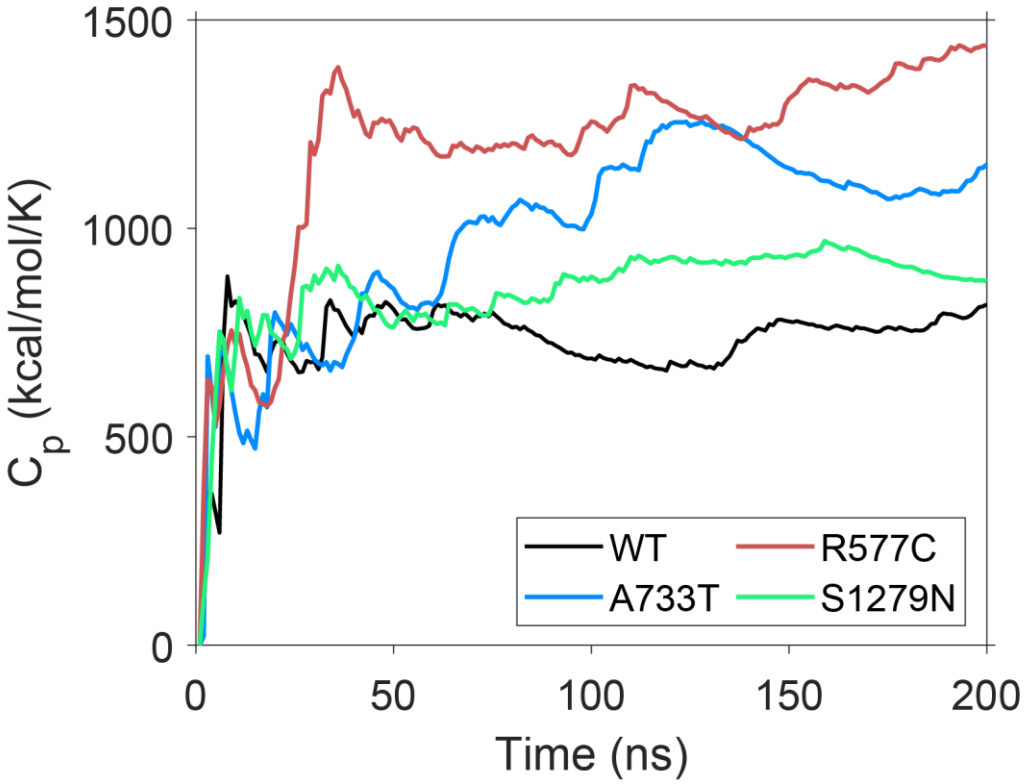


Figure 3.3.2. WT and mutant MutSα protein's average heat capacity results. WT, mutant A733T, mutant R577C and mutant S1279N are colored as black, blue, red and green, respectively.

3.4 Interaction Analysis

The ATPase domains are the domains though which MSH2 and MSH6 proteins interact with each other. Tested mutations are either on and close to the ATPase domain. Interaction analysis between ATPase domains was performed to observe how the mutations would affect their interactions. For the H bond interaction analysis, the

cut-off was taken as 3.5 Å, and interactions observed with a frequency of 15% or more were taken into consideration. 11 hydrogen bonds were observed in WT protein, nine in A733T mutant protein, 15 in R577C mutant and 12 in S1279N mutant. While the least number hydrogen bonds were observed for A733T, the highest number was observed in S1279N. The frequency of hydrogen bonds throughout the trajectory decreased in mutants compared to WT.

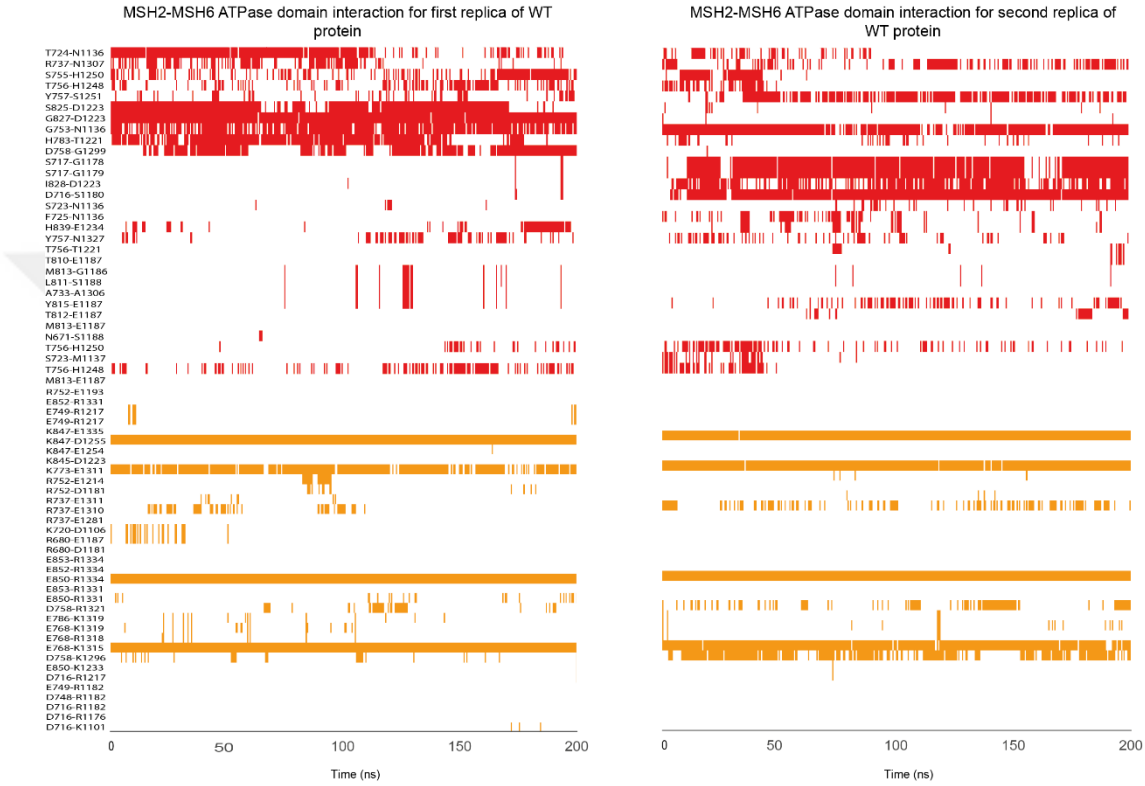


Figure 3.4.1. Hydrogen bond (red) and salt bridge (yellow) interactions between ATPase domains of WT MSH2 and MSH6 proteins.

For the salt bridge interaction analysis in the ATPase domain, the cut-off was taken as 6 Å, and salt bridges that could be observed with a frequency of 15% or more throughout the trajectory were included in the analysis. Four of seven salt bridges observed in the WT protein were observed with high frequency. Seven salt bridges were observed in the A733T mutant protein, among which 3 were observed with high frequency, while two were observed with moderate frequency. For the R577C mutant, two out of seven salt bridges were observed with high frequency. In the S1279N mutant protein, there are nine salt bridge interactions.

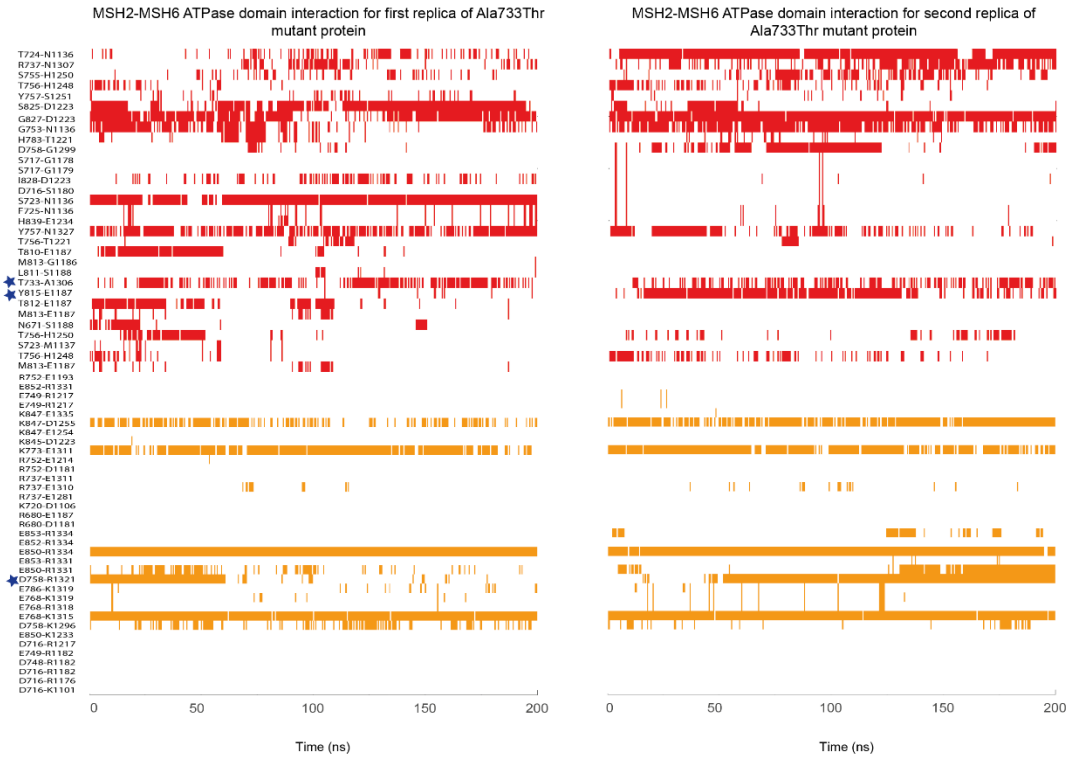


Figure 3.4.2. Hydrogen bond (red) and salt bridge (yellow) interactions between ATPase domains of MSH2 and MSH6 proteins with A733T mutation.

Interactions in the WT ATPase domain along the trajectory lengths are shown in Figure 3.4.1. H bonds that were observed with a frequency higher than 15% at any point along the trajectory are Asn671-Gly1218, Thr724-Asn1136, Arg737-Asn1307, Ser755-Hsd1250, Thr756-Hsd1248, Tyr757-Ser1251, Ser825-Asp1223, Gly827-Asp1223, Gly753-Asn1136, Hsd783-Thr1221, and Asp758-Gly1299. In addition, salt bridges observed with a frequency of more than 15% were observed as Asp758-Lys1296, Glu768-Lys1315, Glu850-Arg1334, Lys773-Glu1311, Lys847-Asp1255, and Arg737-Glu1311.

The interactions observed in the A733T mutant protein are given in Figure 3.4.2. Unlike WT, only two hydrogen bonds were frequently observed. One of them is the Thr733-Ala1306 hydrogen bond. Another observed hydrogen bond was between the Tyr815-Glu1187 residues. Especially in the second replica of MD simulation, this hydrogen bond was observed with high frequency. The Asp758-Arg1321 salt bridge observed in the mutant protein was not observed in the WT protein.

The interaction graph over MD simulation time for the R577C mutant is provided in Figure 3.4.3. Both new interaction and also losses of interaction in the ATPase domain compared to WT were observed. Tyr757-Ser1251 interaction was rarely observed in

the mutant protein. Furthermore, it hydrogen bonds Ser717-Gly1178, and Asp716-Ser1180, which were observed mainly in the second replica of the WT protein, were absent in the mutant protein. Hydrogen bonds Phe725-Asn1136, Tyr757-Asn1327 and Thr756-His1248 that were not observed for WT, were observed to form. Glu853-Arg1334 and Asp758-Arg1321 salt bridges not present in WT were formed in this mutant, while the Glu850-Arg1334 interaction present in mutant was observed with a considerably lower frequency.

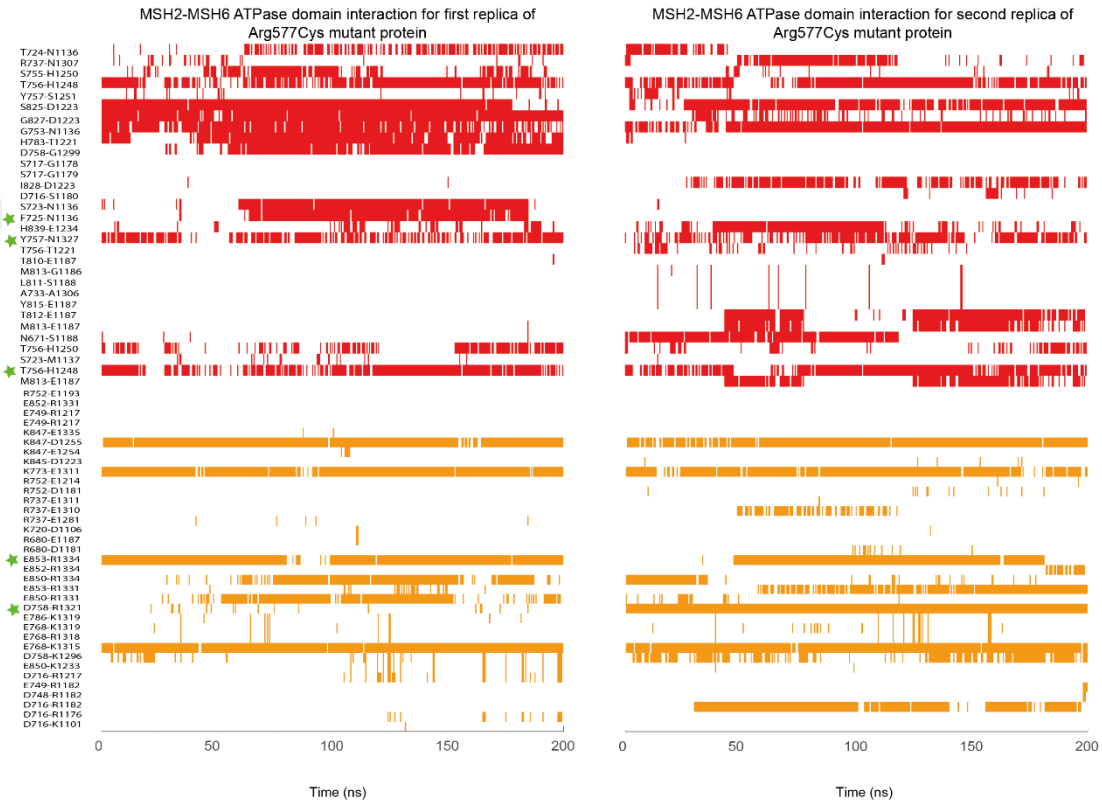


Figure 3.4.3. Hydrogen bond (red) and salt bridge (yellow) interactions between ATPase domains of MSH2 and MSH6 proteins with R577C mutation.

The interaction graph of the S1279N mutant is given in Figure 3.4.4. Gly827-Asp1223 interaction observed in WT appears less frequently in the mutant protein. It was also observed that the mutant protein's hydrogen bonds of Ser717-Gly1178 and Asp716-Ser1180 were absent in the WT. The hydrogen bond interaction in the mutant protein was higher in number compared to WT. The hydrogen bonds formed with the highest frequency were Ile828-Asp1223, His839-Glu1234 and Thr756-His1248. These interactions have not been observed in the WT protein. Futhermore, the observed number of salt bridges were more than those observed in the WT. Salt bridges, which

were not observed in WT but were observed with high frequency in S1279N were Glu852-Arg1334, Glu850-Arg1331, and Asp758-Arg1321.

Figure 3.4.4. Hydrogen bond (red) and salt bridge (yellow) interactions between ATPase domains of MSH2 and MSH6 proteins with S1279N mutation.

3.5 Principle Component Analysis (PCA)

In this study Principle Component Analysis was performed and the two principal movements in the MutSα protein are shown in Figures 3.5.1 and 3.5.2. The most dominant movement, i.e. PC1, for WT and mutant proteins occurred mostly in the MSH2 protein. The motion vector is inward/outward in the WT and A733T mutant proteins (Figure 3.5.1 A, B). In the R577C mutant, the PC1 in the lever and connector domains of MSH2 was observed inward/outwards, while the mobility in the clamp region was observed in the outward/inward direction to the protein (Figure 3.5.1, C). In the S1279N mutant, the most dominant movement was observed in the clamp domains of MSH2 and MSH6 proteins (Figure 3.5.1, D).

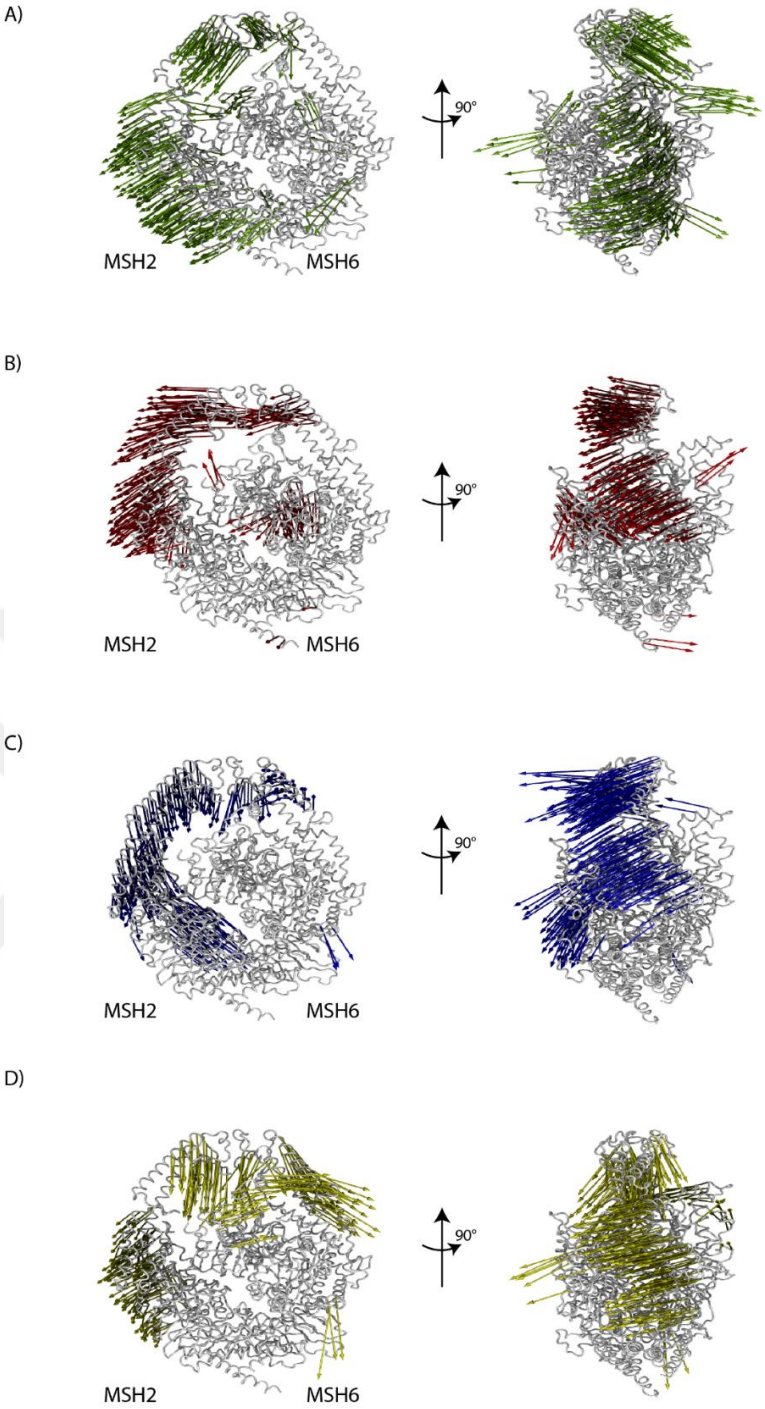


Figure 3.5.1. PC1 of WT and mutant MutSα protein obtained by PCA analysis. A) Frontal and side view of the most important movements in WT MutSα B) Front and side views of the most important movements in the A733T mutant C) R577C mutant D) S1279N mutant.

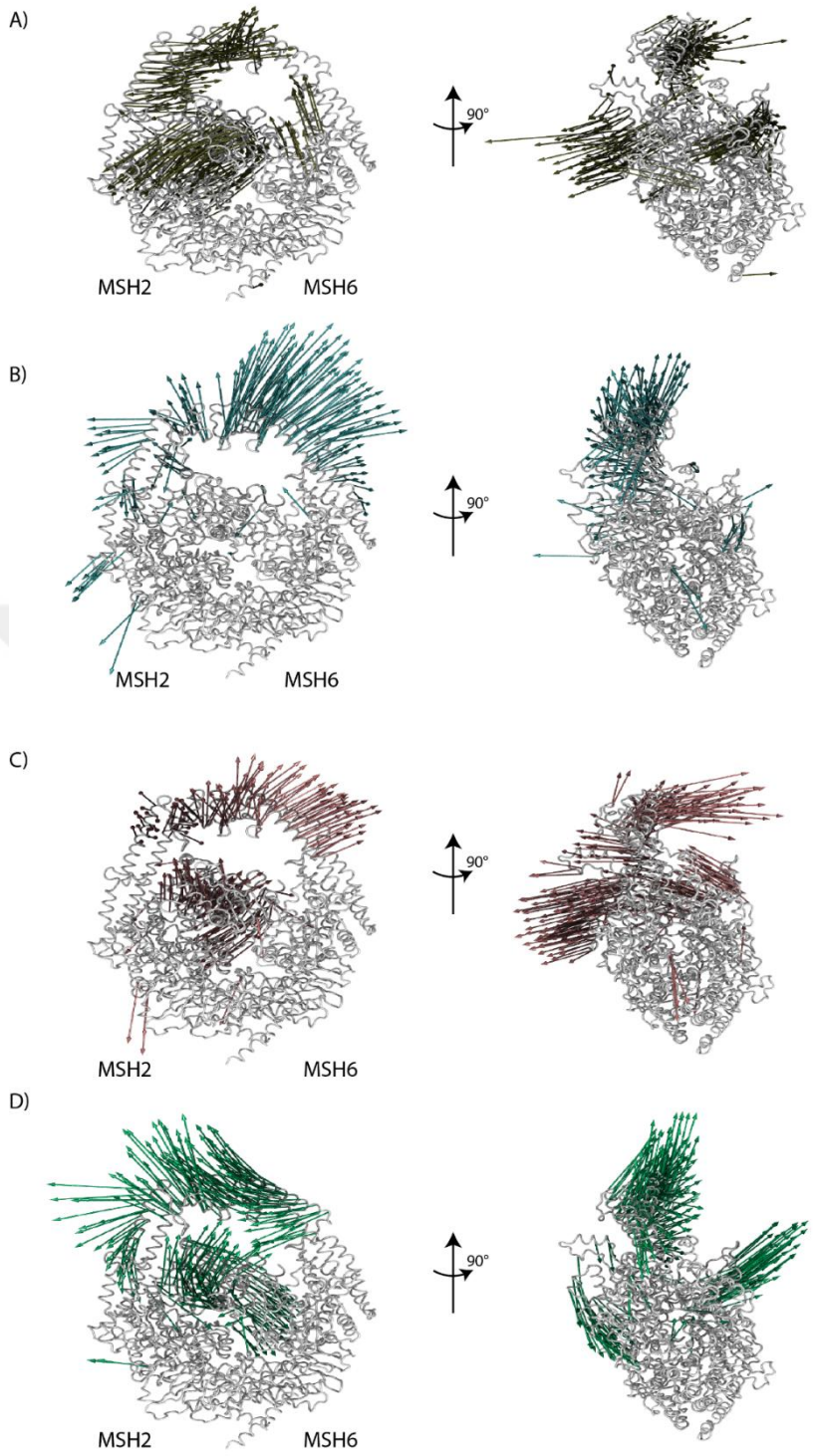


Figure 3.5.2. PC2 of WT and mutant MutSα protein obtained by PCA analysis. A) Frontal and side view of the most important movements in WT MutSα B) A733T mutant C) R577C mutant D) S1279N mutant.

The second most dominant movements, i.e. PC2, in MutSα are shown in Figure 3.5.2. Most of the movement of PC2 was observed in MSH6. In WT, the movements were observed in the mismatch binding, connector and clamp domains in MSH2 (Figure 3.5.2, A). In the A733T mutant, the MSH6 clamp domain was the region where the

PC2 movements were most common, and it appears to be in the inward/outward directions of the clamp domain (Figure 3.5.2, B). In the R577C mutant protein, PC2 movements that are effective in the inward/outward movement of the MSH6 clamp domain were observed. In addition, it has been observed that the PC2 is also effective in the MSH2 mismatch binding domain (Figure 3.5.2, C). The PC2 in the S1279N MutSα protein mutant is shown in the Figure 3.5.2 D panel, where PC2 movemetns are large in MSH2 and MSH6's clamp and mismatch binding domains. While the movement in the clamp domain shows the inward/outward movement of the region, the mismatch binding domain shows the movement outward/inwward of the protein.

The locations where both replicas of the proteins were sampled in space during the simulation with PCA analysis are shown as the heat map (Figure 3.5.3). The selected space was divided into 50x50 grids. Conformations were clustered into grids. Number of conformations were normalized and distributions were obtained. The purple star is the region where the crystal structure is located.

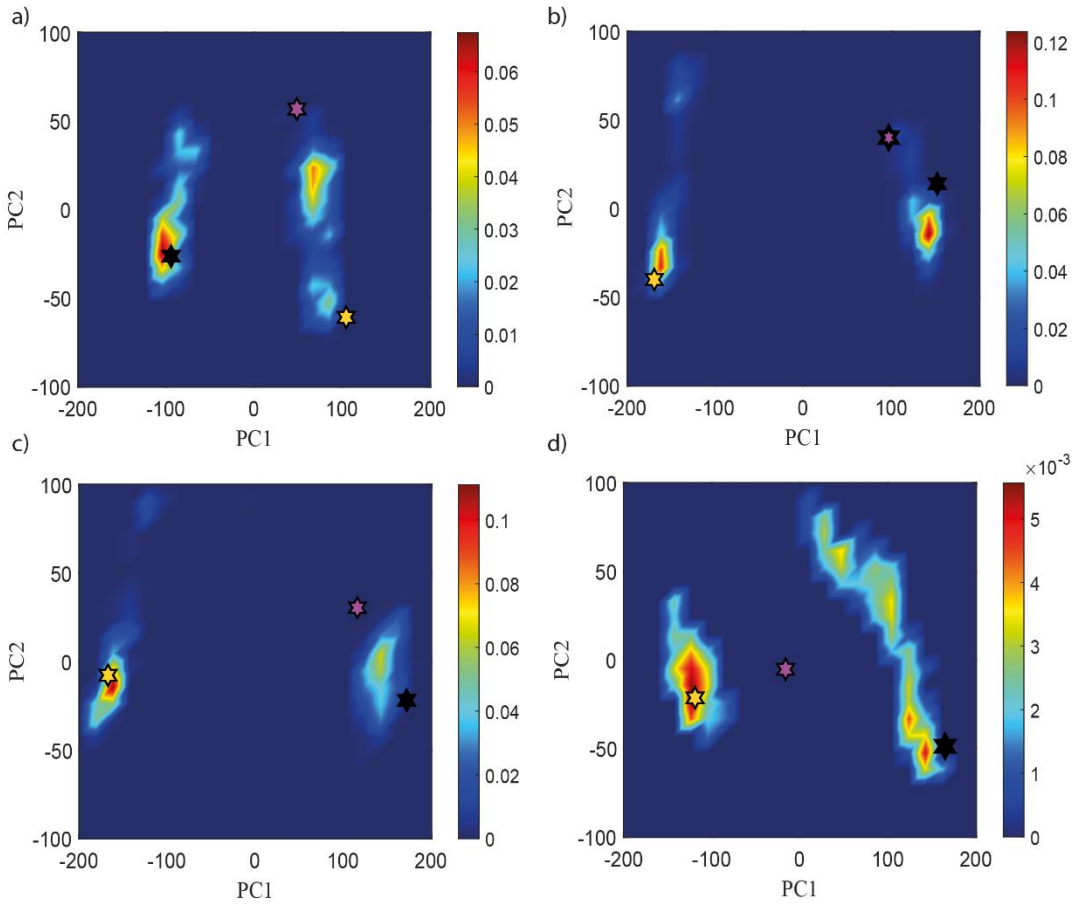


Figure 3.5.3. During the MD simulations, the regions sampled by the proteins in space are shown. The magenta star shows the initial positions of the crystal structure, yellow and black star replicas in space. a) WT, b)A733T, c)R577C and d)S1279N.

4. CONCLUSION

Colorectal cancers are among the most common types of cancer. When early diagnosis of colorectal cancer is detected, the 5-year survival rate is around 90%. Some types of colon cancers are known to be inherited. As a result of hereditary transfer of pathogenic variants in cancer susceptibility genes, it causes the emergence of similar or related cancer types in next generations. Many of the cancer-associated genes control DNA damage repair or cell cycle. When the MutSα complex formed by MSH2 and MSH6 proteins encounters damaged DNA, it recognizes the damaged area on the DNA and binds to defect and initiates the MMR pathway. The MutSα complex is formed by the combination of MSH2 and MSH6 monomers, which consist of mismatch binding domain, connector domain, lever domain, clamp domain and ATPase domains. Domains that have various roles in DNA binding, ATP hydrolysis, acting on DNA, interacting with other proteins and initiating repair are very important for the correct functioning of the MutSα complex.

Various variations were idnetified through genetic screening tests of people having cancer family history performed by Levent Doğanay and his research group within the scope of the TUBITAK 1003 project that also funded this thesis work. In this thesis, and the effects of these variations on protein structure and dynamics were investigated. Point mutationsA733T, R577C, and S1279N were introduced and MD simulations were performed. Changes in protein structure and dynamics and also thermodynamics with respect to WT were performed. For this purpose, RMSD, RMSF and heat capacity analyzes of the protein were performed based on MD simulations trajectories.

Similar RMSD values were observed during the MD simulation in both replicas of the WT structure. Also the RMSD values sampled in replicas of the A733T mutant's MD simulation gave comparable values. The difference between the RMSD values of the R577C mutant was comparably higher between the replicas. For the S1279N mutant the RMSD values diverged between the replicas after 150 ns. This difference between replicas may suggest that S1279N might be sampling a larger conformational

space. The average RMSD values of WT protein were generally lower than the mutants. Therefore, it can be said that mutations cause changes in protein conformation and dynamics. By extending the MD simulation time, a better understanding of the extent of the conformational changes would be probably observed.

Since the MutSα protein is such a large protein, it is difficult to understand the effect of mutations on the whole protein and the stability of conformations. Since RMSF analysis is a way of effectively quantifying the fluctuations of each residue in the protein structure throughout the simulation length, it provides insight regarding the rigidity of the domains in the proteins. The changes in the domains of the fluctuations during the simulation were examined separately. Figure 3.2.1 shows the changes in fluctuations caused by mutations in the mismatch binding domain. In particular, the fluctuations in MSH2 increased compared to WT. Mutations may have caused conformational deviation in MSH2 compared to WT. In MSH6, on the other hand, it was observed that the mutations causing the highest changes in RMSF were R577C and S1279N. Changes in the connector domain were more effective in the domain in MSH2. Fluctuations in the lever domain appear to be much more affected by the A733T mutation than R577C and S1279N. The mutation decreased fluctuation in the highly mobile regions of the WT protein. The other two mutations showed similar and high fluctuations between residues 825-875 of MSH6.

Previous studies have shown that the clamp and lever domains undergo large conformational changes in DNA-free MutSα, and these domains move in the opposite direction, causing the protein to open at the DNA binding site (Mukherjee and Feig, 2009). In this study, mutations in the structure caused the clamp region to gain a more emphasized degree of movement than WT (Figure 3.2.4). In the MD simulations the clamp regions were closed in the WT and mutant DNA-free structure. Different results may have been obtained from the above mentioned work in the literature due to the fact that the missing structures were completed in the mentioned study or the simulated conditions were slightly different. The mutations seem to have caused much more activity in this region. Although clamp domains also fluctuated largely in the WT structure, it can be said that mutations increase the fluctuation here even further. Thus, rigidity appears to be reduced in this domain. Although mutations are

at or close to the ATPase region of the protein, they may also allosterically affect clamp domain.

Two of the mutations (A733T and S1279N) occur in the ATPase domain of both MSH proteins. No considerable change were observed in MSH2. It is seen that there are fluctuations in the ATPase domains where the R577C and S1279N mutations occur. Since the ATPase domain is located at the C-Terminal, an increase in fluctuation in the terminal residues can be expected, but it can also be thought that the higher fluctuation compared to WT and A733T may be also partly caused by mutations.

Heat capacity values averaged over replicas showed that the Cp values of all mutations, especially the R577C mutation, were higher than WT. The increase in Cp valuemay be related to a higher susceptibility to unfolding. In particular, MD simulations showed that the R577C mutation caused more opening in the structure of the protein. This may be potentially the partial cause in the differences on binging or recognizing mechanism to defective DNA.

As a result of the interaction analysis, the composition and frequency of salt bridges and hydrogen bonds in the structure were affected. The A733T mutation caused additional hydrogen bond formation at the location where the mutation occurred. Single point mutations both prevented some of the interactions and also resulted in new additional interactions in the ATPase region.

The two most dominant movements in the protein were determined as the principal vectors by the PCA analysis. The first two PCs mainly affected different regions of the proteins. Heat maps were generated to investigate the conformational space the proteins sampled during MD simulations. The sampled regions varied between replicates of each system. While PC1 mainly affected MSH2, the PC2 mostly affected MSH6 clamp domain. For the WT in the absence of DNA opening from the clamp to some degree was observed, potentially facilitating DNA scanning and binding to mismatched bases.This thesis showed via MD simulations that A733T, R577C, and S1279N mutation affect MutSα structure and dynamics. Especially, the effect on the clamp domain structure and dynamics could potentially affect the protein’s DNA binding effectivity. Thus, suggesting an atomic insight to these mutations relevance to colon cancer.

REFERENCES

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(24), 13629-13634.

Alani, E., Lee, J. Y., Schofield, M. J., Kijas, A. W., Hsieh, P. Yang, W. 2003. "Crystal Structure and Biochemical Analysis of the MutS.ADP.Beryllium Fluoride Complex Suggests a Conserved Mechanism for ATP Interactions in Mismatch Repair * 210", Journal of Biological Chemistry, 278(18), 16088-16094.

Allen, D. J., Makhov, A., Grilley, M., Taylor, J., Thresher, R., Modrich, P. Griffith, J. D. 1997. "MutS mediates heteroduplex loop formation by a translocation mechanism", The EMBO journal, 16(14), 4467-4476.

Aparicio, T., Baer, R. Gautier, J. 2014. "DNA double-strand break repair pathway choice and cancer", DNA Repair (Amst), 19, 169-175.

Bateman, A. C. 2021. "DNA mismatch repair proteins: scientific update and practical guide", Journal of Clinical Pathology, 74(4), 264-268.

Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M. MacKerell, A. D. 2012. "Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain $\chi 1$ and $\chi 2$ Dihedral Angles", Journal of Chemical Theory and Computation, 8(9), 3257-3273.

Bischof, H., Rehberg, M., Stryeck, S., Artinger, K., Eroglu, E., Waldeck-Weiermair, M., Gottschalk, B., Rost, R., Deak, A. T., Niedrist, T., Vujic, N., Lindermuth, H., Prassl, R., Pelzmann, B., Groschner, K., Kratky, D., Eller, K., Rosenkranz, A. R., Madl, T., Plesnila, N., Graier, W. F. Malli, R. 2017. "Novel genetically encoded fluorescent probes enable real-time detection of potassium in vitro and in vivo", Nature Communications, 8(1), 1422.

Blackwell, L. J., Bjornson, K. P. Modrich, P. 1998. "DNA-dependent activation of the hMutSαlpha ATPase", J Biol Chem, 273(48), 32049-32054.

Blackwell, L. J., Martik, D., Bjornson, K. P., Bjornson, E. S. Modrich, P. 1998. "Nucleotide-promoted release of hMutSαlpha from heteroduplex DNA is consistent with an ATP-dependent translocation mechanism", J Biol Chem, 273(48), 32055-32062.

Bowers, J., Sokolsky, T., Quach, T. Alani, E. 1999. "A mutation in the MSH6 subunit of the Saccharomyces cerevisiae MSH2-MSH6 complex disrupts mismatch recognition", J Biol Chem, 274(23), 16115-16125.

Cadet, J., Berger, M., Douki, T. Ravanat, J. L. 1997. "Oxidative damage to DNA: formation, measurement, and biological significance", Rev Physiol Biochem Pharmacol, 131, 1-87.

Carethers, J. M. Stoffel, E. M. 2015. "Lynch syndrome and Lynch syndrome mimics: The growing complex landscape of hereditary colon cancer", World J Gastroenterol, 21(31), 9253-9261.

Cooper, A. 2005. "Heat capacity effects in protein folding and ligand binding: a re-evaluation of the role of water in biomolecular thermodynamics", Biophysical Chemistry, 115(2), 89-97.

Culligan, K. M., Meyer-Gauen, G., Lyons-Weiler, J. Hays, J. B. 2000. "Evolutionary origin, diversification and specialization of eukaryotic MutS homolog mismatch repair proteins", Nucleic Acids Res, 28(2), 463-471.

Das Gupta, R. Kolodner, R. D. 2000. "Novel dominant mutations in Saccharomyces cerevisiae MSH6", Nat Genet, 24(1), 53-56.

David, C. C. Jacobs, D. J. 2014. "Principal component analysis: a method for determining the essential dynamics of proteins", Methods Mol Biol, 1084, 193-226.

Debrunner, P. G. Frauenfelder, H. 1982. "Dynamics of Proteins", Annual Review of Physical Chemistry, 33(1), 283-299.

Dexheimer, T. S. 2013. "DNA Repair Pathways and Mechanisms". DNA Repair of Cancer Stem Cells. Mathews, L. A., Cabarcas, S. M.Hurt, E. M.: Dordrecht: Springer Netherlands.

Dizdaroglu, M. 2005. "Base-excision repair of oxidative DNA damage by DNA glycosylases", Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 591(1), 45-59.

Drotschmann, K., Clark, A. B., Tran, H. T., Resnick, M. A., Gordenin, D. A. Kunkel, T. A. 1999. "Mutator phenotypes of yeast strains heterozygous for mutations in the MSH2 gene", Proc Natl Acad Sci U S A, 96(6), 2970-2975.

Drummond, J. T., Li, G. M., Longley, M. J. Modrich, P. 1995. "Isolation of an hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells", Science, 268(5219), 1909-1912.

Dudáš, A. Chovanec, M. 2004. "DNA double-strand break repair by homologous recombination", Mutation Research/Reviews in Mutation Research, 566(2), 131-167.

Dufner, P., Marra, G., Räschle, M. Jiricny, J. 2000. "Mismatch recognition and DNA-dependent stimulation of the ATPase activity of hMutSαlpha is abolished by a single mutation in the hMSH6 subunit", J Biol Chem, 275(47), 36550-36555.

Edelbrock, M. A., Kaliyaperumal, S. Williams, K. J. 2013. "Structural, molecular and cellular functions of MSH2 and MSH6 during DNA mismatch repair, damage signaling and other noncanonical activities", Mutat Res, 743-744, 53-66.

Fishel, R. Kolodner, R. D. 1995. "Identification of mismatch repair genes and their role in the development of cancer", Current Opinion in Genetics & Development, 5(3), 382-395.

Fishel, R. Lee, J.-B. 2016. "Mismatch Repair". DNA Replication, Recombination, and Repair: Molecular Mechanisms and Pathology. Hanaoka, F.Sugasawa, K.: Tokyo: Springer Japan.

Fishel, R., Lescoe, M. K., Rao, M., Copeland, N. G., Jenkins, N. A., Garber, J., Kane, M. Kolodner, R. 1993. "The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer", Cell, 75(5), 1027-1038.

Friedberg, E. C. 2003. "DNA damage and repair", Nature, 421(6921), 436-440.

Fromme, J. C. Verdine, G. L. 2004. "Base Excision Repair". Advances in Protein Chemistry Academic Press.

Fukui, K. 2010. "DNA Mismatch Repair in Eukaryotes and Bacteria", Journal of Nucleic Acids, 2010, 260512.

Geng, H., Sakato, M., DeRocco, V., Yamane, K., Du, C., Erie, D. A., Hingorani, M. Hsieh, P. 2012. "Biochemical Analysis of the Human Mismatch Repair Proteins hMutSα MSH2(G674A)-MSH6 and MSH2-MSH6(T1219D)", Journal of Biological Chemistry, 287(13), 9777-9791.

Gillet, L. C. J. Schärer, O. D. 2006. "Molecular Mechanisms of Mammalian Global Genome Nucleotide Excision Repair", Chemical Reviews, 106(2), 253-276.

Gorbalenya, A. E. Koonin, E. V. 1990. "Superfamily of UvrA-related NTP-binding proteins. Implications for rational classification of recombination/repair systems", J Mol Biol, 213(4), 583-591.

Gradia, S., Acharya, S. Fishel, R. 1997. "The human mismatch recognition complex hMSH2-hMSH6 functions as a novel molecular switch", Cell, 91(7), 995-1005.

Gradia, S., Acharya, S. Fishel, R. 2000. "The Role of Mismatched Nucleotides in Activating the hMSH2-hMSH6 Molecular Switch *", Journal of Biological Chemistry, 275(6), 3922-3930.

Groothuizen, F. S., Winkler, I., Cristóvão, M., Fish, A., Winterwerp, H. H., Reumer, A., Marx, A. D., Hermans, N., Nicholls, R. A., Murshudov, G. N., Lebbink, J. H., Friedhoff, P. Sixma, T. K. 2015. "MutS/MutL crystal structure reveals that the MutS sliding clamp loads MutL onto DNA", Elife, 4, e06744.

Hakem, R. 2008. "DNA-damage repair; the good, the bad, and the ugly", Embo j, 27(4), 589-605.

Hanawalt, P. C. Spivak, G. 2008. "Transcription-coupled DNA repair: two decades of progress and surprises", Nature Reviews Molecular Cell Biology, 9(12), 958-970.

Heinen, C. D. 2016. "Mismatch repair defects and Lynch syndrome: The role of the basic scientist in the battle against cancer", DNA Repair (Amst), 38, 127-134.

Hoeijmakers, J. H. 2009. "DNA damage, aging, and cancer", N Engl J Med, 361(15), 1475-1485.

Hsieh, P. Yamane, K. 2008. "DNA mismatch repair: molecular mechanism, cancer, and ageing", Mech Ageing Dev, 129(7-8), 391-407.

Humphrey, W., Dalke, A. Schulten, K. 1996. "VMD: Visual molecular dynamics", Journal of Molecular Graphics, 14(1), 33-38.

I**accarino, I., Marra, G., Palombo, F. Jiricny, J.** 1998. "hMSH2 and hMSH6 play distinct roles in mismatch binding and contribute differently to the ATPase activity of hMutSαlpha", Embo j, 17(9), 2677-2686.

Iyer, R. R., Pluciennik, A., Burdett, V. Modrich, P. L. 2006. "DNA Mismatch Repair: Functions and Mechanisms", Chemical Reviews, 106(2), 302-323.

Jacob, S. Praz, F. 2002. "DNA mismatch repair defects: role in colorectal carcinogenesis", Biochimie, 84(1), 27-47.

Jiricny, J. 1998. "Eukaryotic mismatch repair: an update", Mutation Research/DNA Repair, 409(3), 107-121.

Junop, M. S., Obmolova, G., Rausch, K., Hsieh, P. Yang, W. 2001. "Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair", Mol Cell, 7(1), 1-12.

Kim, S., Oshima, H., Zhang, H., Kern, N. R., Re, S., Lee, J., Roux, B., Sugita, Y., Jiang, W. Im, W. 2020. "CHARMM-GUI Free Energy Calculator for Absolute and Relative Ligand Solvation and Binding Free Energy Simulations", J Chem Theory Comput, 16(11), 7207-7218.

Kolodner, R. D. Marsischky, G. T. 1999. "Eukaryotic DNA mismatch repair", Current Opinion in Genetics & Development, 9(1), 89-96.

Kraemer, K. H., Patronas, N. J., Schiffmann, R., Brooks, B. P., Tamura, D. DiGiovanna, J. J. 2007. "Xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome: a complex genotype-phenotype relationship", Neuroscience, 145(4), 1388-1396.

Krokan, H. E. Bjørås, M. 2013. "Base excision repair", Cold Spring Harb Perspect Biol, 5(4), a012583.

Krokan, H. E., Nilsen, H., Skorpen, F., Otterlei, M. Slupphaug, G. 2000. "Base excision repair of DNA in mammalian cells", FEBS Lett, 476(1-2), 73-77.

Kunkel, T. A. Erie, D. A. 2005. "DNA mismatch repair", Annu Rev Biochem, 74, 681-710.

Kusakabe, M., Onishi, Y., Tada, H., Kurihara, F., Kusao, K., Furukawa, M., Iwai, S., Yokoi, M., Sakai, W. Sugasawa, K. 2019. "Mechanism and regulation of DNA damage recognition in nucleotide excision repair", Genes and Environment, 41(1), 2.

Lamers, M. H., Winterwerp, H. H. Sixma, T. K. 2003. "The alternating ATPase domains of MutS control DNA mismatch repair", Embo j, 22(3), 746-756.

Leach, F. S., Nicolaides, N. C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., Peltomäki, P., Sistonen, P., Aaltonen, L. A. Nyström-Lahti, M. 1993. "Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer", Cell, 75(6), 1215-1225.

Leach, F. S., Nicolaides, N. C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., Peltomäki, P., Sistonen, P., Aaltonen, L. A., Nyström-Lahti, M., Guan, X. Y., Zhang, J., Meltzer, P. S., Yu, J.-W., Kao, F.-T., Chen, D. J., Cerosaletti, K. M., Fournier, R. E. K., Todd, S., Lewis, T., Leach, R. J., Naylor, S. L., Weissenbach, J., Mecklin, J.-P., Järvinen, H., Petersen, G. M., Hamilton, S. R., Green, J., Jass, J., Watson, P., Lynch, H. T., Trent, J. M., de la Chapelle, A., Kinzler, K. W. Vogelstein, B. 1993. "Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer", Cell, 75(6), 1215-1225.

Li, G.-M. 2008. "Mechanisms and functions of DNA mismatch repair", Cell Research, 18(1), 85-98.

Li, X. Heyer, W. D. 2008. "Homologous recombination in DNA repair and DNA damage tolerance", Cell Res, 18(1), 99-113.

Lin, D. P., Wang, Y., Scherer, S. J., Clark, A. B., Yang, K., Avdievich, E., Jin, B., Werling, U., Parris, T., Kurihara, N., Umar, A., Kucherlapati, R., Lipkin, M., Kunkel, T. A. Edelmann, W. 2004. "An Msh2 point mutation uncouples DNA mismatch repair and apoptosis", Cancer Res, 64(2), 517-522.

Lindahl, T. Barnes, D. E. 2000. "Repair of endogenous DNA damage", Cold Spring Harb Symp Quant Biol, 65, 127-133.

Loconte, D. C., Patruno, M., Lastella, P., Di Gregorio, C., Grossi, V., Forte, G., Ingravallo, G., Varvara, D., Bagnulo, R., Simone, C., Resta, N. Stella, A. 2014. "A rare MSH2 mutation causes defective binding to hMSH6, normal hMSH2 staining, and loss of hMSH6 at advanced cancer stage", Human Pathology, 45(10), 2162-2167.

Manhart, C. M. Alani, E. 2017. "DNA replication and mismatch repair safeguard against metabolic imbalances", Proc Natl Acad Sci U S A, 114(22), 5561-5563.

Marsischky, G. T., Filosi, N., Kane, M. F. Kolodner, R. 1996. "Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair", Genes Dev, 10(4), 407-420.

Martin, S. A., McCabe, N., Mullarkey, M., Cummins, R., Burgess, D. J., Nakabeppu, Y., Oka, S., Kay, E., Lord, C. J. Ashworth, A. 2010. "DNA Polymerases as Potential Therapeutic Targets for Cancers Deficient in the DNA Mismatch Repair Proteins MSH2 or MLH1", Cancer Cell, 17(3), 235-248.

McCammon, J. 1984. "Protein dynamics", Reports on Progress in Physics, 47(1), 1.

Mendillo, M. L., Hargreaves, V. V., Jamison, J. W., Mo, A. O., Li, S., Putnam, C. D., Woods, V. L. Kolodner, R. D. 2009. "A conserved MutS homolog connector domain interface interacts with MutL homologs", Proceedings of the National Academy of Sciences, 106(52), 22223-22228.

Modrich, P. 2006. "Mechanisms in eukaryotic mismatch repair", J Biol Chem, 281(41), 30305-30309.

Modrich, P. Lahue, R. 1996. "MISMATCH REPAIR IN REPLICATION FIDELITY, GENETIC RECOMBINATION, AND CANCER BIOLOGY", Annual Review of Biochemistry, 65(1), 101-133.

Mukherjee, S. Feig, M. 2009. "Conformational Change in MSH2-MSH6 upon Binding DNA Coupled to ATPase Activity", Biophysical Journal, 96(11), L63-L65.

Nagy, I. Z., Lustyik, G., Nagy, V. Z., Zarándi, B. Bertoni-Freddari, C. 1981. "Intracellular Na+:K+ ratios in human cancer cells as revealed by energy dispersive x-ray microanalysis", J Cell Biol, 90(3), 769-777.

Negureanu, L. Salsbury, F. R. 2012. "Insights into protein - DNA interactions, stability and allosteric communications: a computational study of mutSα-DNA recognition complexes", Journal of biomolecular structure & dynamics, 29(4), 757-776.

Negureanu, L. Salsbury, F. R. 2012. "The molecular origin of the MMR-dependent apoptosis pathway from dynamics analysis of MutSα-DNA complexes", Journal of Biomolecular Structure and Dynamics, 30(3), 347-361.

O'Donovan, P. J. Livingston, D. M. 2010. "BRCA1 and BRCA2: breast/ovarian cancer susceptibility gene products and participants in DNA double-strand break repair", Carcinogenesis, 31(6), 961-967.

Obmolova, G., Ban, C., Hsieh, P. Yang, W. 2000. "Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA", Nature, 407(6805), 703-710.

Onur, E., Tuğrul, B. Bozyiğit, F. 2009. "DNA hasarı and onarım mekanizmaları", Türk Klinik Biyokimya Dergisi, 7(2), 61-70.

P**alombo, F., Gallinari, P., Iaccarino, I., Lettieri, T., Hughes, M., D'Arrigo, A., Truong, O., Hsuan, J. J. Jiricny, J.** 1995. "GTBP, a 160-kilodalton protein essential for mismatch-binding activity in human cells", Science, 268(5219), 1912-1914.

Pâques, F. Haber, J. E. 1999. "Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae", Microbiol Mol Biol Rev, 63(2), 349-404.

Peltomäki, P. 2001. "DNA mismatch repair and cancer", Mutation Research/Reviews in Mutation Research, 488(1), 77-85.

Peltomäki, P. 2003. "Role of DNA Mismatch Repair Defects in the Pathogenesis of Human Cancer", Journal of Clinical Oncology, 21(6), 1174-1179.

Petruseva, I. O., Evdokimov, A. N. Lavrik, O. I. 2014. "Molecular mechanism of global genome nucleotide excision repair", Acta Naturae, 6(1), 23-34.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L. Schulten, K. 2005. "Scalable molecular dynamics with NAMD", J Comput Chem, 26(16), 1781-1802.

Prabhu, N. V. Sharp, K. A. 2005. "Heat capacity in proteins", Annual review of physical chemistry, 56(1), 521-548.

Puumalainen, M.-R., Rüthemann, P., Min, J.-H. Naegeli, H. 2016. "Xeroderma pigmentosum group C sensor: unprecedented recognition strategy and tight spatiotemporal regulation", Cellular and Molecular Life Sciences, 73(3), 547-566.

Sachadyn, P. 2010. "Conservation and diversity of MutS proteins", Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 694(1), 20-30.

Schärer, O. D. 2003. "Chemistry and Biology of DNA Repair", Angewandte Chemie International Edition, 42(26), 2946-2974.

S**chärer, O. D.** 2013. "Nucleotide excision repair in eukaryotes", Cold Spring Harb Perspect Biol, 5(10), a012609.

Schofield, M. J., Nayak, S., Scott, T. H., Du, C. Hsieh, P. 2001. "Interaction of Escherichia coli MutS and MutL at a DNA mismatch", J Biol Chem, 276(30), 28291-28299.

Stojic, L., Brun, R. Jiricny, J. 2004. "Mismatch repair and DNA damage signalling", DNA Repair (Amst), 3(8-9), 1091-1101.

Su, S. S. Modrich, P. 1986. "Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs", Proceedings of the National Academy of Sciences, 83(14), 5057-5061.

Tiraby, J. G. Fox, M. S. 1973. "Marker discrimination in transformation and mutation of pneumococcus", Proc Natl Acad Sci U S A, 70(12), 3541-3545.

Treffers, H. P., Spinelli, V. Belser, N. O. 1954. "A Factor (or Mutator Gene) Influencing Mutation Rates in Escherichia Coli", Proc Natl Acad Sci U S A, 40(11), 1064-1071.

Trenner, A. Sartori, A. A. 2019. "Harnessing DNA Double-Strand Break Repair for Cancer Treatment", Frontiers in Oncology, 9.

Tubbs, A. Nussenzweig, A. 2017. "Endogenous DNA Damage as a Source of Genomic Instability in Cancer", Cell, 168(4), 644-656.

van Gent, D. C., Hoeijmakers, J. H. Kanaar, R. 2001. "Chromosomal stability and the DNA double-stranded break connection", Nat Rev Genet, 2(3), 196-206.

Velho, S., Fernandes, M. S., Leite, M., Figueiredo, C. Seruca, R. 2014. "Causes and consequences of microsatellite instability in gastric carcinogenesis", World J Gastroenterol, 20(44), 16433-16442.

Verdine, G. L. Bruner, S. D. 1997. "How do DNA repair proteins locate damaged bases in the genome?", Chemistry & biology, 4(5), 329-334.

Vilenchik, M. M. Knudson, A. G. 2003. "Endogenous DNA double-strand breaks: Production, fidelity of repair, and induction of cancer", Proceedings of the National Academy of Sciences, 100(22), 12871-12876.

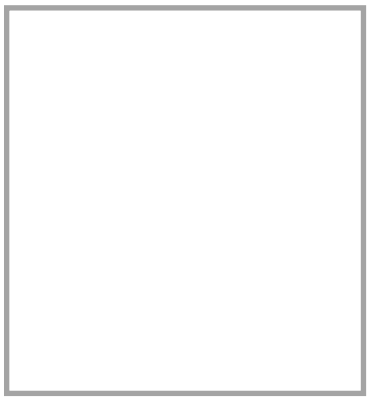
Vítor, A. C., Huertas, P., Legube, G. de Almeida, S. F. 2020. "Studying DNA Double-Strand Break Repair: An Ever-Growing Toolbox", Frontiers in Molecular Biosciences, 7.

Wang, B., Francis, J., Sharma, M., Law, S. M., Predeus, A. V. Feig, M. 2016. "Long-Range Signaling in MutS and MSH Homologs via Switching of Dynamic Communication Pathways", PLOS Computational Biology, 12(10), e1005159.

Warren, J. J., Pohlhaus, T. J., Changela, A., Iyer, R. R., Modrich, P. L. Beese, Lorena S. 2007. "Structure of the Human MutSα DNA Lesion Recognition Complex", Molecular Cell, 26(4), 579-592.

Woods, M. O., Williams, P., Careen, A., Edwards, L., Bartlett, S., McLaughlin, J. R. Younghusband, H. B. 2007. "A new variant database for mismatch repair genes associated with Lynch syndrome", Hum Mutat, 28(7), 669-673.

CURRICULUM VITAE



Name Surname : Clara Xazal Buran

EDUCATION :

- **B.Sc.** **:** 2020, Istanbul Technical University, Department of Molecular Biology and Genetics
- **M.Sc.** **:** Continuing, Istanbul Technical University, Graduate School of Science, Engineering and Technology, Department of Molecular Biology-Genetics and Biotechnology

PROJECTS:

- Kolon and Meme Kanserlerinde Erken Uyarı Sistemi Oluşturulması and Kişisel Sağlık Sistemi (E- Nabız) Veri Tabanına Entegrasyonu Projesi**,** TÜBİTAK project no. 318S129**.**