ACIBADEM MEHMET ALI AYDINLAR UNIVERSITY
INSTITUTE OF HEALTH SCIENCES

# COMPUTATIONAL INSIGHTS FROM ERAP1 AND ERAP2 VARIANTS FOR THE PATHOGENESIS OF ANKYLOSING SPONDYLITIS

YUNUS EMRE DİLEK

M.Sc. THESIS

DEPARTMENT OF MEDICAL BIOTECHNOLOGY

SUPERVISOR
Assoc. Prof. Günseli Bayram Akçapınar

ISTANBUL-2022

ACIBADEM MEHMET ALI AYDINLAR UNIVERSITY
INSTITUTE OF HEALTH SCIENCES

# COMPUTATIONAL INSIGHTS FROM ERAP1 AND ERAP2 VARIANTS FOR THE PATHOGENESIS OF ANKYLOSING SPONDYLITIS

YUNUS EMRE DİLEK

M.Sc. THESIS

DEPARTMENT OF MEDICAL BIOTECHNOLOGY

SUPERVISOR
Assoc. Prof. Günseli Bayram Akçapınar

ISTANBUL-2022

## DECLARATION

I declare that this thesis work is my own work, I had no unethical behavior at any stages from the planning to the writing of the thesis, I obtained all the information in this thesis in accordance with academic and ethical rules, I cited all the information and comments that were not obtained with this thesis work, and I provided resources in the list of references. I also declare that there was no violation of any patents and copyrights during the study and writing of this thesis.

03/06/2022

Yunus Emre Dilek

## PREFACE AND ACKNOWLEDGEMENT

I acknowledge my advisor Assoc. Prof. Günseli Bayram Akçapınar for her support throughout the entire thesis work from choosing it to presenting it. Dr. Bayram Akçapınar helped me with my first involvement in a project that I genuinely liked to work on and let me take a more directive path in my academic career. I acknowledge, a doctorate scholar of the same project Şeyma Çolakoğlu Özkaya for giving me generic details and some references about the thesis subject and helping me with many project related documents. I acknowledge Protein Engineering and Bioproduction Lab (PEBL) group members İrem Kara, Sena Kıvrak, and Hüseyin Okan Soykam for instructing me in using some computational tools and software as well as web servers. I thank all the members of PEBL group for their positivity and kindness throughout all time.

This thesis was supported by the Turkish Scientific and Technological Research Council, TÜBİTAK 1001 program [grant number 219S794]. I am grateful to TÜBİTAK for the financial support throughout my thesis.

I thank my jury members Prof. Züftü Tanıl Kocagöz and Asst. Prof. Fatma Tuba Gözet for providing their precious time, effort and for accepting to review my thesis and for their valuable comments.

Finally, I am indebted to my mother and my father for their love and moral support.

# TABLE OF CONTENTS

# LIST OF ABBREVIATIONS

| | |
|---|---|
| **Å** | Angstrom |
| **APN** | Aminopeptidase N |
| **AS** | Ankylosing spondylitis |
| **β2m** | Beta-2 microglobulin |
| **BSA** | Buried surface area |
| **Cα** | Alpha carbon |
| **CD8+** | Cytotoxic T-lymphocyte |
| **CM** | Center of mass |
| **Cryo-EM** | Cryogenic electron microscopy |
| **ER** | Endoplasmic reticulum |
| **ERAP1** | Endoplasmic reticulum aminopeptidase-I |
| **ERAP2** | Endoplasmic reticulum aminopeptidase-II |
| **FF** | Force field |
| **FHC** | Free heavy chain |
| **GWAS** | Genome-wide association study |
| **IRAP** | Insulin regulated aminopeptidase |
| **HLA** | Human leukocyte antigen |
| **K** | Kelvin |
| **LD** | Linkage disequilibrium |
| **MD** | Molecular dynamics |
| **MHC** | Major histocompatibility complex |
| **NK** | Natural killer |
| **ns** | Nanosecond |
| **PBC** | Periodic boundary condition |
| **PCA** | Principal component analysis |
| **PME** | Particle Mesh Ewald |
| **RGYR** | Radius of gyration |
| **RMSD** | Root mean square deviation |
| **RMSF** | Root mean square fluctuation |
| **SASA** | Solvent accessible surface area |

| | |
|---|---|
| **SNP** | Single nucleotide polymorphism |
| **SpA** | Spondyloarthritis |
| **TAP** | Transporter associated with antigen processing |
| **WTCCC** | Wellcome Trust Case Control Consortium |

# LIST OF FIGURES

# LIST OF TABLES

# ÖZET


**Ankilozan Spondilitin Patogeneziyle İlgili ERAP1 ve ERAP2 Varyantlarının Hesaplamalı Analizi**

Ankilozan Spondilit, artrit ve kemikleşmeye neden olan kronik enflamatuvar ve otoimmün hastalığıdır. Hastalığın oluşum nedeni belirsiz iken, ERAP1 ve ERAP2 genleri, HLA-B27 alelinden sonra, AS için ikinci sırada risk faktörü oluşturmaktadır. Bunlar, adaptif bağışıklık sisteminde yer alan ERAP proteinlerini kodlamaktadır. Bu proteinler, antijenik peptitlerin MHC-I'e uygun olarak bağlanması için bu peptitleri kesmekten sorumludur. Daha sonra peptitler, hücre yüzeyinde, MHC-I tarafından sitotoksik T lenfositlerine sunulur. Hipotezimize göre ERAP proteinlerindeki varyasyonlar ile dimerizasyon, enzim yapısında ve fonksiyonunda değişikliğe yol açmaktadır, bu da hücre yüzeyinde değişik çeşitlilikte bir peptit repertuvarına neden olmaktadır. Bu çerçevede, farklı ERAP allotiplerinin, ERAP monomer, homodimer ve heterodimer modellerinin yapı-fonksiyon ilişkisini anlamak için MD simülasyonları, PBC altında CHARMM36m kuvvet alanı kullanılarak 310.15 K'de yürütülmüştür. Dengeleme ve üretim simülasyonları sırasıyla NVT ve NPT grupları altında 2 ve 150 ns'de yürütülmüştür. Varyasyonların ve dimerizasyonun enzimin yapısına etkisi RMSD, RMSF, PCA ve RGYR hesaplamaları ile analiz edilmiştir. Sonuçlar, kapalı konformasyonların genellikle hem iki enzim hem de seçilen allotipler için daha kararlı olduğunu göstermiştir. Homodimerizasyon, dimer arayüzündeki tuz köprüsü ve hidrojen bağı etkileşimleri ile desteklenmiştir. Heterodimerizasyon, kısmen azalmış tuz köprüsü ve hidrojen bağı etkileşimleriyle daha az stabil bir yapıya neden olmuştur. Allotip yapılarındaki varyasyonlar, dimer yapılarında önemli değişimlere yol açmıştır. Bu sebeple, kapalı konformasyonun stabilizasyon için çok önemli olduğu bulunmuştur, dimerleşmenin ise yapısal ve fonksiyonel farklılıklara katkı sağladığı sonucuna varılmıştır.



**Anahtar Sözcükler:** Adaptif bağışıklık sistemi, Ankilozan spondilit, ERAP1, ERAP2, Moleküler dinamik simülasyonları

# ABSTRACT

**Computational Insights from ERAP1 and ERAP2 Variants for the Pathogenesis of Ankylosing Spondylitis**

Ankylosing Spondylitis is a chronic inflammatory autoimmune disease causing arthritis and ossification in joints. While its etiology is unclear, ERAP1 and ERAP2 genes were considered as second risk factors of AS only behind HLA-B27. These encode for ERAP proteins involved in adaptive immune system. They are responsible for proper trimming of antigenic peptides for their appropriate binding to MHC-I. Peptides are then presented to CD8+ T cells on the cell surface by MHC-I. Our hypothesis is that variations and dimerization in ERAP proteins can cause a change in structure and function of the enzymes resulting in altered peptide repertoire on cell surface. Within this scope, to understand the structure-function relationship of different ERAP allotypes, ERAP monomer, homodimer and heterodimer models, MD simulations were performed at 310.15 K using CHARMM36m ff under PBC. Equilibration and production simulations were run under NVT and NPT ensembles for 2 and 150 ns, respectively. The impact of variations and dimerization on the structure of enzymes were analyzed via RMSD, RMSF, PCA and RGYR calculations. Results demonstrated that closed conformations were generally more stable for both enzymes and selected allotypes. Homodimerization was supported through salt bridge and H-bonding interactions in dimer interface. Heterodimerization resulted in a less stable structure with partially diminished aforementioned interactions. Variations in allotype structures resulted in substantial changes in dimer structures. Therefore, closed conformation is found to be crucial for stabilization whereas dimerization contributes to structural and functional differences.



**Keywords:** Adaptive immune system, Ankylosing spondylitis, ERAP1, ERAP2, Molecular dynamics simulations

# 1 INTRODUCTION AND AIM

Ankylosing Spondylitis (AS) is a common type of spondyloarthritis (SpA) in which long-term inflammation occurs at the joints of the spine. As time progresses, other axial joints are also affected from the disease and stiffness of diseased joints worsens which causes chronic pain. In later development of the disease, ossification of joints grows leading to structural and functional impairments of the body (1–5).

Pathogenesis of AS is currently unknown. It is considered to be dependent on both genetic and environmental factors. Genetic factors were determined to be the risk factor for the disease at more than 90% (2,4–6). Among the genetic factors examined, human leukocyte antigen-B27 (HLA-B27*) allele was found to have a significant effect on the development of disease with about a third of all genetic factors related to the disease (1,3–5). While HLA-B27 is the major risk factor, there were only a few HLA-B27-positive individuals that developed the disease, suggesting that more genes were involved in susceptibility to disease (2,7). In later studies, new gene groups called endoplasmic reticulum aminopeptidase-I (ERAP1) and -II (ERAP2) were discovered and ERAP1 was associated with AS as the second major risk factor after HLA-B27 (3–6,8,9). Moreover, Vitulano *et al.* indicated an epistatic interaction between HLA-B27 and ERAP1 single nucleotide polymorphisms (SNPs) (10).

Peptide cleaving ERAP1 and ERAP2 enzymes are from M1 family of zinc aminopeptidases. In adaptive immune system, they are involved in antigen processing and major histocompatibility complex-I (MHC-I) assembly pathway (3,6,11,12).

In a classical antigen processing and MHC-I assembly pathway, a complex cytosolic antigenic protein is broken down into peptides by proteasome. While C-terminal end of antigenic peptides are sufficiently cleaved by proteasome, N-terminal end has extended parts left in long peptides. For appropriate binding to MHC-I, these peptides are first transported into endoplasmic reticulum (ER) via transporter associated with antigen processing (TAP) protein. Afterwards, they are further

* HLA-B27 is a MHC-I allele that has the greatest genetic risk among other HLA alleles (2).

trimmed by ERAP enzymes. Newly-formed peptides with optimal length are loaded onto MHC-I, establishing peptide–MHC-I complex. At last, the complex is transported on cell surface, making up the peptide repertoire[1] of the cell (Figure 1) (3,8,11,12).

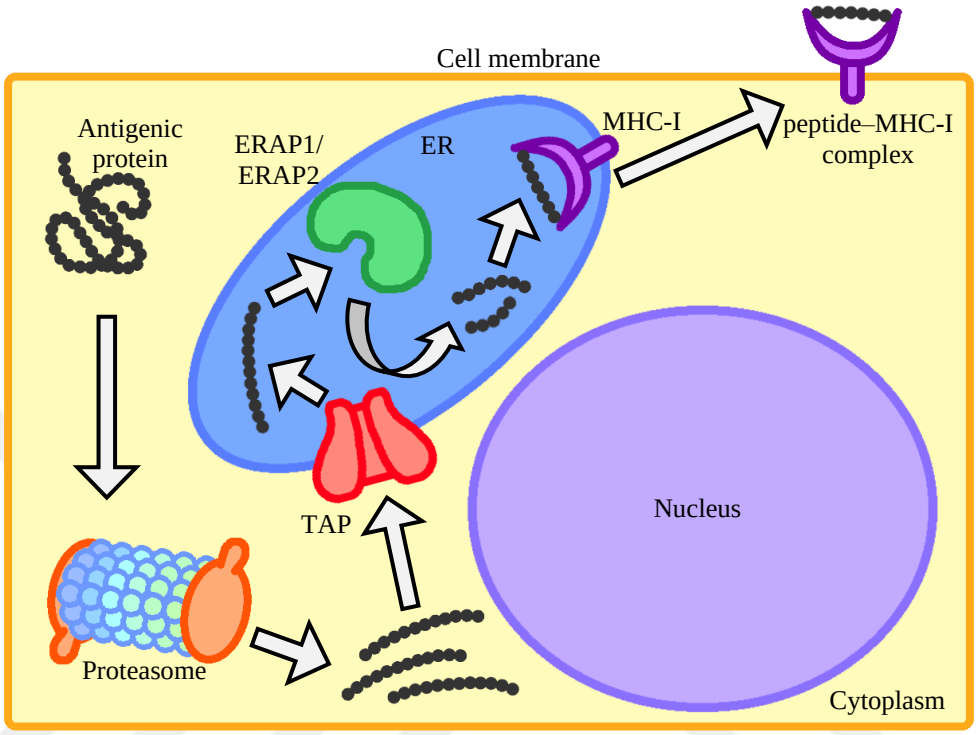



Figure 1. Classical antigen processing and MHC-I assembly pathway. Figure adapted from Agrawal *et al.* (5).

Several theories were brought forward regarding pathogenesis of AS to explain the role of HLA-B27 as well as ERAP1/2. Polymorphisms in ERAP1 and ERAP2 can give rise to the protein variants with altered function. These variants can affect the trimming of an antigenic peptide by the enzymes causing different response. In one of the theories, abnormal peptide forms a complex with HLA-B27. Consequently, the arthritogenic[2] peptide presented on cell surface is detected by cytotoxic T-lymphocytes (CD8+) that induces aberrant immune response. For the second theory, binding of abnormal peptide dissociates beta-2 microglobulin (β2m) subunit of HLA-B27. Upon dissociation, HLA-B27 heavy chains are released in free or homodimer form. Free heavy chains (FHCs) are either misfolded and accumulated in endoplasmic reticulum

[1] Encompasses all antigenic peptides presented by MHC-I.

[2] A peptide that is capable of causing arthritis.

(ER) causing ER stress or transported on membrane. Homodimers are transported on cell surface where specific binding of HLA-B27 homodimers to natural killer (NK) cells and T lymphocytes activates immune response. In another theory, misfolded peptide is accumulated in ER, creating unfolded protein response and ER stress. These responses eventually enable autophagy pathways (Figure 2) (2–5,8,13).

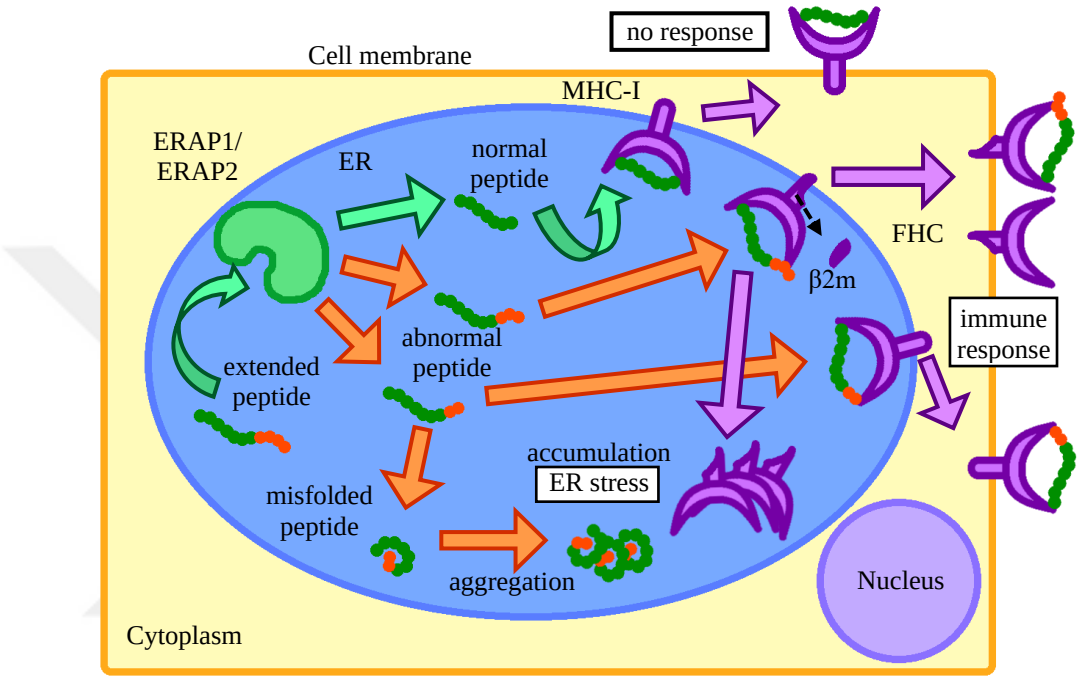



Figure 2. Theories regarding pathogenesis of AS involving ERAP and HLA-B27. Figure adapted from Haroon *et al.* (8).

Our hypothesis is that polymorphisms in ERAP1 and ERAP2 can change the peptide trimming efficiency or ERAP catalytic activity thereby causing an alteration in MHC-I immunopeptidome profiles. Homo- and heterodimer allotypes of ERAP1 and formation of ERAP1–ERAP2 heterodimers may contribute to this phenomenon. Differentiated peptides can be recognized as non-self by CD8+ that it induces immune response causing the development of AS.

The purpose of this study is to analyze the structural differences between ERAP1 allotypes, homo- and heterodimers along with ERAP2 and their possible contribution to AS mechanisms.

# 2 BACKGROUND

## 2.1 Overview of ERAP and ERAP-Related Proteins

A protein called ERAP1 belongs to oxytocinase subfamily of M1 aminopeptidases[1]. Along with ERAP1, ERAP2 and insulin regulated aminopeptidase (IRAP), as well, belong to this subfamily and they have 49% and 43% sequence homology to ERAP1, respectively (15). All three proteins are found localized in intracellular compartment of the cell where ERAP1/2 proteins exhibit ER localization and IRAP protein exhibits localization in vesicles and on cell membrane (16). Among other cellular functions, ERAP1/2 perform endogenous antigen processing by editing antigenic peptides from their N-terminal end to reduce them to their optimal length to load on MHC-I (11,17). A similar antigen processing role for cross-presentation[2] of generated peptides is available in IRAP besides other roles (15). Out of the three proteins identified in humans, only ERAP1 and IRAP are present in mice. ERAP2 has specificity to different peptides that complements with ERAP1 during antigen processing in humans (18,19).

## 2.2 Structures of ERAP1 and ERAP2

Structure of ERAP1 is comprised of four domains, general functions of which are just like other M1 aminopeptidases (Figure 3). Domain I (between 0–254 residues) has a beta-sheet region, part of which interacts with domain IV from the loop and it is important for steric hindrance effect during catalytic reaction (20,21). Near the protein N-terminus, a hydrophobic interaction domain is located that can be found tethered to ER membrane which is currently unstructured in all crystal models (22). Domain II (between 255–529 residues) has the catalytic site of the protein (Figure 4). It contains conserved GAMEN motif and zincin motif ($HEXXHX_{18}E$). Asn321 stabilizes the loops in the region and Glu320 aids the enzymatic reaction. Conserved catalytic

[1] Protease enzymes that catalyze N-terminal aminoacid removal from polypeptides. M1 family of this group requires a zinc ion to work (14).

[2] An ability of antigen presenting cells (APCs) to present antigenic peptides with MHC-I molecules to CD8+.

residue Tyr438 is found further away from other catalytic residues and it is essential for stabilizing intermediates (20). Zinc binding residues His353, His357 and Glu376 interact with the catalytic Zn(II) ion to keep it coordinated (20,21). There is also an ER retention maintaining sequence that retains ERAP1 in ER which is currently unstructured in ERAP1 models (23). Domain III (between 530–614 residues) has a beta-sheet region that works like a hinge between neighbor domains in order to change the conformational state. Domain IV (between 615–940 residues) includes regulatory sites of the protein. It mainly features an alpha-helical structure and it is proposed to be the domain where C-terminus of a substrate is linked to (24). Protein–protein interaction region is thought to be located around this domain as Wong *et al.* reported a protein of the same family called Aminopeptidease N (APN) that forms a dimer structure via interaction on the same domain (25) and has about 33% sequence homology to ERAP proteins (26).

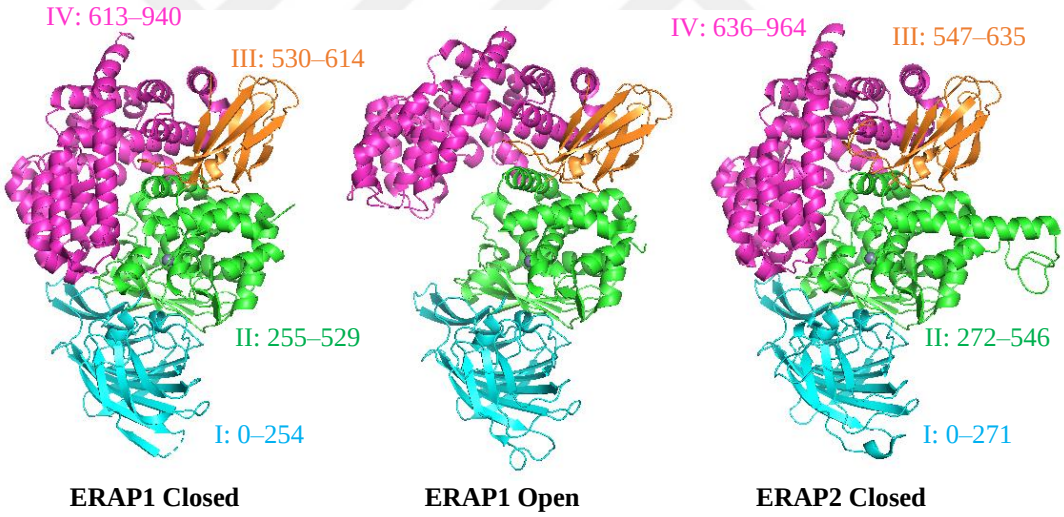



Figure 3. Structure of ERAP proteins. ERAP1 closed (left, PDB ID: 2YD0), ERAP1 open (middle, PDB ID: 3MDJ) and ERAP2 closed (right, PDB ID: 5AB0) conformation states colored by domain.

Similar structural arrangement of ERAP1 is found in ERAP2 because of close homology within the sequence (Figure 3). Position of common motifs in the catalytic site are found shifted 17 aminoacids further in the sequence.

Internal cavity forms in the closed conformation state of ERAP1 and ERAP2 and it is surrounded by domain I, II and IV (Figure 4). Domain IV accommodates a substantial part of the cavity compared to other known proteins in the same family.

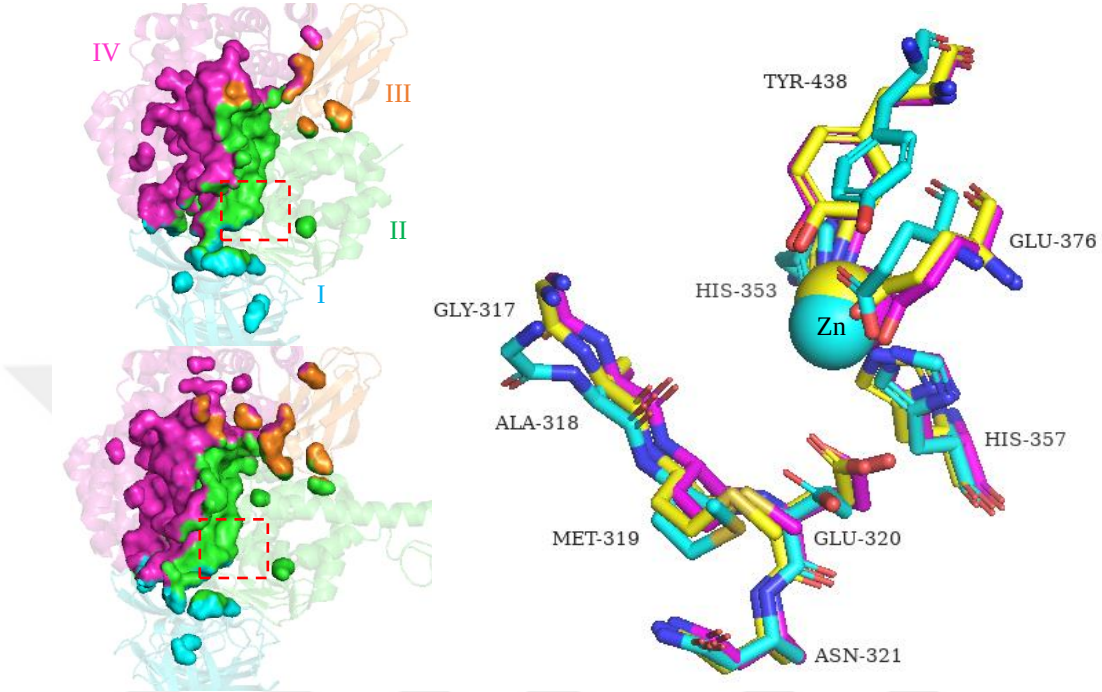



Figure 4. Internal cavity and catalytic site of ERAP proteins. ERAP1 closed (top left, PDB ID: 2YD0) and ERAP2 closed (top right, PDB ID: 5AB0) conformation states colored by domain. Catalytic sites are indicated with the red dashed rectangle. Superimposition of catalytic sites in ERAP1 closed (left, magenta, PDB ID: 2YD0), ERAP1 open (left, cyan, PDB ID: 3MDJ) and ERAP2 closed (left, yellow, PDB ID: 5AB0) conformation states.

While structures of ERAP proteins are currently available with their soluble domains, their N-terminal sequences are largely unstructured. In ERAP1, this region corresponds to a transmembrane domain which tethers ERAP1 to ER membrane (Figure 5). While this region does not contain any ectodomain cleavage site, there is a signal peptide sequence (7).

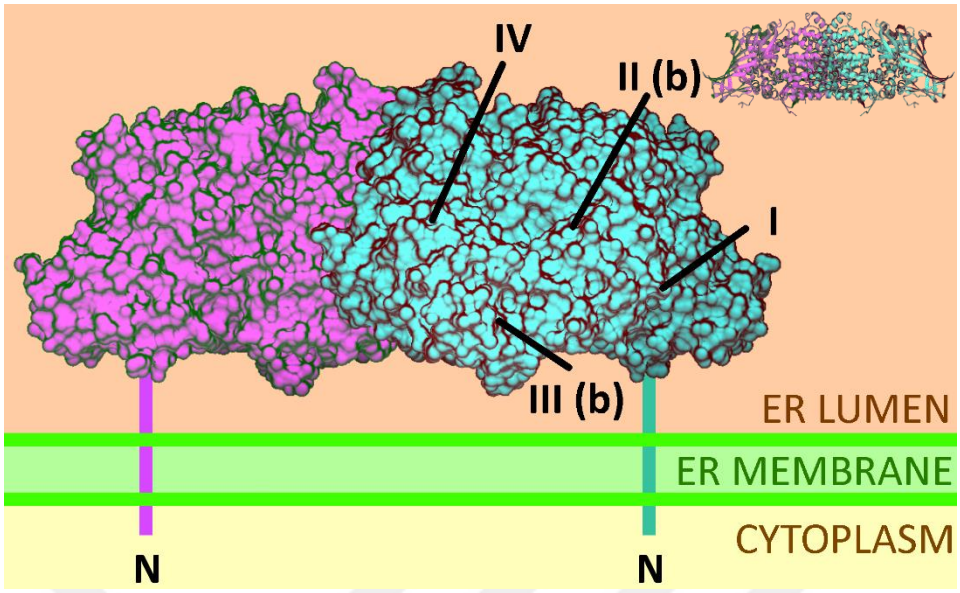



Figure 5. Hypothetical structure of membrane-bound ERAP1 dimer. N-terminus and domains are indicated with single letters and roman numerals, respectively (b: domain located in the back side) and chains are colored. Cartoon representation is shown at the top right corner.

### 2.3 Mechanism of Action in ERAP1 and M1 Family

To date, no substrate–ERAP1 complex structure was known in pre-catalysis phase. On the other hand, an inhibitor called bestatin was exhibited to form a complex with ERAP1 in the crystal structure, in both closed (21) and open states (20). According to the solved structures, N-terminus of the inhibitor is positioned within the active site suggesting N-terminus location of the substrate in the same region. Studies involving other proteins of M1 family were also published in bestatin-bound (27,28) and peptide analogue-bound forms (29,30). Therefore, studies in this area are mostly focused on the inhibitor–enzyme complex structures along with single point mutations based on Hap2 structure. Structures of individual allotypes, effects of combinations of SNPs forming the allotypes on the enzyme structure and function are still missing from the literature.

There are several regulatory sites situated in ERAP1 (30,31). These sites are located at the opposite side of the internal cavity neighboring the catalytic site domain and they usually correspond to C-terminus of long peptides (24,32) and some other

ligands (30,32–34). While long peptides interacting with both catalytic and regulatory sites from both terminal ends can induce a catalytic reaction depending on the affinity of the enzyme (20,24), small peptides interacting with catalytic and regulatory sites can have an inhibitory effect on ERAP1 affinity to long peptides and an activating effect on ERAP1 affinity to shorter peptides, respectively (20,30,34).

Functionally, ERAP1 has two distinct conformational states with different enzymatic activities (Figure 3) (20,21,31). In open conformational state, ERAP1 is hypothesized to be inactive and it is available for substrate binding (21). Upon binding, ERAP1 changes its shape into a more active, closed conformational state. Here, domain I and II gets closer to domain IV by the rotation on domain III (hinge) to close up the inside from any solvent access (10,20). More deeply, translation and rotation of helix H4a is coupled with rotation of helix H5. Some residues lining the S1 specificity pocket (such as Phe433) and Tyr438 are repositioned accordingly and Tyr438 gets closer to the bound substrate, Zn(II) ion and other active site residues in closed conformational state (20,31). This starts changes in active site leading to a catalytic reaction that cleaves the N-terminus of the substrate by hydrolysis (35). After the reaction, ERAP1 returns to its open state and releases the product (21). These conformational changes are consistent with other M1 aminopeptidases (20).

In theory, the mechanism involves a zinc ion and a water molecule. A peptide bound to ERAP1 within the active site is polarized with the Zn(II) ion from its carbonyl group triggering a nucleophilic attack by the water molecule on its amino group. This breaks the scissile bond of the peptide and the transition becomes unstable that it generates two product molecules by destroying the reactant peptide (20,36,37).

When it comes to peptide trimming, two theoretical mechanisms are debated (Figure 6) (15,38). In ERAP1 molecular ruler mechanism, peptides are trimmed optimally for binding MHC-I effectively. In MHC-I template mechanism; MHC-I binds a peptide of interest first, N-terminal extension of the bound peptide is trimmed by ERAP1 later. Several evidences were reported for both mechanisms. In some studies, ERAP1 trimmed several peptides in a length-dependent manner causing

peptides of certain length to be in optimal size for MHC-I presentation and disregarding peptides out of the range (20,24,32,39,40). In other studies, peptide partially disassociated from MHC-I with a protruding N-terminal end was trimmed by ERAP1 efficiently (41,42). Moreover, further trimming of an optimal size peptide was protected by MHC-I in some cases (43,44). However, no ERAP1 with a fully-open conformation state has been obtained from crystal structure models. While molecular ruler mechanism was strongly evidenced (45), both mechanisms could be involved in peptide trimming concertedly based on peptide preference.

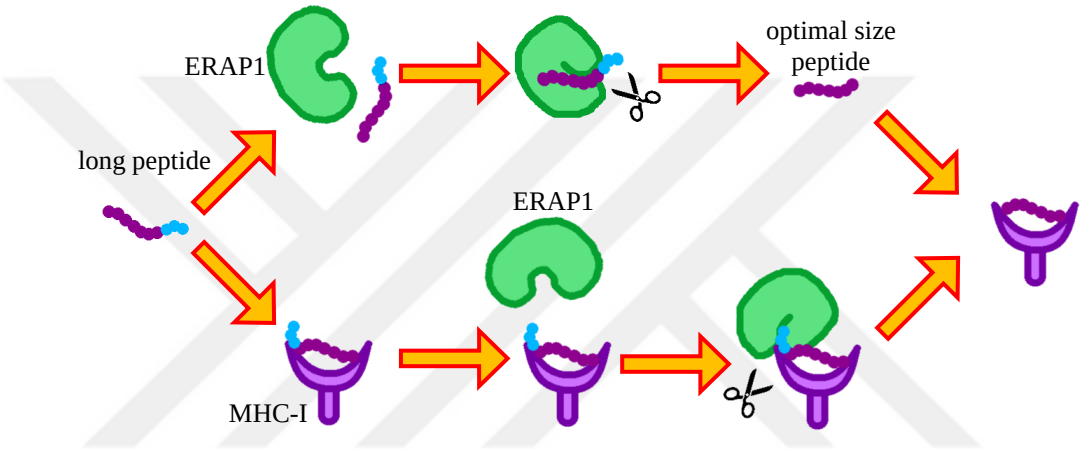



Figure 6. Two theoretical mechanisms of peptide trimming. Molecular ruler mechanism (top) and MHC-I template mechanism (bottom). Figure adapted from Colbert et al. (45).

In molecular ruler mechanism, ERAP1 sets a ground for MHC-I peptide repertoire by trimming peptides of 9–16 amino acid length to 8–10 amino acid length (24,39,46,47). Allosteric binding of the peptide and ERAP1 helps activation of this mechanism (32). N- and C-terminal ends (33), together with the alignment of internal residues (48) determine substrate specificity of ERAP1. While ERAP1 has an affinity to hydrophobic residues such as Leu at N-terminus, it has weak interactions with hydrophilic and charged residues. As for Pro, it is never hydrolyzed by ERAP1. However, this preference is coupled with internal sequence of the substrate. Residues surrounding the internal cavity of ERAP1 carry negative charges in overall with the preference for binding positively charged residues such as Arg. Some hydrophobic pockets are also present in the cavity (4,16,20,24,35,48,49).

Despite the similarities in the whole structure and domain organization, ERAP2 has distinct specificity and substrate preference compared to ERAP1 due to the structural differences in active site and peptide binding sites. For instance, ERAP2 has better trimming activity with shorter peptides that ERAP1 cannot trim and ERAP2 activity decreases as peptides get longer. Furthermore, ERAP2 has preference for positively charged residues such as Arg at N-terminus unlike ERAP1 (50,51). In the internal cavity, more positive electrostatic potential was observed (52), as distinct from ERAP1 possessing an internal cavity with strong negative potential (48).

Interdependence of ERAP1 and ERAP2 comes with another matter in terms of peptide processing. *In vivo* study of Saveanu *et al.* found strong findings to a possible heterodimeric form of ERAP1 and ERAP2 structures (19). Subsequently, *in vitro* studies attempted to reproduce the heterodimers as similarly as the previous study and observed induced ERAP1 activation upon heterodimerization (19,42,49). Peptide trimming efficiency was increased by the improvement in ERAP1 substrate affinity in complement with ERAP2 while a substantial decrease in ERAP2 substrate affinity was detected (49). In other *in vitro* study, heterodimers showed trimming activity for MHC-I-bound long peptides and increased stability of peptide–MHC-I complex (42). However, ERAP1 and ERAP2 enzymes are mostly monomeric in the enzyme pool in which only 10–30% of their structures are found in heterodimer form (50). In addition, it has not yet been cleared up for what extent both ERAP proteins interact *in vivo*. In regards to molecular structure, no topographic model of the aforementioned heterodimer has been provided as well as the site at which the dimerization occurs. Nevertheless, in a very recent paper, Papakyriakou *et al.*, performed the first computational analysis of ERAP1/2 heterodimer (51). This study slightly overlaps with the topic of this thesis in terms of the use of homo- and heterodimers for only Hap2 and ERAP2 (N392). However, it does not include other allotypes. Moreover, neither the ER retention signal loop is adequately modeled, nor dimerization was based on molecular docking. Due to the lack of a crystal structure, they used crystallographic homodimer structures of ERAP2 and homologous proteins (IRAP and APN) and substituted their chains for ERAP1 and ERAP2. Exon 10 loop of ERAP1 was structurally grafted from ERAP2. Heterodimer templates with a different binding

mode were simulated. Binding modes used were domain I interactions in asymmetric unit and domain II interactions between a chain of asymmetric unit and a symmetry-related chain from ERAP2 homodimer crystal structure and domain IV interactions from both IRAP and APN homodimer crystal structures. In ERAP1/2 heterodimer simulations, interactions involving exon 10 loop of domain II in both chains received rearrangement within simulation time when domain III of ERAP1 and domain I of ERAP2 joined. Analyses demonstrated ERAP1/2 heterodimer with these interactions as the most stable. However, M1 family proteins are usually found with protein–protein interaction domain in domain IV (22). Therefore, dimer interactions acquired from M1 protein family can be helpful in the way for finding the dimerization site of ERAP1/2.

One recent study from Mattorre *et al.* discovered a shorter ERAP2 sequence formed in a more acidic environment upon autocleavage and expressed by macrophages and some cancer cells (52). Interestingly, this shorter structure observed to be bound to IRAP protein and localized in both cell membrane and endosomes. While this event occurred independently of allotype variations, studies about the role of ERAP2 in cellular functions may require more elaboration.

### 2.4 Polymorphisms of ERAP1 and ERAP2

Naturally, ERAP1 is a highly polymorphic protein with variable levels of protein expression and enzymatic activity among human population. Many polymorphisms in ERAP1 are located near catalytic site (at positions 346 and 349), peptide binding region (at positions 725 and 730) or regions that affect conformational rearrangements (at positions 528 and 575). These polymorphisms affect ERAP1 activity and specificity (Figure 7). Several *in vitro* studies demonstrated major effects of K528R (rs30187) SNP on conformational switch of the enzyme. On the other hand, Q730E (rs27044) SNP signified a change in ERAP1 preference for substrates by increasing its affinity to shorter peptides (15,53,54). These SNPs are located at the regulatory site of ERAP1 distant from the catalytic site. Therefore, they may have an indirect effect on immunopeptidome (55).

In ERAP2, polymorphisms are poor and non-synonymous changes are limited. Thus far, there is one known SNP (K392N, rs2549782) proven to change ERAP2 enzymatic activity (56). Another SNP (A/G allele, rs2248374) known for destabilizing RNA transcript of the enzyme is also available. This variant leads to loss in ERAP2 expression but it is a rare case in the general population (58–60).

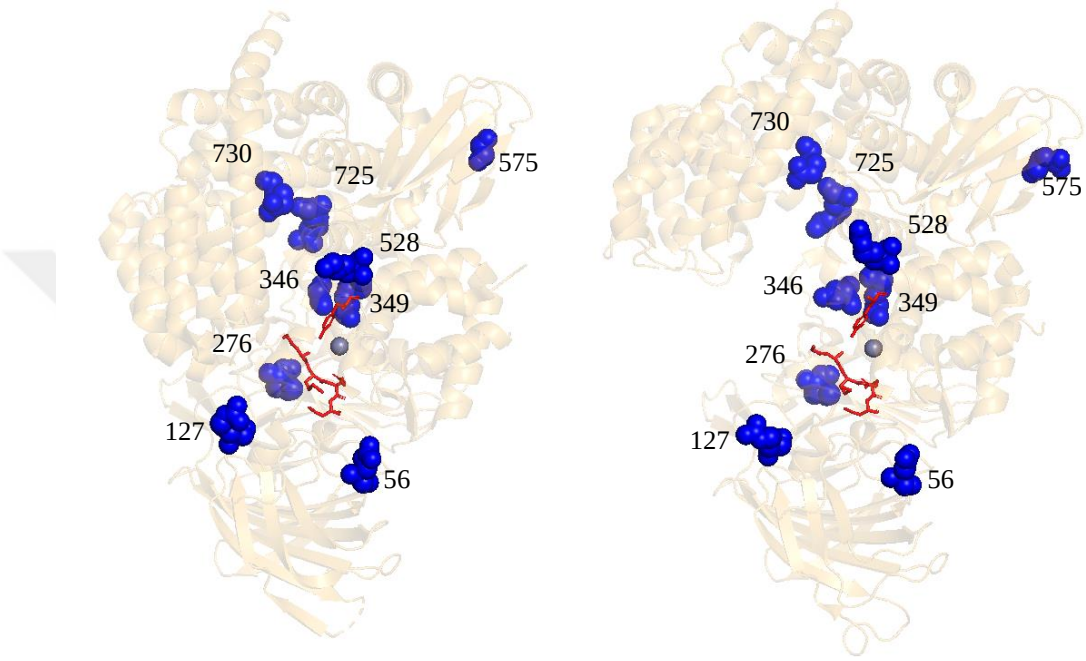



Figure 7. Location of common variants in ERAP1 structure. ERAP1 closed (left, PDB ID: 2YD0) and ERAP1 open (right, PDB ID: 3MDJ) were shown. Whole protein structures (transparent, orange cartoon), their SNPs (blue spheres with residue numbers), active site residue (red licorice) and Zn(II) (grey sphere) were represented.

## 2.5 Association of ERAP1 and ERAP2 Variants with AS

In order to avoid misconceptions, it is important to point out that haplotypes and allotypes are different terms which were used interchangeably in many studies. Haplotypes describe only single SNPs whereas allotypes describe combination of SNP groups/haplotypes (57,58). Hereupon, ERAP1 common SNP groups will be referred as allotypes in this study.

A comprehensive study was done by several research groups to report effects of ERAP1 SNP variants. In a study WTCCC (Wellcome Trust Case Control Consortium)

performed, K528 (rs30187) and Q730 (rs27044) were associated with AS as risk factors, and N575 (rs10050860), V349 (rs2287987) and Q725 (rs17482078) were associated with AS as protective factors (59). Harvey *et al.* found 6 AS-associated SNPs (R127, M349, K528, D575, R725, and Q730) of European ancestry (60). Maksymowich *et al.* studied AS patients in Canada and recorded 3 ERAP1 SNPs (P127R, R528K and D575N) as significantly associated for AS (61). In ERAP1, SNPs associated as AS protective confer loss-of-function in ERAP1 causing abundance of long peptides in MHC-I immunopeptidome (47), whereas AS-predisposing SNPs are functionally more active, cause overtrimming and low level of MHC-I-presented peptides. Although many more SNPs were determined, their association with AS was not distinguished due to tight linkage disequilibrium (LD[*]). This means distinct combinations of complex residues are also in charge of association.

Table 1. Allotypes of common ERAP1 missense SNPs. Table adapted from Ombrello *et al.* (51).

| | **SNP Positions** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Allotypes** | 56 | 127 | 276 | 346 | 349 | 528 | 575 | 725 | 730 |
| Hap1 (ancestral) | E | P | I | G | M | K | D | R | Q |
| Hap2 | E | R | I | G | M | K | D | R | Q |
| Hap3 | E | R | I | G | M | K | D | R | E |
| Hap4 | E | R | I | G | M | R | D | R | E |
| Hap5 | E | R | I | D | M | R | D | R | E |
| Hap6 | E | P | I | G | M | R | D | R | E |
| Hap7 | K | P | I | G | M | R | D | R | E |
| Hap8 | E | P | M | G | M | R | D | R | E |
| Hap9 | E | P | M | G | M | R | N | R | E |
| Hap10 | E | P | I | G | V | R | N | Q | E |

[*] Happens when two or more alleles are non-randomly associated. HLA genes are widely impacted by this concept.

For the investigation of SNP groups associated with AS, allotypes were also identified. In WTCCC research, Hap1–Hap10 were determined as 10 common ERAP1 allotypes including 99% of all ERAP1 variants in European population (Table 1) (53,58–60). Hutchinson *et al.* demonstrated common ERAP1 allotypes in a wider population data and used two allelic combinations of individuals to compare enzyme activity versus population frequency. Notably, they classified Hap2–Hap2, Hap2–Hap8 as high activity-high frequency and Hap10–Hap10 as low activity-moderate frequency with clear evidence, distinct from other allele groups (58).

Allotypes including K528-D575-E730 SNP group (KDE) confers risk for AS, whereas allotypes including P127-I276-R528 SNP group (PIR) protects against AS (53). While KDE is only present in Hap3, PIR is present in Hap6, 7 and 10. Evans *et al.* reported independent effects of K528 and N575 in a genome-wide association study (GWAS[*]). Their analyses showed that AS risk was 3–4 times reduced in individuals with R528 and N575 homozygote (RN) (9). Kadi *et al.* identified K528-D575-R725 (KDR) as risk group and R528-N575-Q725 (RNQ) as protective group, and classified Hap1–3 as risk allotypes and Hap10 as protective allotype (62). Subsequently, Bettencourt *et al.* identified M349-K528-D575-R725 (MKDR) as risk group and V349-R528-N575-Q725 (VRNQ) as protective group, once again, classifying Hap1–3 and Hap10 in the same way. Reeves *et al.* investigated effects of SNPs in combinatorial fashion. Results in the analyses demonstrated that Hap2 encodes an ERAP1 molecule that trims peptides efficiently leading to high production of peptide–MHC-I complex. On the contrary, Hap10 encoding a different ERAP1 structure trimmed peptides poorly and caused low production of peptide–MHC-I complex (57). Considering all these results together, two independent effects are evident in AS association; one favoring AS risk, the other favoring AS protection.

In a study with ERAP2 variants, Yao *et al.* reported G allele of rs2248374 as protective against AS by causing loss of ERAP2 expression (63). Wiśniewski *et al.* studied Polish population to evaluate AS risk of both ERAP1 and ERAP2 variants. Here, ERAP2 rs2248374 variant alone did not induce any effect on ERAP2 for AS.

[*] A method for determining effects of genetic variants in particular diseases.

However, ERAP2 A allele (known for encoding a full-length ERAP2 (64)) in combination with high activity ERAP1 (M349-K528-E730) increased risk of AS. By contrast, ERAP2 G allele (known for encoding a non-functional ERAP2 (64)) together with low activity ERAP1 (V349-R528-E730) was protective against AS (65).

All in all, AS development is associated with ERAP1/2 polymorphisms in a complex manner. Pathogenic role of MHC-I strongly depends on these polymorphisms due to associations between MHC-I and ERAP1/2 in regards to peptide binding (66,67). Unfortunately, high degree of ERAP1 polymorphism and various ERAP1 variants among human population points at the adversity of designating each polymorphic effect on shaping MHC-I immunopeptidome. Therefore, observing structural changes of ERAP1 and ERAP2 allotypes in both monomer and dimer forms can help getting to know the disease mechanism further by moving one step closer.

# 3 MATERIALS AND METHODS

Experimental studies performed on ERAP1/2 proteins are still limited in terms of structure and function. Size of ERAP proteins is large and they can exist as both soluble and membrane-bound proteins in cells (7,68). Due to the unsolved restrictions of membrane proteins, current structural data available for these enzymes were only captured in crystal structure. On the other hand, the lack of discovery of ERAP substrates led to construction of protein–ligand complexes mostly constituting of peptide analogues with low specificity.

Recent technological advancements in computational methods can be beneficial for revealing the underlying mechanisms of AS pathogenesis in relation to ERAP proteins with a focus on their structural differences. Following diagram outlines applied computational methods in this study (Figure 8).

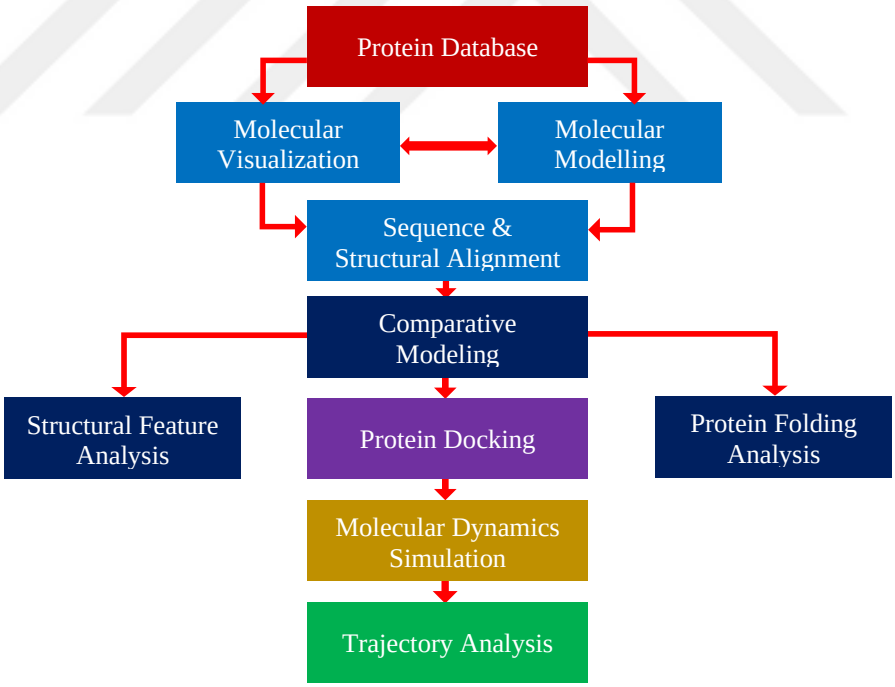



Figure 8. An overview of computational methods used in this study.

To summarize the methods of the study, crystal structures of proteins from protein structure database were downloaded. Using molecular visualization and molecular

modelling techniques together, best structures were selected and modified accordingly. Sequence and structural alignments were performed in preparation to comparative modeling. Afterwards, sequence gaps were remodeled by comparative modeling. Best remodeled structures were selected and evaluated in terms of electrostatic potential and hydrophobicity by using structural feature analysis. For the aim of predicting ERAP dimer structures, protein docking was carried out. In initial analyses, stability was calculated for both monomers and dimers to narrow down the selected structures for MD simulation. Monomers and dimers of ERAP1 and ERAP2 structures simulated with an attempt to understand structural variations using this technique. Finally, MD trajectories were analyzed to determine the resulting structural changes.

## 3.1 Data Collection, Molecular Modelling and Comparative Modeling

### 3.1.1 Extraction of protein structures and sequences

Protein sequence of human ERAP1 and ERAP2 proteins were provided from UniProt website using the following entries (69) in UniProtKB: Q9NZ08 and Q6P179. All available crystal structures of relevant proteins were also downloaded from RCSB PDB website (70). These include chain A of 2YD0 (46–940 amino acids, ligand-bound ERAP1 closed structure), chain B of 3MDJ (46–934 amino acids, ligand-bound ERAP1 open structure), chain B of 3QNF (46–935 amino acids, ERAP1 open apo-enzyme structure) and chain A of 5AB0 (54–964 amino acids, ligand-bound ERAP2 closed structure). Missing sequences were identified (excluding N- and C-terminal sequences; 111–114 and 486–513 amino acids in 2YD0, 111–113, 417–433, 486–514, 552–555, 864–867 and 893–906 amino acids in 3MDJ, 111–114, 426–433, 552–558, 758–799 and 811–914 amino acids in 3QNF, full-length in 5AB0). General information about protein chain interactions, ligands and ions were collected from PDBsum website to locate active site and zinc binding residues (71). PyMOL (v2.5.1) and VMD (v1.9.3) softwares were used for visualization of protein structures, modelling and analysis (72,73).

All available ERAP1 and ERAP2 structures were sorted by their quality and quantity where only a selection of structures was used in this study. Selected structures with their PDB ID are shown above (Table 2). Chains used from PDBs were chain A of 2YD0 and 5AB0 and chain B of 3MDJ and 3QNF. All selected ERAP1 structures are comprised of Hap2 sequence except 3MDJ which had a Hap5 sequence initially and ERAP2 5AB0 structure has N392 SNP. All desired allotypes (ERAP1 Hap1–10 and ERAP2 N392 & K392) were generated by mutations in original structures.

Table 2. Protein structures used in this MD study and modelled allotypes.

| PDB ID(s) | Allotype | Modelled Allotypes | Description |
|---|---|---|---|
| 2YD0 | Hap2 | Hap1–10 | ERAP1 closed conformation |
| | Hap2 | Original | ERAP1 closed conformation w/ bestatin ligand |
| 3MDJ | Hap5 | Hap2 | ERAP1 open conformation |
| | Hap5 | Hap2 | ERAP1 open conformation w/ bestatin ligand |
| 3QNF | Hap2 | Original | ERAP1 open conformation |
| 5AB0 | N392 | Original/K392 | ERAP2 closed conformation |
| 2YD0–2YD0 | Hap2 | Original/Hap8 | ERAP1 closed conformation homodimers, Hap2–Hap2 and Hap8–Hap8 |
| 2YD0–5AB0 | Hap2 | Original/Hap8 | ERAP1/2 closed conformation heterodimers, Hap2–N392, Hap8–N392 and Hap8–K392 |
| | N392 | Original/K392 | |

### 3.1.2 Optimization of existing structures

Even though selected structures were determined as the most eligible, these structures were still lacking some residues and bonds; therefore, they needed more optimization for accuracy. Heteroatoms and protein chains to include were selected, missing disulfide bonds were generated (404-443, 736-743 and hypothetical 486-498 for ERAP1 structures and 421-460, 504-514 and 759-766 for ERAP2 structures), some missing residues were added (excluding large gaps) and original structures were mutated into desired allotypes by using CHARMM-GUI web server (v3.1–3.7) (74). A few terminal sequences (935–940 amino acids in 3MDJ and 936–940 amino acids

in 3QNF) were added to ERAP1 open structures to match with 2YD0 ERAP1 closed structure. Bestatin ligand originally present in 2YD0 and 3MDJ structures was both kept for analyzing ligand–protein interaction and removed together with other heteroatoms (excluding Zn(II) ion) for analyzing the apo-protein. Ligand from 5AB0 structure was removed.

### 3.1.3 Comparative modeling

#### 3.1.3.1 Homology modeling

Missing residues that were still remaining after first optimization were added by Modeller (v9.24–9.25) homology modeling software (75). These included residues 111–114, 486–513 and 553–557 in ERAP1 2YD0 closed structure, residues 417–433, 486–514, 552–555, 864–867 and 893–906 in ERAP1 3MDJ open structure and residues 110–114, 427–431, 486–514, 551–558, 821–828, 856–871 and 892–908 in ERAP1 3QNF open structure. However, N-terminal membrane embedded residues between 1 and 45 amino acids were excluded due to their high flexibility and membrane positions. Open and closed target structures of ERAP1 were remodeled by using ERAP1 sequence alignment as a reference in order to *de novo* model missing loops with large gaps. Remodels with the lowest DOPE[1] scores were selected out of 50 models in the next step of the experiment as they are considered better in model assessment.

#### 3.1.3.2 Protein threading

As an alternative to homology modeling, I-TASSER server (v5.1) was run to model ERAP1/2 structures using *ab initio* protein modelling (76). Best 5 predicted structure models were compared in regards to their C-score[2].

[1] A quality score designed for selecting the best structure model.

[2] A confidence score quantitatively measured from each model to evaluate their quality.

## 3.2 Protein Docking

A molecular docking web server called HADDOCK (v2.4) saves computation time in exchange for receiving an experimental data of residues known to have a contact within the complex. Docking process involves different stages. Firstly, structure is energetically minimized in rigid-body state (it0). Afterwards, semi-flexible and explicit solvent refinements (it1 and itw) are run. At the end of each stage, best structures determined from HADDOCK scores are held. HADDOCK score combines a default weighted total of energies (such as van der Waals, electrostatic, desolvation and restraint violation) together with buried surface area (77).

At the beginning of the study, a possible binding site for ERAP1/2 dimers was not successfully determined. In the end, a reported APN homodimer structure (a protein of the same family) with a protein–protein interaction site was found and settled for possible dimerization site of ERAP1/2 proteins because of its homology (25). In accordance with this purpose, HADDOCK server was used to predict the binding poses of protein chains for ERAP1 homodimer and ERAP1/2 heterodimer structures (77). Residues to interact were selected which are located in a possible protein–protein interaction site for both ERAP1 and ERAP2 chains. This corresponds to all domain IV residues near and on surface (APPENDIX 1). In HADDOCK, these possible interacting residues were written in “active residues” section while other parameters were left in default. As distinct from rigid docking, not only fixed structures are rotated and moved around in a single stage but also determined parts of structures in “active residues” section are granted to be flexibly oriented in a later stage. Best complexes from the best cluster of each result was downloaded and compared according to their HADDOCK scores and common residue contacts.

## 3.3 Analysis of Structural Features

ERAP1/2 monomer structures were examined to validate their consistency with previous studies. APBS Electrostatics plugin in PyMOL was used to measure electrostatic potential surface of ERAP1/2 monomers (78). In PyMOL, “color_h.py”

script was run to measure hydrophobic surface of ERAP1/2 monomers (APPENDIX 2) (79). This script was readily downloaded and slightly changed for regular use (80). Secondary structure of hydrophobic transmembrane region and exon 10 loop region in ERAP1 was predicted by using Jpred (v4), PSIPRED (v4.0), RaptorX and also I-TASSER server (76,81–83). Glycosylation of the transmembrane region was predicted by NetNGlyc (v1.0) and NetOGlyc (v4.0) servers (84,85).

### 3.4 Energy Changes in Protein Folding and Interaction Energy

With the aim of calculating energy changes of protein structures, original ERAP1 crystal structure with the allotype Hap2 (2YD0) was mutated to produce Hap1–10 allotypes and original ERAP2 N392 allotype (5AB0) was mutated to produce K392 allotype by using FoldX software (v5.0). Effects of mutations on structures of ERAP1 and ERAP2 were investigated with the calculation of relative binding energy difference ($\Delta\Delta G$) in the same software. Values not between ±0.5 kcal/mol indicate large stabilization/destabilization effects. Additionally, interaction energies ($\Delta G_{binding}$) of ERAP1 homodimer and ERAP1/2 heterodimer were calculated (86). This energy was obtained by subtracting the sum of individual chain energies ($\Delta G_A + \Delta G_B$) from the energy of the complex ($\Delta G_{AB}$): $\Delta G_{binding} = \Delta G_{AB} - (\Delta G_A + \Delta G_B)$.

### 3.5 Simulation Configuration and Molecular Dynamics Simulations

In order for simulation of protein structures, a simulation box with a suitable environment was provided using CHARMM-GUI web server (v3.1–3.7) (74). First of all, each system was solvated with 10 angstroms (Å) thick TIP3P water and neutralized with 0.15 M KCl (87). Secondly, CHARMM36m force field (FF) was applied under periodic boundary conditions (PBCs). Afterwards, configuration files were downloaded for each system.

MD simulation was chosen as the computational method for analyzing motion of atoms in a given time. First of all, energy minimization was done. During MD simulations, several more steps were carried out. Temperature of each system was

gradually increased from 0 Kelvin (K) to human body temperature 310.15 K in heating step. Subsequently, equilibration step was run in NVT ensemble for 2 nanoseconds (ns). Finally, production step was performed in NPT ensemble for 100–150 ns depending on the system. NVT is characterized by constant number of particles (N), temperature (T) and volume (V) while NPT is characterized by constant N, T and pressure (P). As an electrostatics method, Particle Mesh Ewald (PME) was used. All MD simulations were carried out via NAMD software (v2.9) (88).

### 3.6 Trajectory Analysis

With the intent of understanding structural and functional impacts of homo- and heterodimerization; ERAP1/2 monomers, ERAP1 homodimer and ERAP1/2 heterodimer complexes were analyzed via MD simulations. All analyses were performed for 150 ns except all ERAP1 open conformation monomer structures (100 ns) and the third run of Hap8–N392 heterodimer (100 ns due to limited time). Analyses were done by using root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), principal component analysis (PCA), radius of gyration (RGYR), center of mass (CM) distances (in terms of chains, domains, residues and Zn(II) ion), angle theta, solvent accessible surface area (SASA) and weak bonding interactions (residues within contact distance, salt bridge and H-bond).

Analysis of RMSD was executed to analyze changes in average protein movement in the course of time. In Grcarma interface, “Fitting” task was used. Here, only “CA” (c-alpha) was checked in atom selection, all protein chains in chain selection and “All” in residue selection while other settings were left in default to perform RMSD (89). Analysis of RMSF was carried out to observe average movement of each residue individually in the course of time. A custom script called “rmsf.tcl” was executed in which all c-alpha atoms were selected and “measure rmsf” command from VMD was called (APPENDIX 3) (90). As an alternative approach to RMSF, PCA was performed to visualize each residue movement in a stability gradient manner. For selection, c-alpha atoms were defined. Analysis of RGYR was used to investigate protein folding and compactness through time. In Grcarma interface, “Radius of gyration” task was

used. In this panel; atom selection, chain selection and residue selection parameters were set as how they were done in RMSD analysis and other settings were left in default to run RGYR (87). In addition to these methods, CM distance between chains of dimers as well as CM distance between domain I and IV of monomers and dimers were calculated. In active site, alpha carbon (Cα) distance of Tyr438 to all GAMEN residues was calculated. Theta analysis was executed to describe basic structural motions. Here, theta (θ) was obtained from the interdomain angle defined with CMs of domain I-II, III and IV then converted into degrees by using a custom script (91) (APPENDIX 4). In SASA, a probe radius set at 1.4 Å was traced around the protein surface to calculate exposed surface area. This gives an idea on conformational dynamics as a function of energy difference. For SASA calculation, "sasa.tcl" script made by Falsafi and Karimi was used (APPENDIX 5) (92). As input, all protein atoms of a structure were used for this script which takes given selection and uses "measure sasa" command from VMD. For the evaluation of interchain interaction strength in dimers, intermolecular interactions were explored. Among these interactions, salt bridge and H-bond analysis were conducted. In salt bridge analysis, common residue pairs were also determined and grouped in sets. For both salt bridge and H-bond analysis, cutoff distances were set at 3 Å. Additionally, the cutoff distance of 3 Å was set in which any inter-residue contact within the range was determined.

Calculations of RMSD and RGYR were performed by using Grcarma software (89). On the other hand, RMSF, angle theta, CM distance, SASA, contacting residues, salt bridge and H-bond calculations were done via VMD custom commands (93). Lastly, PCA was employed in RStudio software through Bio3D package (94).

# 4 RESULTS

## 4.1 No Stable Secondary Structure Predicted in ER Retention Loop of ERAP1

Sequence alignment of ERAP1 and ERAP2 was run by Clustal Omega (v1.2.4) (95) server and rendered by Jalview (v2.11.2.2) (96) software (APPENDIX 6). Unstructured sequences were comprised of membrane-bound hydrophobic residues near N-terminus in both enzymes and residues located in exon 10 loop which contains ER retention signal in ERAP1. These sequences were not well-conserved between ERAP proteins (Figure 9).

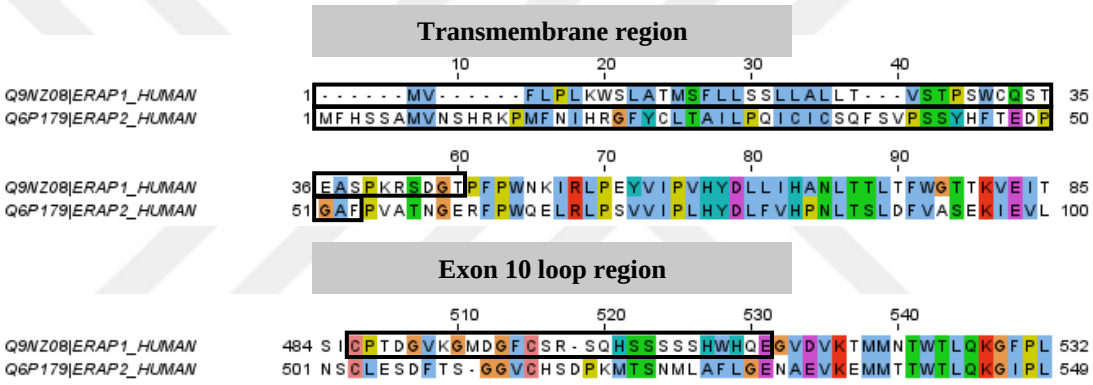



Figure 9. Sequence alignment results of transmembrane (top) and exon 10 loop (bottom) regions of ERAP1 and ERAP2. Amino acids with similar properties are colored. Black rectangle highlights the unstructured sequences in both enzymes. Full alignment results are available in APPENDIX 6.

Results from secondary structure prediction of multiple servers were obtained for the determination of structural features. All results agreed that there is an alpha helix structure forming between 10–24 amino acids in ERAP1 indicating an alpha-helical transmembrane domain around this region. Similar structure was also confirmed by I-TASSER results between 16–31 amino acids in ERAP2. However, sequences located at exon 10 encoded region did not show any stable secondary structure in ERAP1. In I-TASSER results, exon 10 loop region exhibited an alpha-helical structure with very low confidence scores only between unstructured 487–489 aa and 493–499 aa sequences (Figure 10).

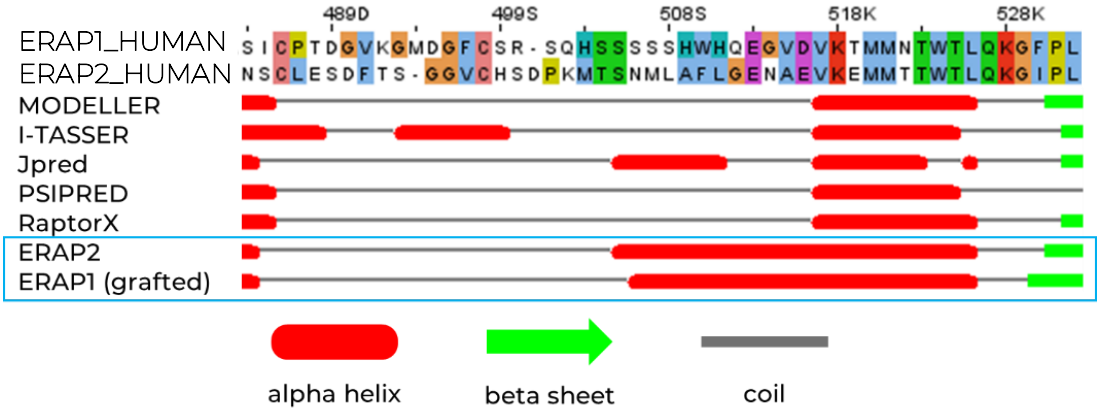



Figure 10. Secondary structure prediction results of exon 10 loop region in ERAP1. Sequences inside blue square show original ERAP2 (5AB0) and ERAP1 (with a grafted exon 10 loop structure from ERAP2 5AB0 by Papakyriakou *et al.* (51)).

Structural alignments were done by both Modeller software (75) and I-TASSER server (76). In Modeller, exon 10 loop was modeled in ERAP1 structure whereas, in I-TASSER, entire structure was modeled and models with best scores were listed for both ERAP proteins. Modeller modeled exon 10 region with no secondary structure in ERAP1 models while I-TASSER produced a very short alpha-helical structure around the sequence with repeated serine residues (505–508 amino acids) in all best 5 models of ERAP1. In transmembrane region of ERAP1 and ERAP2 models from protein threading, residues were positioned in a way that resembles an alpha-helical structure. However, they were not arranged similarly in 3D space. Following these results, ERAP1 remodeled protein structure with the best score in Modeller was selected for next step and ERAP2 protein structure was left unchanged.

For primary ERAP1 allotypes (Hap2, Hap3, Hap8 and Hap10), ERAP2 allotypes (N392 and K392), ERAP1 homodimer (Hap2–Hap2) and ERAP1/2 heterodimer (Hap2–N392), secondary structures were analyzed by STRIDE protein secondary structure assigment server (97). For every 10 ns in a simulation time, a structure file was created to run on STRIDE server in order to observe secondary structure changes in each simulation. Results at 50 ns intervals were included (APPENDIX 7).

## 4.2 Docking Results Show Acceptable Poses in Best Clusters

In docking prediction of ERAP1 homodimer and ERAP1/2 heterodimer structures, outer surface of domain IV including interacting residues of APN chains were selected for modelling ERAP dimer interactions (Figure 11).

Docking runs were performed for Hap2–Hap2, Hap2–N392, Hap8–Hap8, Hap8–K392 and Hap8–N392 before MD simulations. Many other docking runs (involving all ERAP1 homodimers with identical allotypes, N392 and K392 coupled ERAP1/2 heterodimers with all ERAP1 allotypes) were also done to calculate their interaction energy.

All docking results of best clusters had reasonable scores for structures that were chosen for dimer simulations because of large cluster size and binding poses of protein complexes in resemblance to APN homodimer (Figure 12, 13, 14, 15 and 16). Likewise, structure poses and common contacts were also very similar.

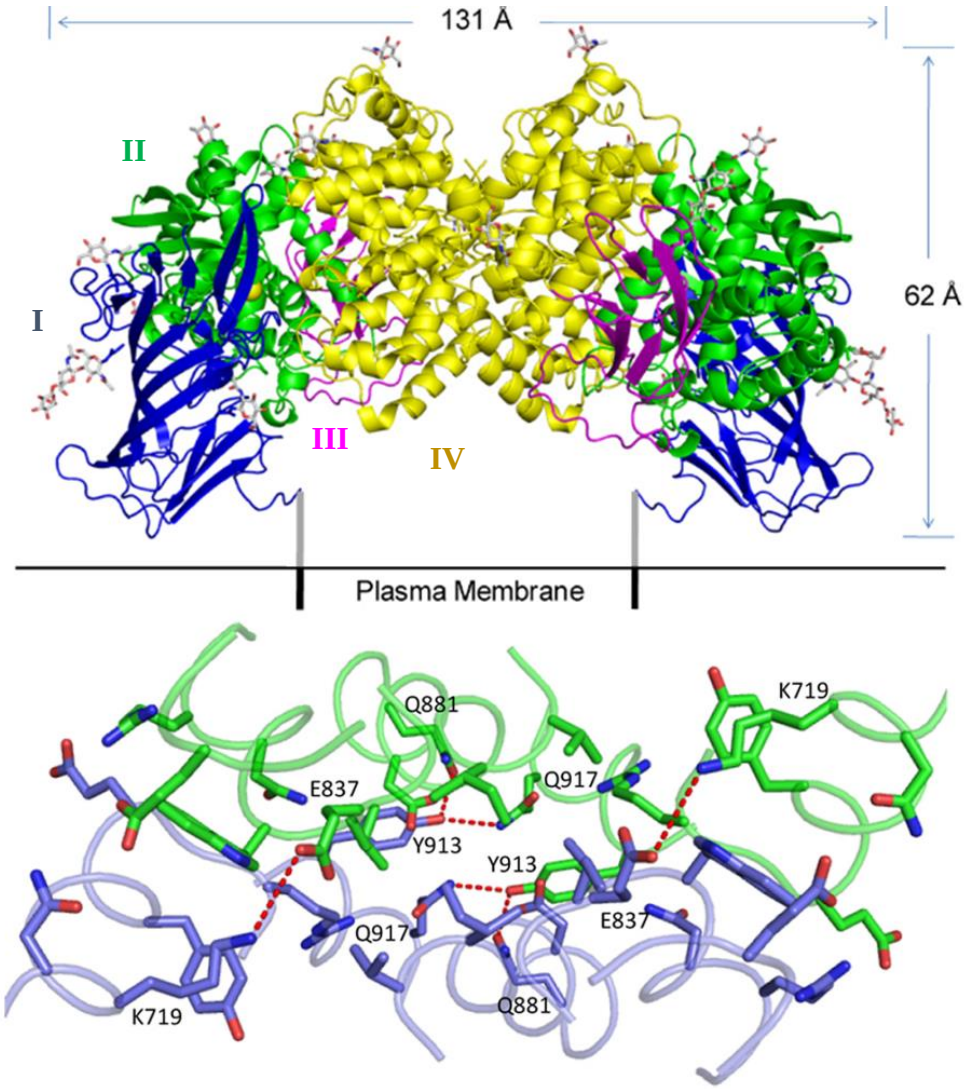
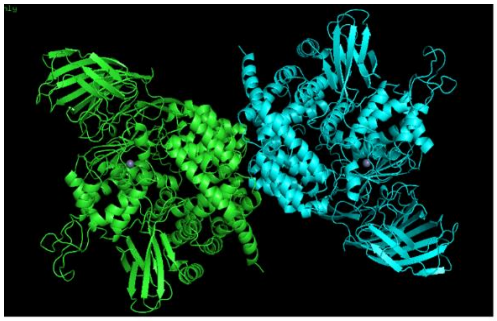



Figure 11. Structure of APN homodimer (top) and interacting residues in dimer interface (bottom). Domains were numbered from N-terminus to C-terminus of chains and colored accordingly. Interacting residues were labeled and colored by chain. Interacting atoms were connected with red dashed line. Figures modified and re-used from Figure 1 and 2 of Wong *et al.*, respectively (Creative Commons Attribution 4.0 International License) (25).

For calculation of energy changes in homo- and heterodimer structures by the help of interaction energy, one series of homodimers and two series of heterodimers were prepared and predicted by molecular docking (Table 3, 4 and 5). Docking scores were acceptable for every complex.



**Cluster 1**

| | |
|---|---|
| HADDOCK score | -148.0 +/- 5.5 |
| Cluster size | 128 |
| RMSD from the overall lowest-energy structure | 0.9 +/- 0.7 |
| Van der Waals energy | -65.4 +/- 10.9 |
| Electrostatic energy | -860.1 +/- 61.1 |
| Desolvation energy | 45.9 +/- 4.1 |
| Restraints violation energy | 435.3 +/- 30.5 |
| Buried Surface Area | 3524.8 +/- 198.7 |
| Z-Score | -1.9 |

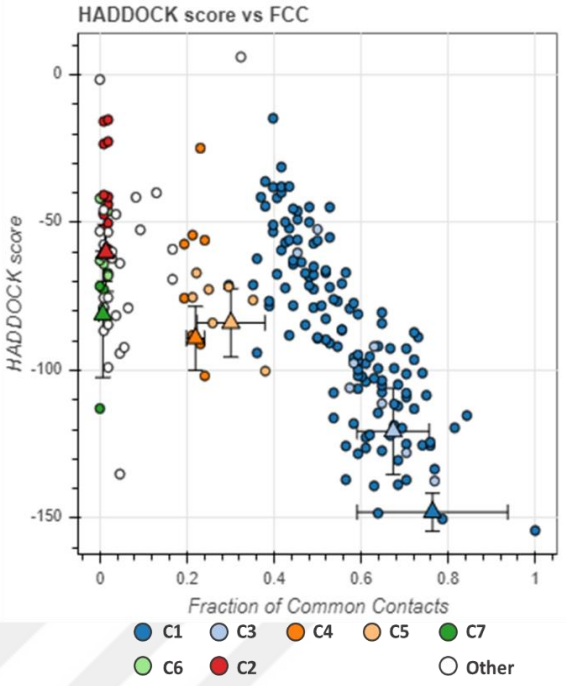



Figure 12. ERAP1 closed Hap2–Hap2 homodimer docking results. Docking results of the best cluster (bottom left), best structure from the cluster (top left), and common contacts of all clusters (right) (C: cluster, 1–7: cluster number).

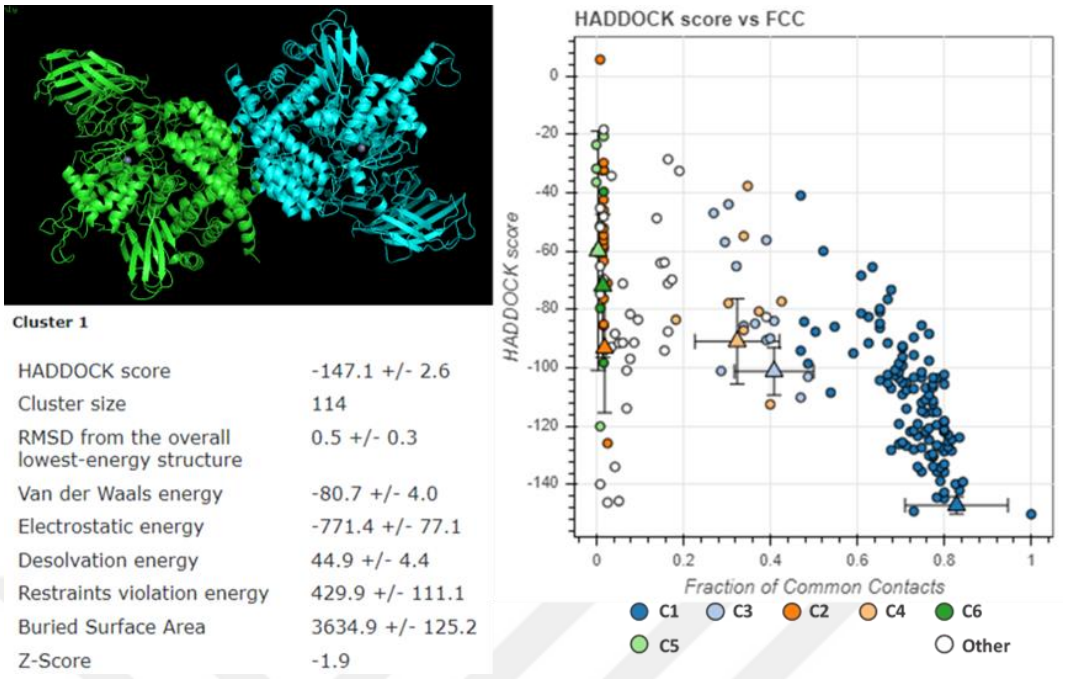



Figure 13. ERAP1/2 closed Hap2–N392 heterodimer docking results. Docking results of the best cluster (bottom left), best structure from the cluster (top left), and common contacts of all clusters (right) (C: cluster, 1–6: cluster number).

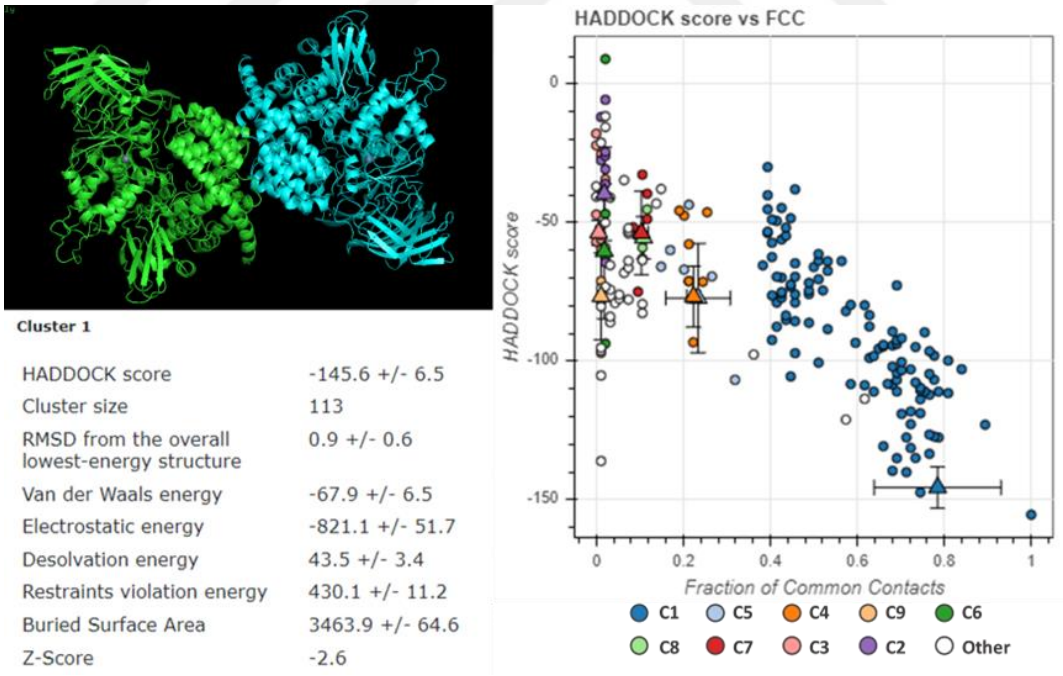
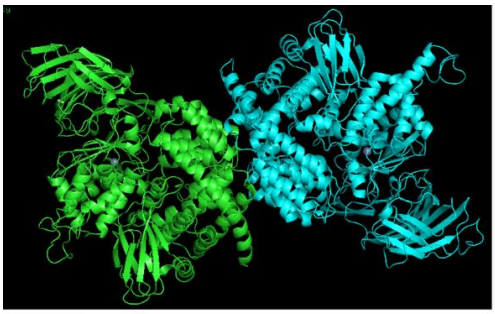



Figure 14. ERAP1 closed Hap8–Hap8 homodimer docking results. Docking results of the best cluster (bottom left), best structure from the cluster (top left), and common contacts of all clusters (right) (C: cluster, 1–9: cluster number).



Cluster 1

| | |
|---|---|
| HADDOCK score | -155.7 +/- 3.1 |
| Cluster size | 119 |
| RMSD from the overall lowest-energy structure | 0.4 +/- 0.2 |
| Van der Waals energy | -84.8 +/- 2.6 |
| Electrostatic energy | -803.9 +/- 39.2 |
| Desolvation energy | 45.6 +/- 2.3 |
| Restraints violation energy | 443.2 +/- 84.6 |
| Buried Surface Area | 3732.2 +/- 90.5 |
| Z-Score | -2.5 |

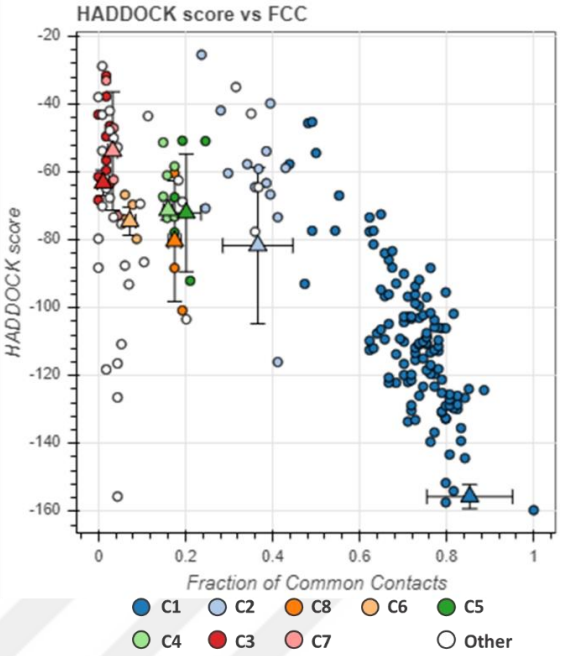
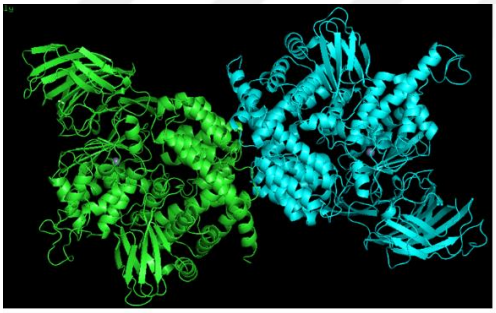



Figure 15. ERAP1/2 closed Hap8–K392 heterodimer docking results. Docking results of the best cluster (bottom left), best structure from the cluster (top left), and common contacts of all clusters (right) (C: cluster, 1–8: cluster number).



Cluster 1

| | |
|---|---|
| HADDOCK score | -161.4 +/- 6.8 |
| Cluster size | 137 |
| RMSD from the overall lowest-energy structure | 0.7 +/- 0.2 |
| Van der Waals energy | -87.6 +/- 8.2 |
| Electrostatic energy | -846.8 +/- 7.3 |
| Desolvation energy | 50.6 +/- 1.0 |
| Restraints violation energy | 450.2 +/- 44.3 |
| Buried Surface Area | 3888.7 +/- 36.3 |
| Z-Score | -2.2 |

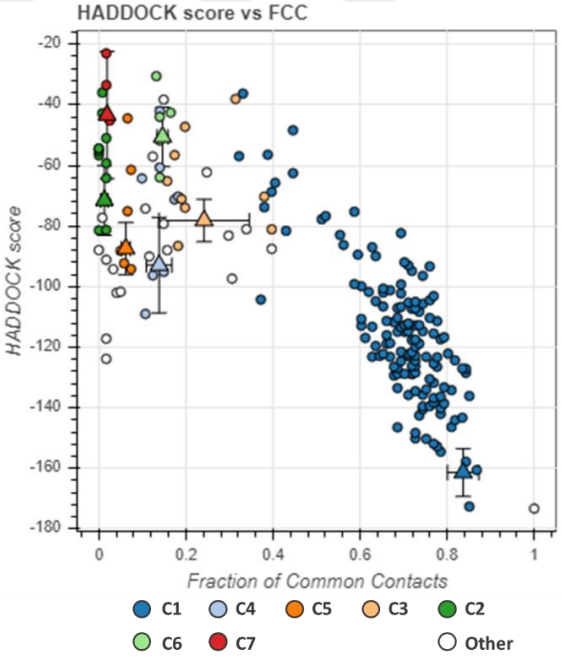



Figure 16. ERAP1/2 closed Hap8–N392 heterodimer docking results. Docking results of the best cluster (bottom left), best structure from the cluster (top left), and common contacts of all clusters (right) (C: cluster, 1–7: cluster number).

Table 3. Docking results of ERAP1 homodimer structures in preparation for energy changes calculation. Chain A (ChA) and B (ChB) are indicated. Reference structure was used for ERAP1 (PDB ID: 2YD0).

| **ChA-ChB** | **1-1** | **2-2** | **3-3** | **4-4** | **5-5** | **6-6** | **7-7** | **8-8** | **9-9** | **10-10** |
|---|---|---|---|---|---|---|---|---|---|---|
| Cluster no | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 |
| HADDOCK score | -157.9 ± 12.1 | -151.8 ± 2.1 | -141.2 ± 5.2 | -138.9 ± 5.3 | -141.1 ± 1.1 | -134.0 ± 7.4 | -144.7 ± 5.4 | -147.8 ± 8.0 | -143.3 ± 10.7 | -147.6 ± 4.9 |
| Cluster size | 130 | 122 | 117 | 91 | 109 | 86 | 27 | 27 | 105 | 127 |
| RMSD from the overall lowest-energy structure | 1.0 ± 0.7 | 1.2 ± 0.8 | 0.8 ± 0.3 | 0.9 ± 0.6 | 0.7 ± 0.5 | 0.5 ± 0.4 | 0.5 ± 0.3 | 0.9 ± 0.7 | 1.3 ± 0.7 | 0.8 ± 0.7 |
| Van der Waals energy | -70.0 ± 11.2 | -61.3 ± 7.0 | -69.3 ± 10.3 | -67.4 ± 7.3 | -58.3 ± 4.4 | -62.0 ± 13.4 | -66.7 ± 8.4 | -66.2 ± 6.0 | -68.4 ± 4.3 | -67.9 ± 9.4 |
| Electrostatic energy | -860.3 ±36.8 | -908.9 ± 39.6 | -822.2 ± 78.8 | -767.7 ± 55.5 | -855.0 ± 37.7 | -793.2 ± 124.7 | -847.0 ± 50.7 | -848.6 ± 74.3 | -796.4 ± 69.6 | -859.5 ± 82.5 |
| Desolvation energy | 44.0 ± 1.0 | 47.7 ± 2.9 | 48.3 ± 4.2 | 44.3 ± 2.1 | 48.4 ± 2.7 | 45.7 ± 7.0 | 43.1 ± 4.7 | 44.6 ± 5.3 | 43.2 ± 6.0 | 47.5 ± 4.9 |
| Restraints violation energy | 401.7 ± 82.8 | 436.1 ± 32.0 | 443.5 ± 29.6 | 378.0 ± 38.9 | 397.0 ± 26.4 | 409.4 ± 24.0 | 483.1 ± 32.1 | 434.6 ± 10.2 | 411.9 ± 47.8 | 447.2 ± 36.9 |
| Buried Surface Area | 3488.3 ± 149.2 | 3474.5 ± 70.3 | 3442.5 ± 119.3 | 3436.4 ± 164.9 | 3386.3 ± 85.7 | 3177.8 ± 101.4 | 3380.3 ± 123.5 | 3508.7 ± 35.8 | 3510.5 ± 89.8 | 3519.5 ± 112.5 |
| Z-Score | -1,6 | -1.9 | -1.5 | -2.2 | -2.3 | -1.8 | -2.1 | -2.3 | -1.5 | -2.3 |

Table 4. Docking results of N392 coupled ERAP1/2 heterodimer structures in preparation for energy changes calculation. Chain A (ChA) and B (ChB) are indicated. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

| **ChA-ChB** | **N-1** | **N-2** | **N-3** | **N-4** | **N-5** | **N-6** | **N-7** | **N-8** | **N-9** | **N-10** |
|---|---|---|---|---|---|---|---|---|---|---|
| Cluster no | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| HADDOCK score | -140.5 ± 4.0 | -141.1 ± 8.7 | -128.1 ± 3.3 | -134.2 ± 7.1 | -132.8 ± 4.1 | -127.4 ± 4.5 | -135.8 ± 4.9 | -148.4 ± 8.0 | -136.4 ± 2.0 | -135.5 ± 4.8 |
| Cluster size | 107 | 114 | 98 | 115 | 129 | 133 | 125 | 126 | 113 | 102 |
| RMSD from the overall lowest-energy structure | 0.5 ± 0.4 | 0.6 ± 0.4 | 32.5 ± 0.1 | 0.7 ± 0.4 | 0.6 ± 0.4 | 0.5 ± 0.3 | 0.4 ± 0.3 | 0.6 ± 0.3 | 0.8 ± 0.1 | 0.5 ± 0.3 |
| Van der Waals energy | -88.4 ± 0.4 | -86.1 ± 5.9 | -88.5 ± 5.7 | -91.1 ± 11.2 | -89.6 ± 3.6 | -86.4 ± 5.6 | -89.4 ± 8.0 | -91.1 ± 19.3 | -88.0 ± 5.7 | -99.2 ± 5.9 |
| Electrostatic energy | -707.2 ± 57.6 | -734.4 ± 50.2 | -659.4 ± 52.1 | -674.9 ± 41.8 | -710.8 ± 43.5 | -678.3 ± 29.3 | -728.5 ± 77.7 | -769.3 ± 74.1 | -704.9 ± 47.3 | -665.0 ± 57.0 |
| Desolvation energy | 45.9 ± 5.1 | 50.4 ± 5.0 | 47.5 ± 4.2 | 45.6 ± 7.1 | 51.0 ± 3.3 | 48.1 ± 3.7 | 54.8 ± 3.1 | 54.8 ± 1.6 | 46.7 ± 1.1 | 47.0 ± 5.6 |
| Restraints violation energy | 434.9 ± 76.8 | 414.7 ± 73.5 | 447.8 ± 26.2 | 462.3 ± 49.4 | 479.9 ± 38.4 | 465.8 ± 25.6 | 445.5 ± 44.6 | 418.4 ± 68.7 | 458.4 ± 60.8 | 496.6 ± 51.6 |
| Buried Surface Area | 3734.3 ± 204.4 | 3706.9 ± 207.3 | 3617.3 ± 212.6 | 3743.0 ± 158.8 | 3703.2 ± 91.7 | 3632.9 ± 125.3 | 3765.8 ± 173.3 | 3900.2 ± 36.6 | 3485.9 ± 122.5 | 3823.8 ± 130.1 |
| Z-Score | -2.6 | -1.9 | -2.4 | -2.2 | -1.9 | -2 | -2.1 | -2.2 | -2.2 | -2.4 |

Table 5. Docking results of K392 coupled ERAP1/2 heterodimer structures in preparation for energy changes calculation. Chain A (ChA) and B (ChB) are indicated. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

| **ChA-ChB** | **K-1** | **K-2** | **K-3** | **K-4** | **K-5** | **K-6** | **K-7** | **K-8** | **K-9** | **K-10** |
|---|---|---|---|---|---|---|---|---|---|---|
| Cluster no | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| HADDOCK score | -137.0 ± 9.3 | -128.5 ± 1.1 | -132.7 ± 5.8 | -133.7 ± 5.0 | -141.4 ± 8.3 | -133.4 ± 6.6 | -133.1 ± 7.9 | -142.0 ± 11.5 | -144.9 ± 1.6 | -143.2 ± 11.4 |
| Cluster size | 126 | 118 | 130 | 120 | 117 | 134 | 132 | 136 | 140 | 116 |
| RMSD from the overall lowest-energy structure | 0.5 ± 0.3 | 23.2 ± 0.2 | 0.5 ± 0.3 | 0.4 ± 0.3 | 0.6 ± 0.3 | 0.6 ± 0.4 | 0.7 ± 0.4 | 0.5 ± 0.3 | 0.4 ± 0.3 | 0.6 ± 0.3 |
| Van der Waals energy | -84.0 ± 6.6 | -83.7 ± 5.2 | -86.4 ± 13.1 | -86.1 ± 6.2 | -91.5 ± 10.5 | -96.6 ± 6.2 | -87.3 ± 4.6 | -88.7 ± 6.3 | -97.8 ± 7.5 | -86.3 ± 13.3 |
| Electrostatic energy | -709.4 ± 25.8 | -714.7 ± 35.2 | -733.2 ± 20.1 | -689.6 ± 49.5 | -708.0 ± 86.2 | -644.0 ± 48.2 | -703.2 ± 66.2 | -759.2 ± 38.1 | -693.8 ± 55.4 | -778.5 ± 41.5 |
| Desolvation energy | 46.9 ± 2.3 | 52.0 ± 4.2 | 46.9 ± 4.1 | 47.0 ± 5.2 | 47.1 ± 3.5 | 46.7 ± 3.3 | 49.0 ± 6.6 | 54.9 ± 6.5 | 46.0 ± 4.1 | 51.7 ± 3.6 |
| Restraints violation energy | 420.5 ± 104.6 | 460.7 ± 64.0 | 534.5 ± 27.5 | 432.8 ± 82.3 | 446.3 ± 69.2 | 453.0 ± 63.5 | 459.4 ± 26.1 | 435.8 ± 34.1 | 456.5 ± 34.2 | 470.3 ± 32.7 |
| Buried Surface Area | 3692.2 ± 47.7 | 3690.6 ± 92.2 | 3703.6 ± 159.6 | 3611.6 ± 138.1 | 3776.3 ± 211.1 | 3734.2 ± 87.8 | 3625.1 ± 138.7 | 3817.5 ± 120.2 | 3706.5 ± 60.8 | 3795.7 ± 241.3 |
| Z-Score | -1.7 | -1.6 | -2 | -2 | -2.1 | -1.7 | -1.8 | -1.6 | -2.1 | -2.3 |

### 4.3 Stability of ERAP Structures Depends on Allotypic Differences

Stability of ERAP1 and ERAP2 closed monomers as a result of single mutations in remodeled structures were calculated and rounded to three decimal places (Table 6). Remodeled Hap2 structure (PDB ID: 2YD0) was mutated to nine other ERAP1 allotypes. Likewise, remodeled N392 structure (PDB ID: 5AB0) was mutated to ERAP2 K392 allotype. In total energies, only Hap2→Hap10 mutation caused a

significant increase that is more than 0.5 kcal/mol. Mutations to Hap1, Hap8 and Hap9 were inversely significant. Mutation from ERAP2 N392 to K392 caused a negative significance as well.

Table 6. Mutational energy changes in ERAP1 and ERAP2 closed monomers. Wild type (WT) was mutated to a new (Mut) allotype indicated with an arrow. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

| **WT→Mut** | **2→1** | **2→3** | **2→4** | **2→5** | **2→6** | **2→7** | **2→8** | **2→9** | **2→10** | **N→K** |
|---|---|---|---|---|---|---|---|---|---|---|
| Total Energy | -1.285 | 0.197 | 0.335 | 0.169 | -0.495 | -0.408 | -1.263 | -1.472 | 2.267 | -1.450 |
| Backbone H-bond | 0 | 0.575 | 1.491 | 1.287 | 1.396 | 2.403 | 0.470 | 0.654 | 3.021 | 0.005 |
| Sidechain H-bond | 0 | 0.527 | -0.027 | 1.261 | 1.255 | 2.233 | 2.189 | 2.364 | -0.948 | -1.128 |
| Van der Waals | 0.028 | 0.313 | -0.044 | -0.115 | 0.260 | 0.578 | -0.454 | -0.276 | 1.440 | -0.943 |
| Electrostatics | 0.215 | -0.439 | -0.554 | -0.340 | -0.266 | -0.347 | -0.098 | -0.014 | 2.765 | -2.117 |
| Solvation Polar | -0.447 | -0.174 | -0.008 | -0.408 | -1.180 | -2.066 | -0.994 | -1.314 | -2.153 | 2.577 |
| Solvation Hydrophobic | -0.154 | 0.377 | 0.431 | 0.086 | 0.424 | 0.697 | -0.570 | -0.454 | 2.371 | -1.701 |
| Van der Waals Clashes | 0.008 | 0.016 | 0.099 | 0.074 | -0.060 | -0.086 | 0.270 | 0.180 | -0.362 | 1.117 |
| Entropy Sidechain | -0.213 | -0.718 | -0.678 | -0.549 | -1.106 | -2.069 | -1.068 | -1.519 | -1.780 | 1.306 |
| Entropy Mainchain | -0.529 | -0.068 | -0.287 | -0.576 | -0.859 | -1.422 | -0.776 | -0.609 | -1.197 | -0.737 |
| Torsional Clash | -0.183 | -0.047 | 0.086 | -0.006 | -0.188 | -0.160 | -0.096 | -0.380 | -0.150 | 0.196 |
| Backbone Clash | -0.894 | -0.034 | -0.143 | 0.170 | -1.008 | -1.068 | -1.257 | -1.379 | -0.851 | -0.007 |
| Helix Dipole | -0.016 | -0.164 | -0.171 | -0.541 | -0.170 | -0.163 | -0.132 | -0.101 | -0.197 | -0.050 |
| Energy Ionisation | 0.005 | 0 | -0.003 | -0.005 | 0.001 | -0.005 | -0.003 | -0.003 | -0.542 | 0.025 |

Interaction energies of two identical chains in ERAP1 closed homodimer structures were calculated after best complexes were obtained from molecular docking (Table 7). Results showed that interaction energies of Hap1–Hap1, Hap5–Hap5 and Hap9–Hap9 homodimer structures were in negative and Hap2–Hap2 had the highest interaction energy followed by Hap8–Hap8.

Table 7. Interaction energy of ERAP1 homodimer structures. Chain A (ChA) and B (ChB) are indicated. Reference structure was used for ERAP1 (PDB ID: 2YD0).

| ChA-ChB | 1-1 | 2-2 | 3-3 | 4-4 | 5-5 | 6-6 | 7-7 | 8-8 | 9-9 | 10-10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Intraclashes Chain A | 223.76 | 229.48 | 223.71 | 229.62 | 222.19 | 219.86 | 224.12 | 222.49 | 219.05 | 212.97 |
| Intraclashes Chain B | 222.33 | 222.87 | 225.05 | 219.34 | 215.93 | 218.47 | 216.73 | 221.42 | 221.91 | 212.35 |
| Interaction Energy | -1.236 | 6.709 | 1.566 | 2.736 | -2.686 | 1.112 | 1.274 | 4.490 | -2.611 | 2.312 |
| Stability Chain A | 521.56 | 567.41 | 557.88 | 533.51 | 516.93 | 524.48 | 534.73 | 522.33 | 521.21 | 520.12 |
| Stability Chain B | 551.74 | 578.55 | 537.88 | 521.82 | 512.72 | 542.55 | 537.85 | 524.57 | 540.95 | 525.50 |

In order to evaluate dimer interaction, both N392 and K392 coupled ERAP1/2 heterodimer structures were calculated in terms of interaction energy. Structure of N392–Hap8 heterodimer showed the highest energy value (Table 8). Energy of N392–Hap10 was in the negative range close to margin of error. Heterodimer structures paired with K392 demonstrated more changes based on ERAP1 allotype difference

Table 8. Interaction energy of N392 coupled ERAP1/2 heterodimer structures. Chain A (ChA) and B (ChB) are indicated. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

| ChA-ChB | N-1 | N-2 | N-3 | N-4 | N-5 | N-6 | N-7 | N-8 | N-9 | N-10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Intraclashes Chain A | 202.60 | 224.07 | 221.96 | 216.21 | 222.39 | 208.67 | 228.64 | 203.28 | 221.52 | 227.83 |
| Intraclashes Chain B | 211.08 | 219.75 | 220.64 | 215.73 | 223.23 | 200.04 | 225.40 | 219.88 | 220.93 | 220.44 |
| Interaction Energy | 3.093 | 4.240 | 4.875 | 10.697 | 3.884 | 8.084 | 9.345 | 12.007 | 6.377 | -0.575 |
| Stability Chain A | 384.36 | 446.37 | 439.68 | 401.46 | 441.54 | 394.47 | 436.06 | 388.01 | 425.30 | 430.17 |
| Stability Chain B | 202.60 | 224.07 | 221.96 | 216.21 | 222.39 | 208.67 | 228.64 | 203.28 | 221.52 | 227.83 |

(Table 9). Here, K392–Hap2 heterodimer got the highest value of all heterodimer structures.

Table 9. Interaction energy of K392 coupled ERAP1/2 heterodimer structures. Chain A (ChA) and B (ChB) are indicated. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

| ChA-ChB | K-1 | K-2 | K-3 | K-4 | K-5 | K-6 | K-7 | K-8 | K-9 | K-10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Intraclashes Chain A | 228.70 | 230.57 | 225.65 | 226.68 | 228.40 | 228.94 | 228.09 | 228.55 | 227.17 | 228.93 |
| Intraclashes Chain B | 226.79 | 229.35 | 232.35 | 238.24 | 230.61 | 231.73 | 241.69 | 235.85 | 231.63 | 232.86 |
| Interaction Energy | 9.998 | 13.455 | 11.947 | -0.409 | 10.797 | 5.589 | 12.482 | 2.678 | 6.151 | 8.413 |
| Stability Chain A | 438.51 | 449.48 | 416.13 | 445.01 | 431.75 | 432.02 | 463.32 | 436.59 | 436.25 | 455.19 |
| Stability Chain B | 571.02 | 570.37 | 589.36 | 582.46 | 590.75 | 581.36 | 572.56 | 543.95 | 584.20 | 600.66 |

### 4.4 Better Binding Affinity in Heterodimers with Domain IV Binding Mode

The proposed dimerization site from a recent study (51) (exon 10 loop) and the dimerization site that was used in this thesis study (domain IV) were selected in HADDOCK software to observe their binding affinity. Exon 10 loops from the best 5 models of ERAP1 obtained from I-TASSER results were used and grafted on remodeled ERAP1 2YD0 structure from Modeller to predict the dimerization site. ERAP2 5AB0 chain and ERAP1 remodeled chain were added as the first and the second chains, respectively, for heterodimer docking.

In docking results, heterodimer complexes formed by ERAP2 and ERAP1 at domain IV region showed a similar binding mode in best cluster of all predictions (Figure 17, 18, 19, 20 and 21). When domain IV and exon 10 loop regions were compared in heterodimer docking results, all domain IV predictions outran their exon 10 loop counterpart in terms of docking score and cluster size (Figure 22, 23, 24, 25 and 26). Among all heterodimer predictions, only ERAP2 and ERAP1 with exon 10

loop from model 2 adopted a similar binding mode achieved in the aforementioned study (51). However, this binding mode exhibited the lowest docking score and cluster size in the group.

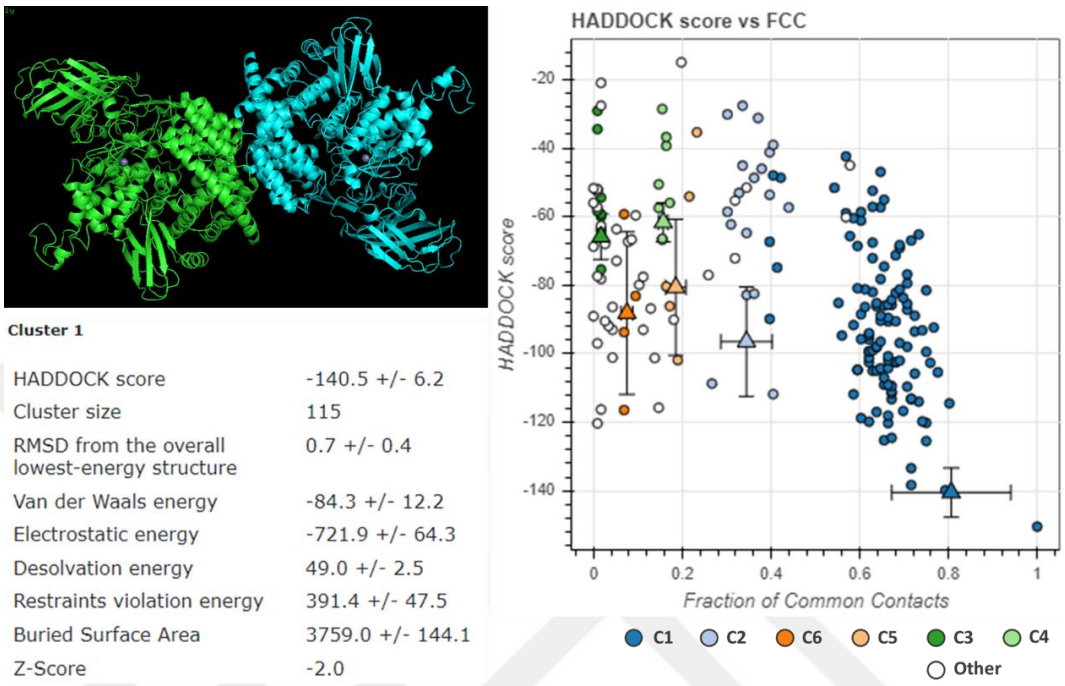
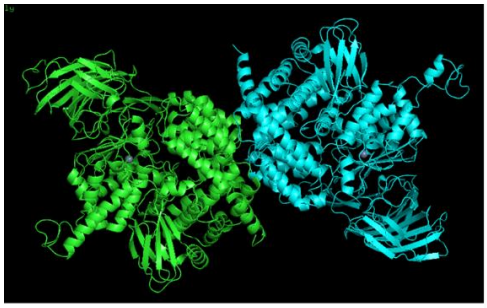



Figure 17. Docking results of ERAP2 and ERAP1 model 1 heterodimer targeted at domain IV region. Docking results of the best cluster (bottom left), best structure from the cluster (top left), and common contacts of all clusters (right) (C: cluster, 1–6: cluster number).



**Cluster 1**

| | |
|---|---|
| HADDOCK score | -130.8 +/- 5.4 |
| Cluster size | 125 |
| RMSD from the overall lowest-energy structure | 6.6 +/- 0.1 |
| Van der Waals energy | -93.2 +/- 7.9 |
| Electrostatic energy | -680.6 +/- 76.7 |
| Desolvation energy | 47.7 +/- 4.6 |
| Restraints violation energy | 508.2 +/- 59.4 |
| Buried Surface Area | 3653.6 +/- 127.4 |
| Z-Score | -1.8 |

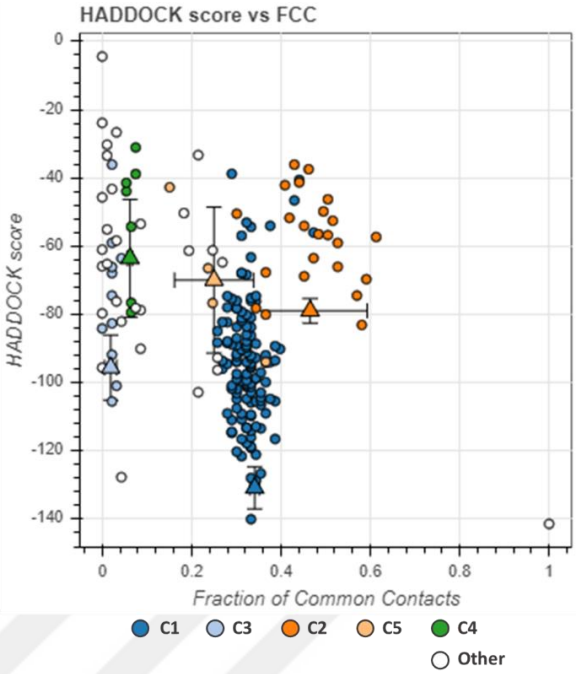
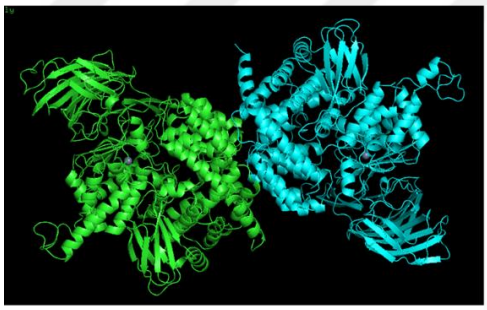



Figure 18. Docking results of ERAP2 and ERAP1 model 2 heterodimer targeted at domain IV region. Docking results of the best cluster (bottom left), best structure from the cluster (top left), and common contacts of all clusters (right) (C: cluster, 1–5: cluster number).



**Cluster 1**

| | |
|---|---|
| HADDOCK score | -141.1 +/- 2.2 |
| Cluster size | 120 |
| RMSD from the overall lowest-energy structure | 25.7 +/- 0.2 |
| Van der Waals energy | -92.5 +/- 6.3 |
| Electrostatic energy | -717.8 +/- 31.1 |
| Desolvation energy | 49.5 +/- 3.8 |
| Restraints violation energy | 455.1 +/- 35.3 |
| Buried Surface Area | 3748.9 +/- 147.8 |
| Z-Score | -2.5 |

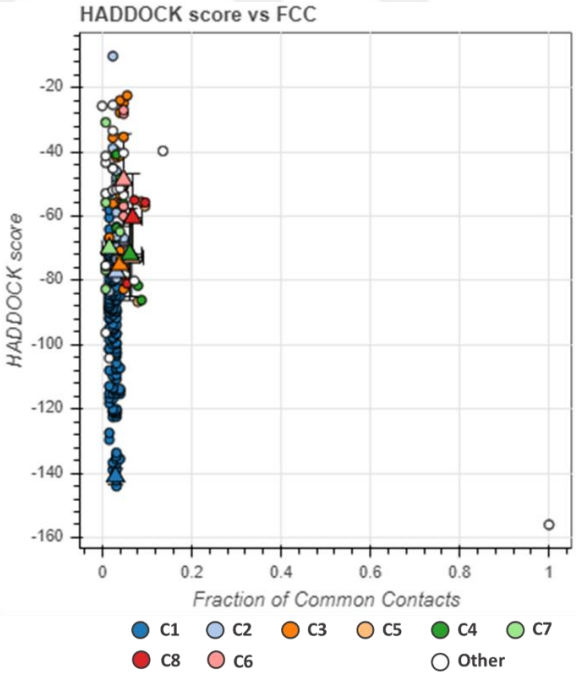
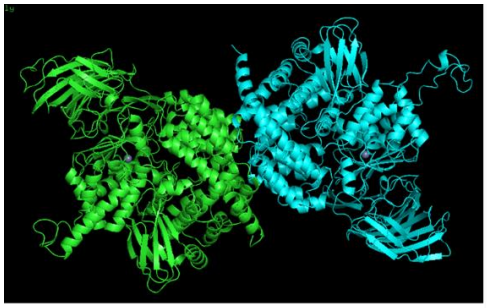



Figure 19. Docking results of ERAP2 and ERAP1 model 3 heterodimer targeted at domain IV region. Docking results of the best cluster (bottom left), best structure from the cluster (top left), and common contacts of all clusters (right) (C: cluster, 1–8: cluster number).



**Cluster 1**

| | |
|---|---|
| HADDOCK score | -149.3 +/- 10.3 |
| Cluster size | 125 |
| RMSD from the overall lowest-energy structure | 0.4 +/- 0.2 |
| Van der Waals energy | -88.9 +/- 7.7 |
| Electrostatic energy | -770.0 +/- 46.5 |
| Desolvation energy | 52.9 +/- 4.1 |
| Restraints violation energy | 407.4 +/- 53.7 |
| Buried Surface Area | 3753.4 +/- 166.7 |
| Z-Score | -1.8 |

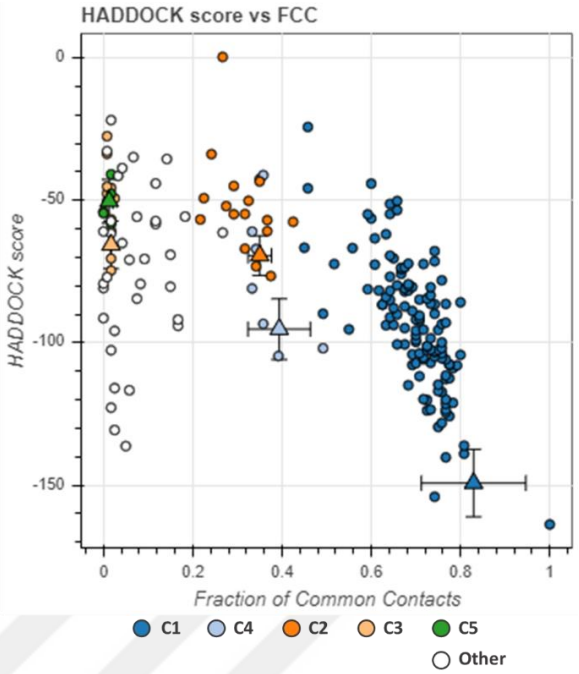
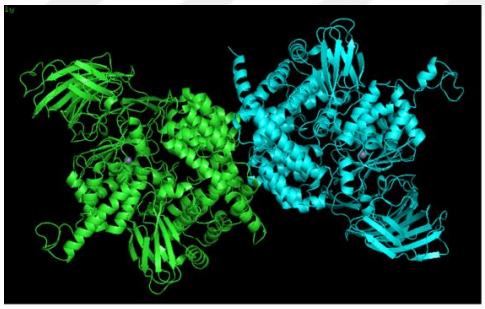



Figure 20. Docking results of ERAP2 and ERAP1 model 4 heterodimer targeted at domain IV region. Docking results of the best cluster (bottom left), best structure from the cluster (top left), and common contacts of all clusters (right) (C: cluster, 1–5: cluster number).



**Cluster 1**

| | |
|---|---|
| HADDOCK score | -149.1 +/- 5.3 |
| Cluster size | 127 |
| RMSD from the overall lowest-energy structure | 0.4 +/- 0.2 |
| Van der Waals energy | -89.6 +/- 10.6 |
| Electrostatic energy | -799.4 +/- 37.3 |
| Desolvation energy | 53.4 +/- 4.8 |
| Restraints violation energy | 470.2 +/- 72.4 |
| Buried Surface Area | 3823.2 +/- 136.5 |
| Z-Score | -2.1 |

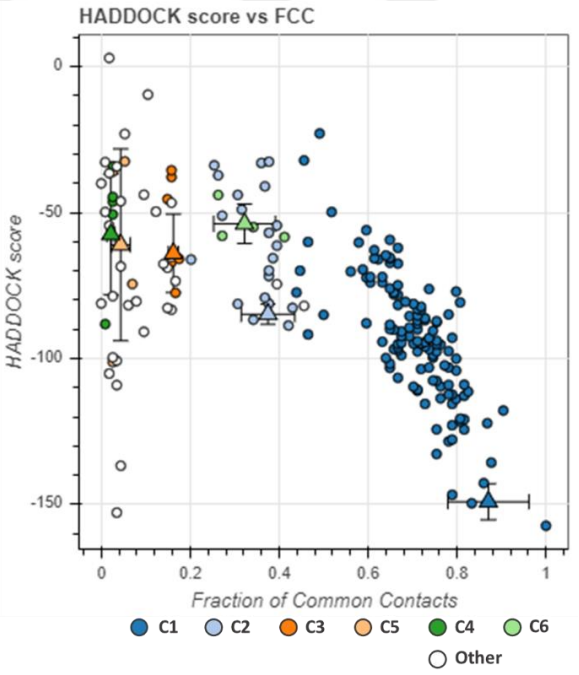
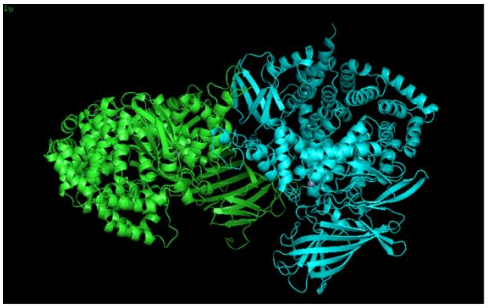



Figure 21. Docking results of ERAP2 and ERAP1 model 5 heterodimer targeted at domain IV region. Docking results of the best cluster (bottom left), best structure from the cluster (top left), and common contacts of all clusters (right) (C: cluster, 1–6: cluster number).



**Cluster 2**

| | |
|---|---|
| HADDOCK score | -135.3 +/- 4.0 |
| Cluster size | 31 |
| RMSD from the overall lowest-energy structure | 2.4 +/- 1.3 |
| Van der Waals energy | -102.7 +/- 12.5 |
| Electrostatic energy | -500.8 +/- 57.0 |
| Desolvation energy | -11.4 +/- 8.0 |
| Restraints violation energy | 789.8 +/- 113.0 |
| Buried Surface Area | 3909.3 +/- 248.1 |
| Z-Score | -1.8 |

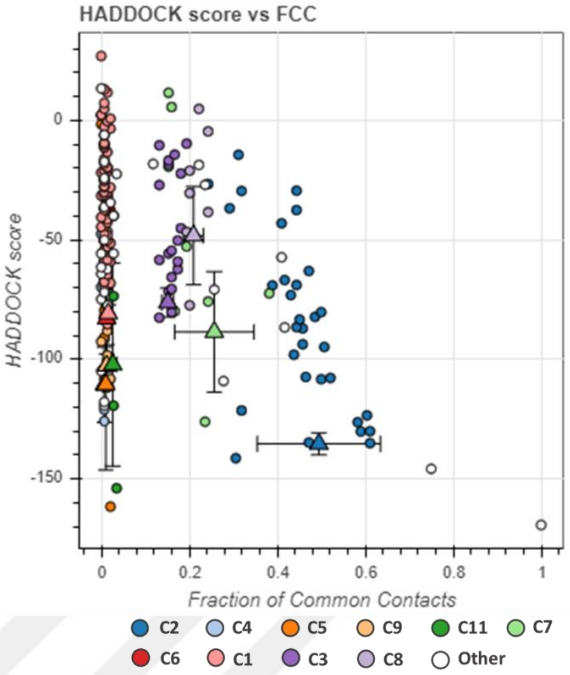
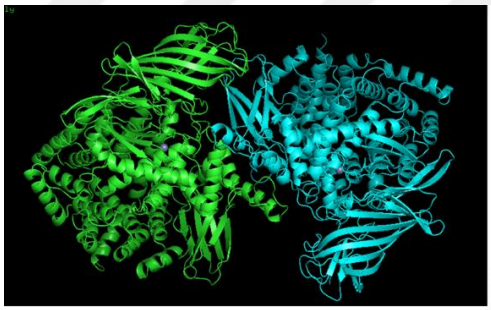



Figure 22. Docking results of ERAP2 and ERAP1 model 1 heterodimer targeted at exon 10 loop region. Docking results of the best cluster (bottom left), best structure from the cluster (top left), and common contacts of all clusters (right) (C: cluster, 1–11: cluster number).



**Cluster 7**

| | |
|---|---|
| HADDOCK score | -119.7 +/- 10.5 |
| Cluster size | 8 |
| RMSD from the overall lowest-energy structure | 5.2 +/- 0.3 |
| Van der Waals energy | -86.6 +/- 8.1 |
| Electrostatic energy | -494.1 +/- 34.7 |
| Desolvation energy | 2.9 +/- 4.7 |
| Restraints violation energy | 628.4 +/- 52.2 |
| Buried Surface Area | 3499.2 +/- 240.2 |
| Z-Score | -1.6 |

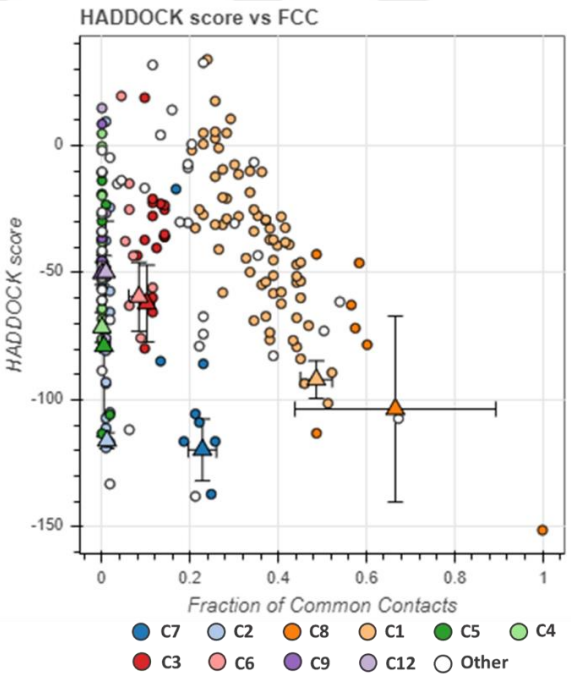
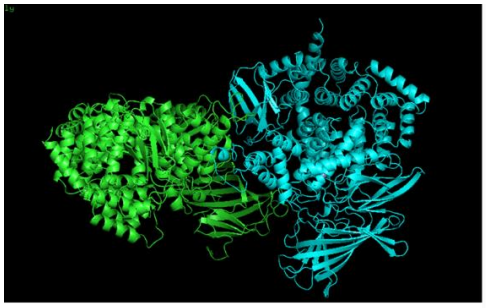



Figure 23. Docking results of ERAP2 and ERAP1 model 2 heterodimer targeted at exon 10 loop region. Docking results of the best cluster (bottom left), best structure from the cluster (top left), and common contacts of all clusters (right) (C: cluster, 1–12: cluster number).



**Cluster 2**

| | |
|---|---|
| HADDOCK score | -123.7 +/- 6.9 |
| Cluster size | 40 |
| RMSD from the overall lowest-energy structure | 3.9 +/- 0.3 |
| Van der Waals energy | -108.8 +/- 5.8 |
| Electrostatic energy | -422.7 +/- 38.8 |
| Desolvation energy | -23.6 +/- 5.5 |
| Restraints violation energy | 932.1 +/- 60.2 |
| Buried Surface Area | 3737.2 +/- 168.6 |
| Z-Score | -1.4 |

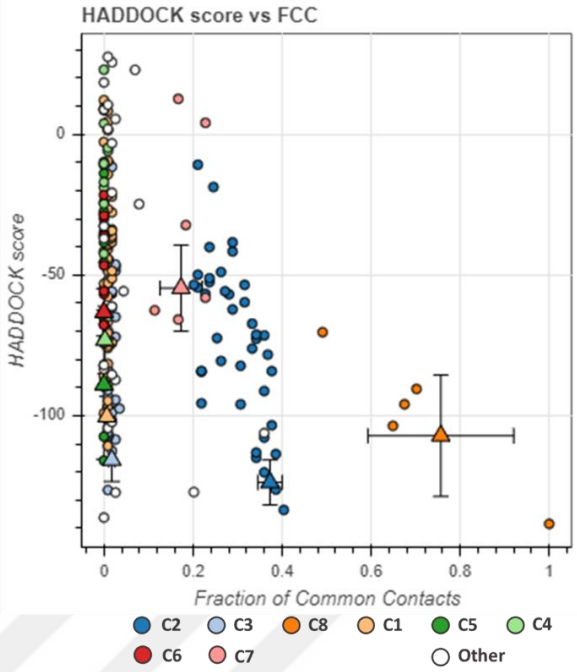
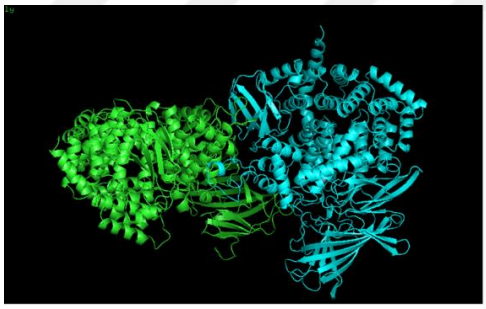



Figure 24. Docking results of ERAP2 and ERAP1 model 3 heterodimer targeted at exon 10 loop region. Docking results of the best cluster (bottom left), best structure from the cluster (top left), and common contacts of all clusters (right) (C: cluster, 1–8: cluster number).



**Cluster 1**

| | |
|---|---|
| HADDOCK score | -129.1 +/- 6.1 |
| Cluster size | 47 |
| RMSD from the overall lowest-energy structure | 0.9 +/- 0.5 |
| Van der Waals energy | -99.5 +/- 5.5 |
| Electrostatic energy | -491.6 +/- 29.4 |
| Desolvation energy | -15.0 +/- 5.7 |
| Restraints violation energy | 836.1 +/- 140.0 |
| Buried Surface Area | 4078.8 +/- 138.9 |
| Z-Score | -2.4 |

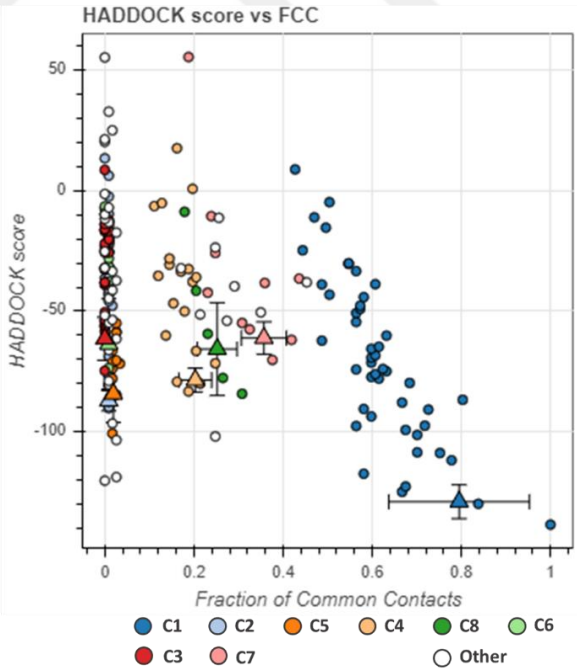
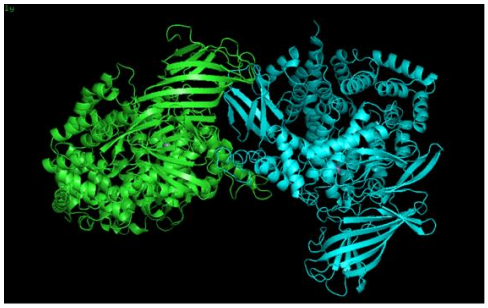



Figure 25. Docking results of ERAP2 and ERAP1 model 4 heterodimer targeted at exon 10 loop region. Docking results of the best cluster (bottom left), best structure from the cluster (top left), and common contacts of all clusters (right) (C: cluster, 1–8: cluster number).



**Cluster 4**

| | |
|---|---|
| HADDOCK score | -130.4 +/- 9.8 |
| Cluster size | 19 |
| RMSD from the overall lowest-energy structure | 32.3 +/- 0.0 |
| Van der Waals energy | -112.9 +/- 10.6 |
| Electrostatic energy | -412.6 +/- 55.0 |
| Desolvation energy | -21.0 +/- 2.8 |
| Restraints violation energy | 859.9 +/- 80.0 |
| Buried Surface Area | 3858.4 +/- 79.9 |
| Z-Score | -1.8 |

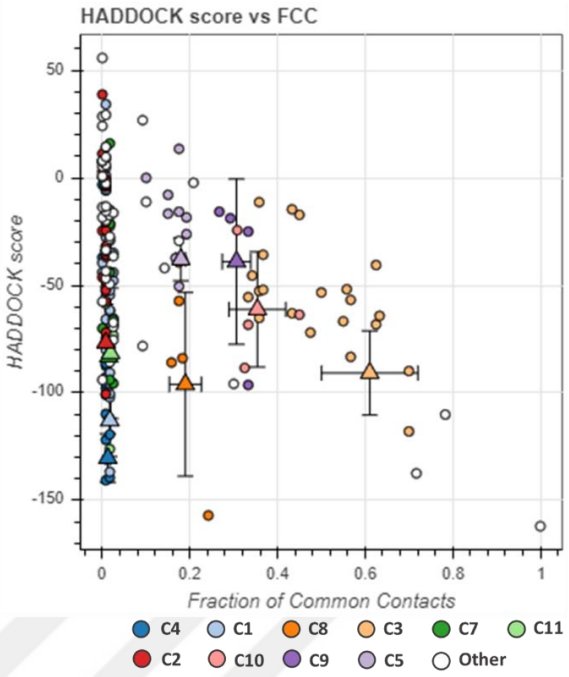



Figure 26. Docking results of ERAP2 and ERAP1 model 5 heterodimer targeted at exon 10 loop region. Docking results of the best cluster (bottom left), best structure from the cluster (top left), and common contacts of all clusters (right) (C: cluster, 1–11: cluster number).

## 4.5 Structural Changes Observed in ERAP1 and ERAP2 Simulations

After completion of simulation runs, snapshots of simulations were taken in 50 ns intervals throughout the simulation. Structures from snapshots of ERAP1 monomers, ERAP2 monomers, ERAP1 homodimers and ERAP1/2 heterodimers were superimposed for structural comparison. For the sake of convenience; Hap2, Hap3, Hap8 and Hap10 were selected for further analysis in this study and grouped as primary ERAP1 allotypes. Allotypes Hap1, Hap4, Hap5, Hap6, Hap7 and Hap9 were categorized as secondary ERAP1 allotypes and largely discarded.

In this study, three parallel runs for each primary ERAP1 allotype was performed. Structures of ERAP1 closed Hap2, Hap3, Hap8 and Hap10 monomers were mostly stable in three different runs for 150 ns (Figure 27). Nonetheless, some motions around domain I and IV and on helix H4a of domain II were observed as seen in Hap2 (run 1 at 150 ns & run 2 at 100 ns), Hap3 (run 1 at 150 ns, run 2 at 100 and 150 ns & run 3 at

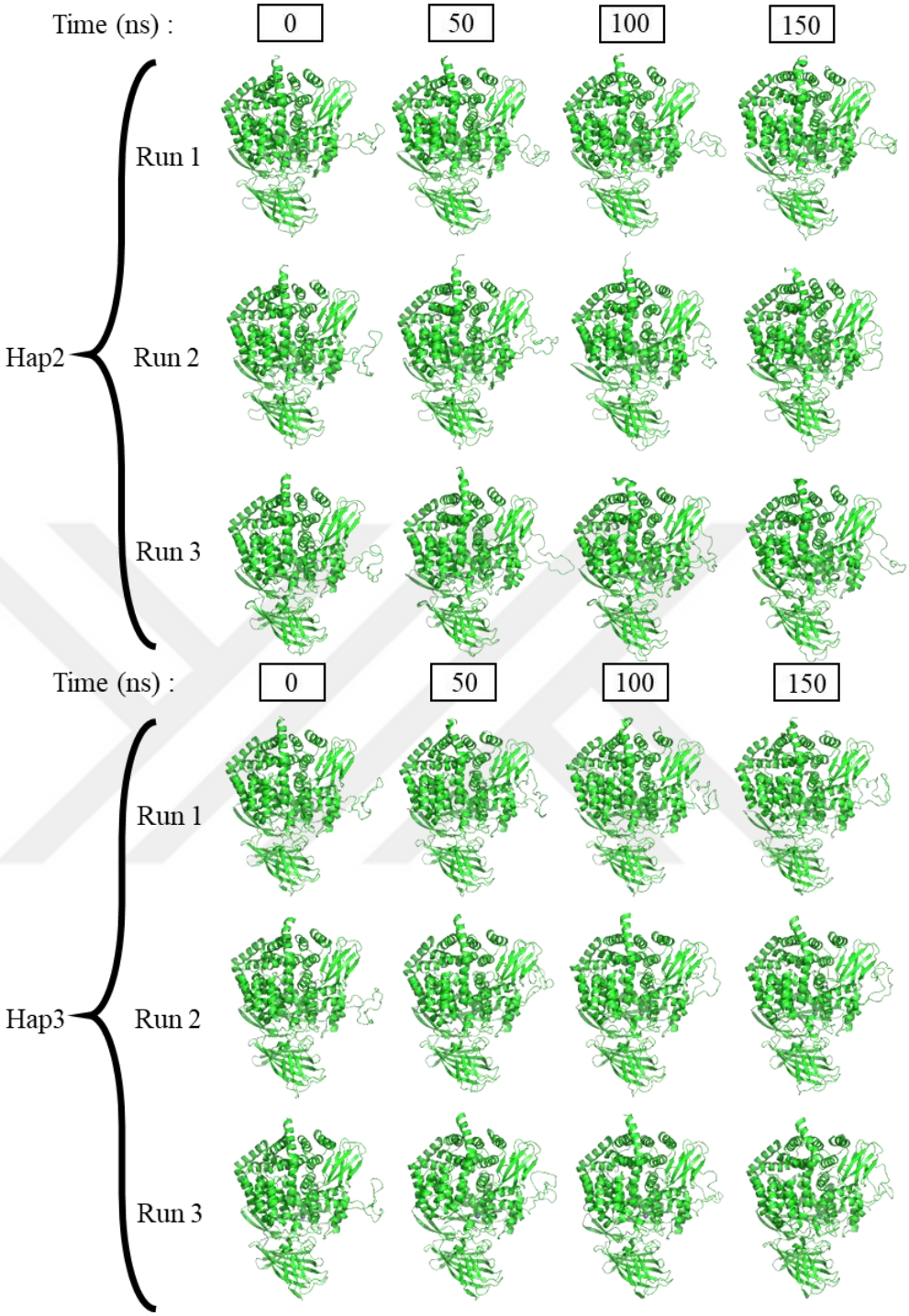



Figure 27. Primary allotypes in ERAP1 closed monomer 3D structures before simulation, at 50, 100 and 150 ns of simulation. The whole ERAP1 structure is represented with green cartoon (PDB ID: 2YD0).

50 ns), Hap8 (run 1 at 100 ns, run 2 in all & run 3 at 100 and 150 ns) and Hap10 (run 1 at 50 and 150 ns, run 2 in all & run 3 at 50 and 100 ns) as well as Zn(II) ion shifting in Hap2 (run 3 at 150 ns). Sequence 485–508 was the most flexible region in all runs.

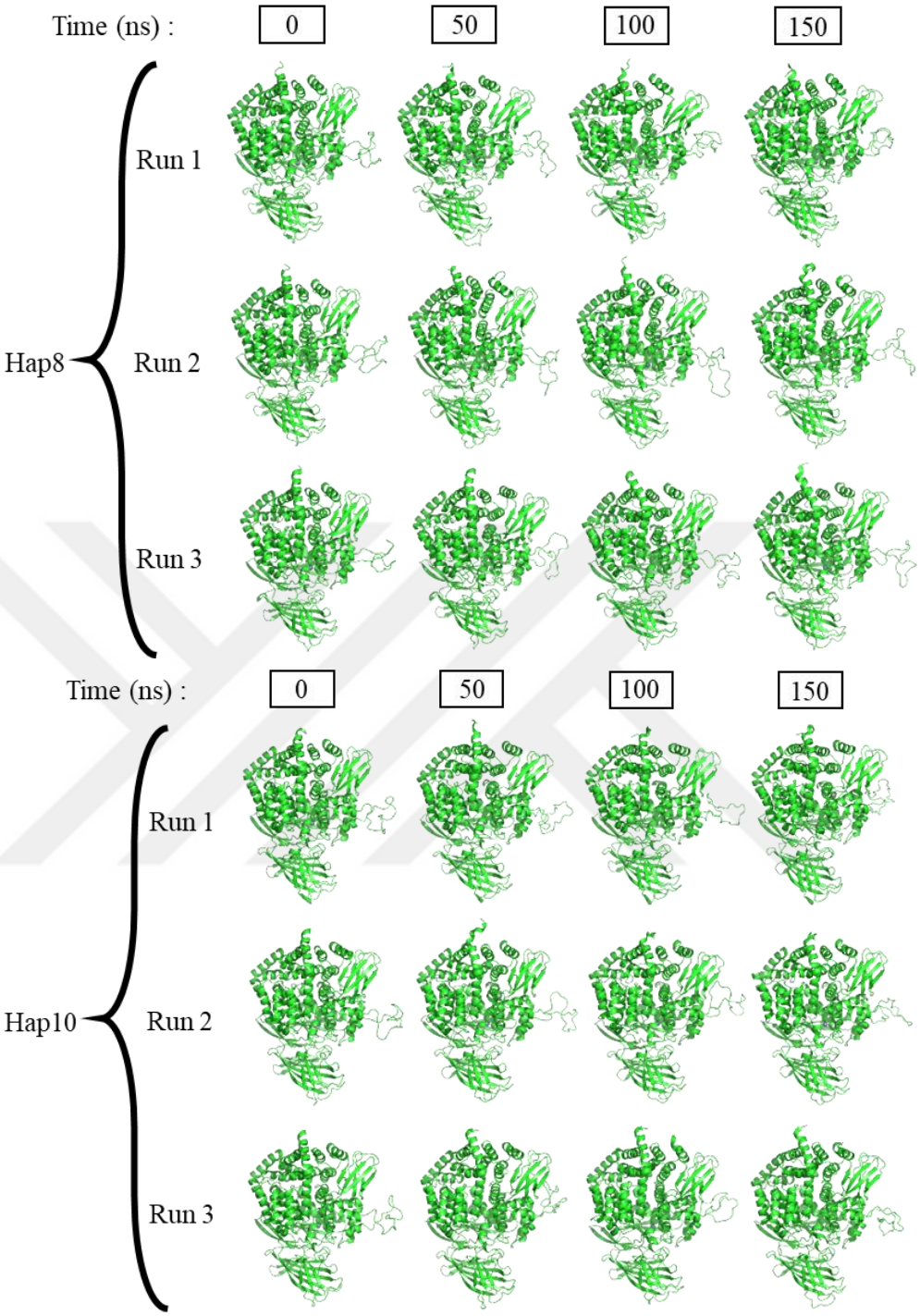



Figure 27. Primary allotypes in ERAP1 closed monomer 3D structures before simulation, at 50, 100 and 150 ns of simulation. The whole ERAP1 structure is represented with green cartoon (PDB ID: 2YD0) (cont.).

As two conformational states were previously observed in ERAP1 studies, structural changes of ERAP1 Hap2 monomer in open form was also examined during simulations (Figure 28). According to the results, all three parallel runs reached to a partially closed structure within 100 ns. Distinctively, run 3 ended up resembling more of a closed conformation. Because open conformations were quite unstable and they were moving towards a closed conformation, they were discarded in trajectory analysis of ERAP1 and ERAP2.

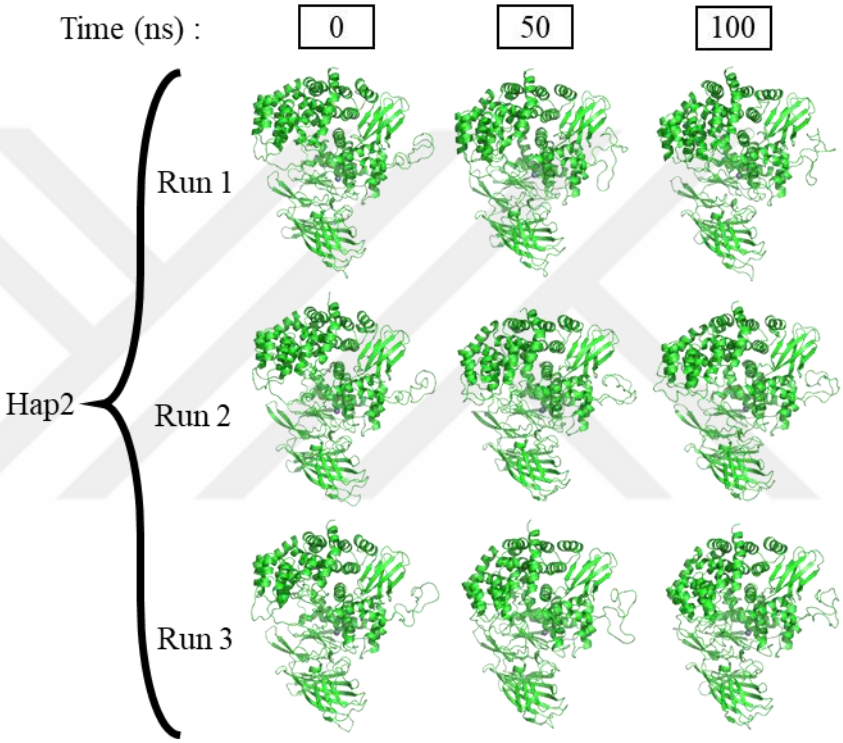



Figure 28. ERAP1 Hap2 open monomer 3D structures before simulation, at 50 and 100 ns of simulation. The whole ERAP1 structure is represented in green cartoon (PDB ID: 3QNF).

ERAP2 closed monomers were captured in screenshots (Figure 29). Structural changes were clearly observed in N392 (run 1 in all, run 2 at 50 ns and 100 ns & run 3 in all) and K392 (run 1 in all, run 2 at 50 and 100 ns & run 3 in all). Even though, the loop within ER retention signal was already structured in ERAP2 crystal models which also has a long α-helix in the same region, it was almost as flexible as the loop at sequence 485–508 in ERAP1 monomers and caused major twists in the α-helix. Strikingly, in run 3 of K392, domain I was distant from the rest late in the simulation.

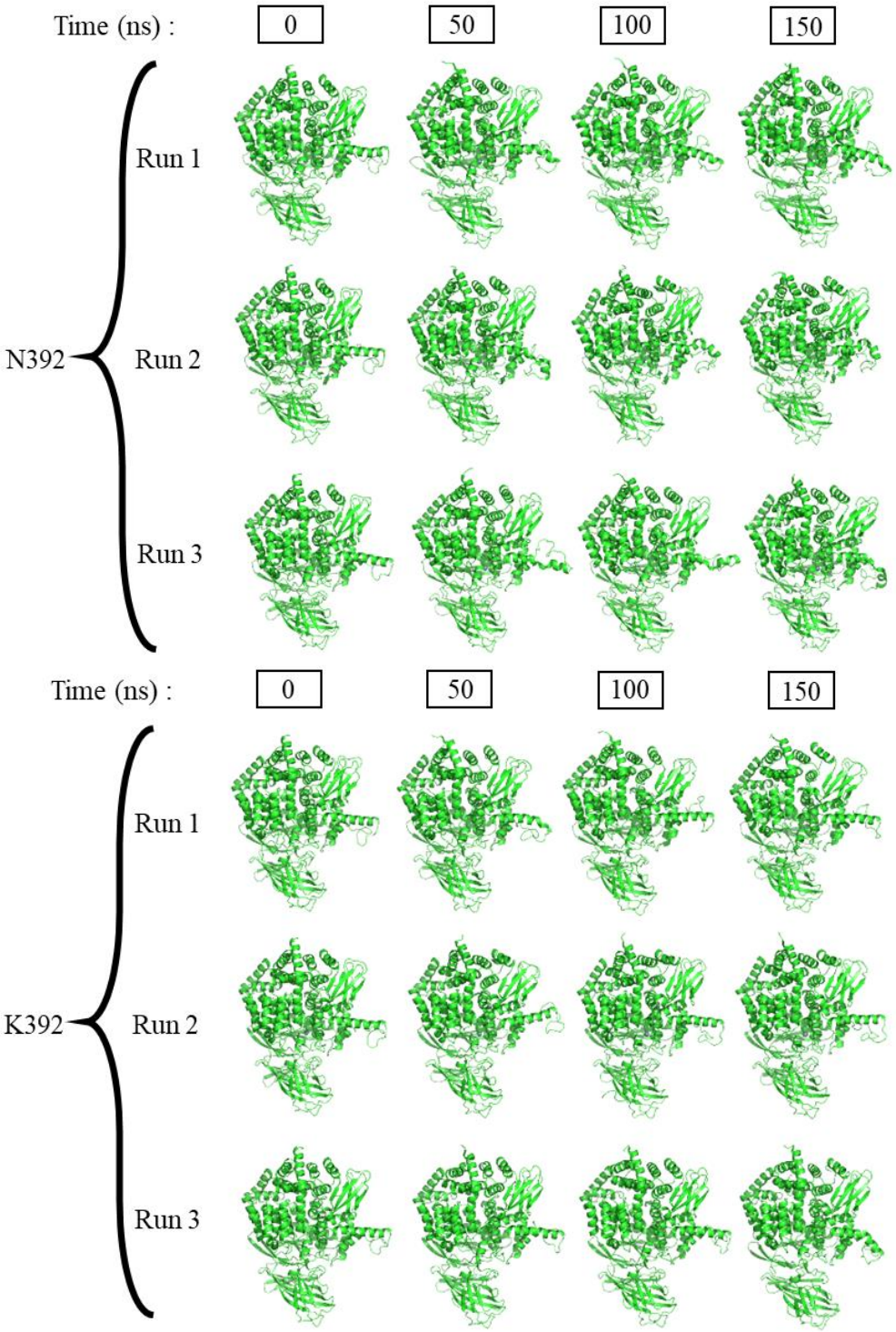



Figure 29. ERAP2 closed monomer 3D structures before simulation, at 50, 100 and 150 ns of simulation. The whole ERAP2 structure is represented in green cartoon (PDB ID: 5AB0).

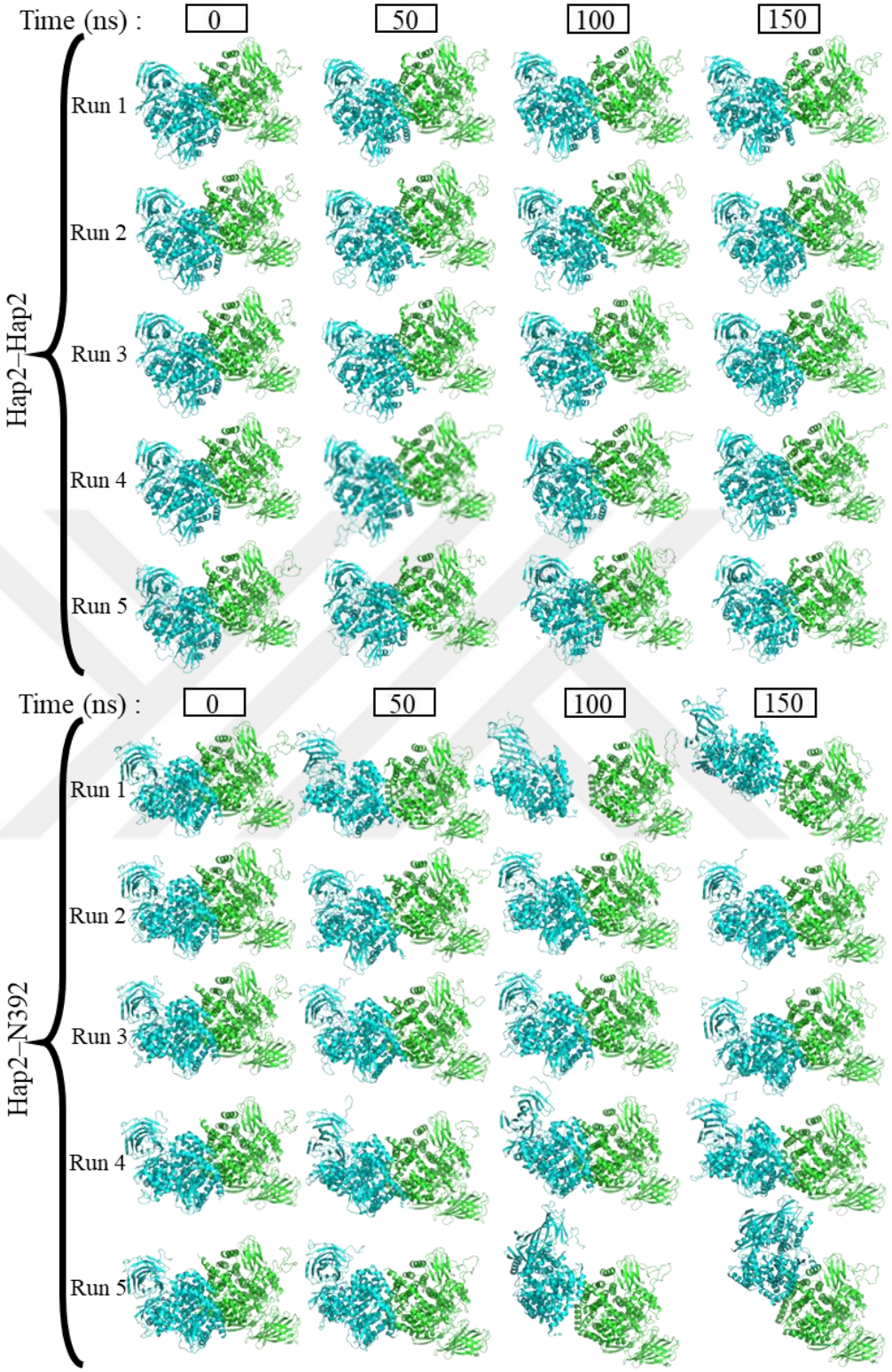



Figure 30. ERAP1 closed Hap2–Hap2 homodimer and ERAP1/2 closed Hap2–N392 heterodimer 3D structures before simulation, at 50, 100 and 150 ns of simulation. Chain A (green) and chain B (cyan) are the first and the second chains of dimer structures (PDB ID: 2YD0 & 5AB0).

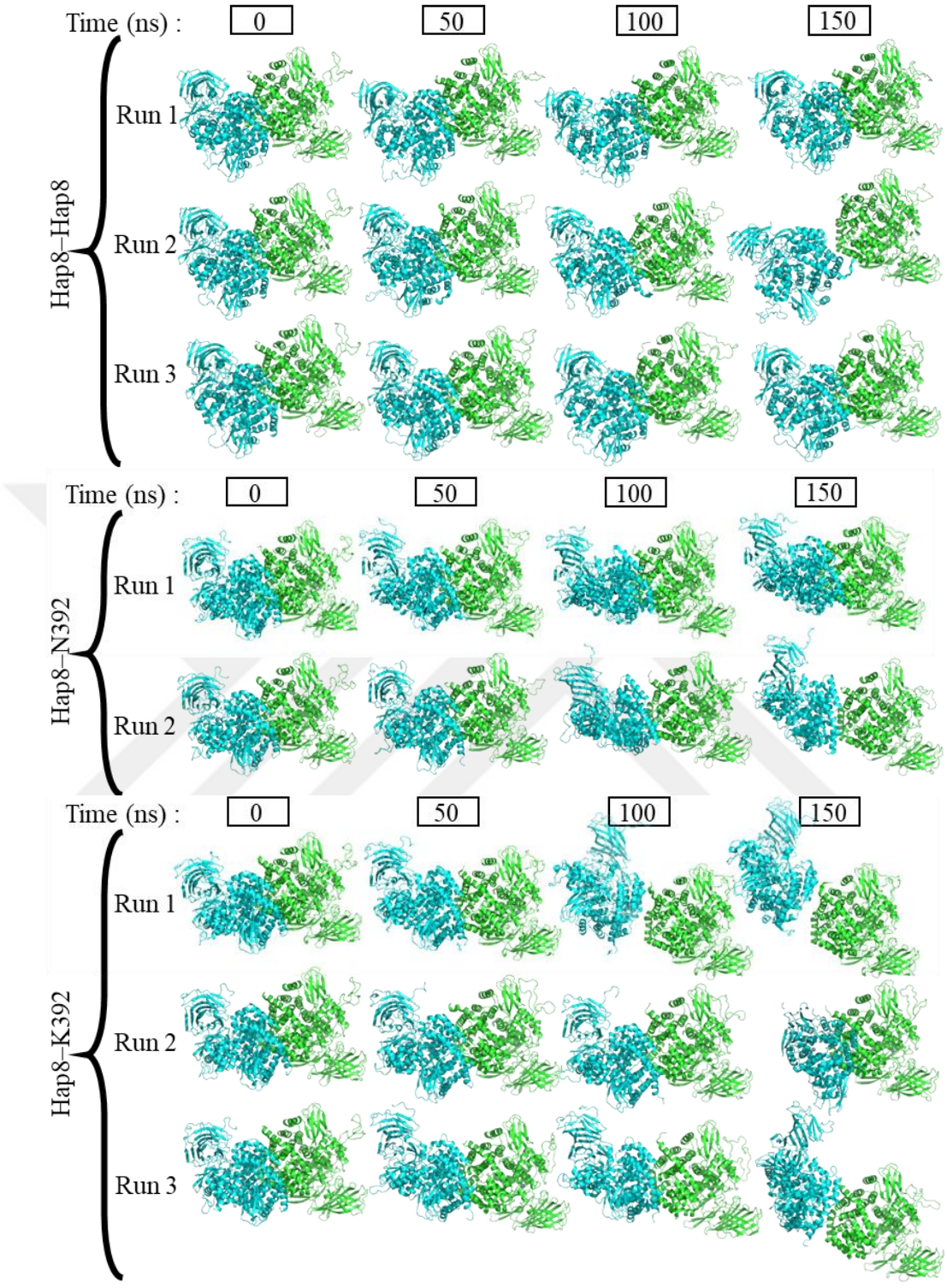



Figure 31. ERAP1 closed Hap8–Hap8 homodimer and ERAP1/2 closed Hap8–N392 and Hap8–K392 heterodimer 3D structures before simulation, at 50, 100 and 150 ns of simulation. Chain A (green) and chain B (cyan) are the first and the second chains of dimer structures (PDB ID: 2YD0 & 5AB0).

In addition to apo-enzyme structures, bestatin-bound ERAP1 Hap2 structures were investigated in both open and closed states. Analysis of these structures were consistent with the results of apo-enzyme structures based on their conformation. As protein–ligand crystal structures were only available with inhibitors, they were discarded from further analysis.

Screenshots of ERAP1 Hap2–Hap2 homodimer and ERAP1/2 Hap2–N392 heterodimer structures were compared together to elicit the relationship in dimeric interactions after performing five different parallel runs for each for 150 ns (Figure 30). Only Hap2 chains, the first chain of dimers, were superimposed for clearance. Changes in both intrachain and interchain interactions were sighted in both structures. Changes in interactions between domain I and IV were observed in Hap2–Hap2 (chain A in run 1 in all, chain A in run 2 at 50 ns & chain A in run 5 at 50 ns) and Hap2–N392 (chain A in run 1 at 50 and 100 ns & chain A in run 4 at 100 and 150 ns). Furthermore, interactions between domain IV of chain A and B were observed in Hap2–Hap2 (run 1 in all, run 2 at 50 ns, run 3 at 50 ns, run 4 at 50 ns & run 5 at 50 and 100 ns) and Hap2–N392 (run 1 in all, run 2 at 50 and 150 ns, run 3 at 50 and 150 ns, run 4 at 50 and 100 ns & run 5 in all). Interestingly, many chain As of structures showed a slight conformational change towards open state while chain Bs of both Hap2–Hap2 homodimer and Hap2–N392 heterodimer structures were almost completely in closed state. Interdomain interactions were significantly disrupted in Hap2–N392 heterodimer structures compared to Hap2–Hap2 homodimers. Especially in run 1 and run 5, Hap2–N392 heterodimers exhibited complete loss of interaction and rearrangement of dimer chains through different residue interactions, respectively.

Along with Hap2–Hap2 and Hap2–N392, combinations of other homodimer and heterodimer structures were also simulated which include Hap8–Hap8 homodimer Hap8–N392 and Hap8–K392 heterodimer structures (Figure 31). While some of the dimer structures were stable, notable structural changes were observed in Hap8–Hap8 (chain A in run 1 at 50 ns and chain B in run 1 at 50 and 100 ns) and Hap8–K392 (chain A in run 1 at 150 ns and chain B in run 1 at 150 ns & chain A of run 3 in all) in terms of intrachain interaction and Hap8–Hap8 (run 1 at 100 ns & run 2 at 50 and 150

ns), Hap8–N392 (run 1 in all & run 2 in all) and Hap8–K392 (run 1 in all, run 2 in all & run 3 at 50 and 150 ns) in terms of interchain interaction. Run 2 of Hap8–Hap8 and all runs of Hap8–K392 had differences in interchain interaction as significant as run 1 and run 5 of Hap2–N392 which presented distinctively unique interacting residues throughout simulation. Finally, a few simulations demonstrated conformational change in chain B only lesser than chain A.

### 4.6 Structural Analysis Demonstrate Differences in Dimer Forms

Trajectory analysis of ERAP1 closed primary allotypes were assessed. In RMSD, all primary allotypes were balanced between values 2 and 4 (Figure 32). In Hap2, stability was in the lowest values with all three runs combined. Run 1 of Hap3 showed similar values as Hap2 while last two parallel runs were at upper limit. All Hap8 runs showed moderate values late in the simulation. Runs of Hap10 were in slightly higher values than most primary allotypes and values were not as balanced as the rest. Nonetheless, all these changes were minimal. In RMSF, most allotype sequences showed ordered values except the loop within ER retention signal which is in similar fashion in all ERAP runs (Figure 33). Some considerable changes were slight increase in domain IV of Hap3 in run 3 and slightly high peaks at residues 113, 172, 425, 557, 628, 717 and 901 in some runs. Unfortunately, these residues were not close to SNP residues. On average, radius of gyration values for all primary allotypes increased only marginally (Figure 34). Additionally, Hap3 run 3 had an uprising line around 50 ns and Hap8 run 2 at the beginning. When CM distance between domain I and IV were checked, Hap2 run 2 had a steep hill in the second third of the simulation (Figure 35). Run 3 of Hap3 had a line similar in trend to its RGYR analysis around 50 ns. Run 2 of Hap8 showed an increase in the second half of the simulation. In SASA analysis, many runs showed an upward tendency along the simulation (Figure 36). The largest increase was observed in run 3 of Hap10. the first runs of Hap3, Hap8 and Hap10 were not affected as much as other runs. Theta angle values were somewhat consistent (Figure 37). But these values varied even among the parallel runs. Only moderate changes were sighted on run 3 of Hap3.

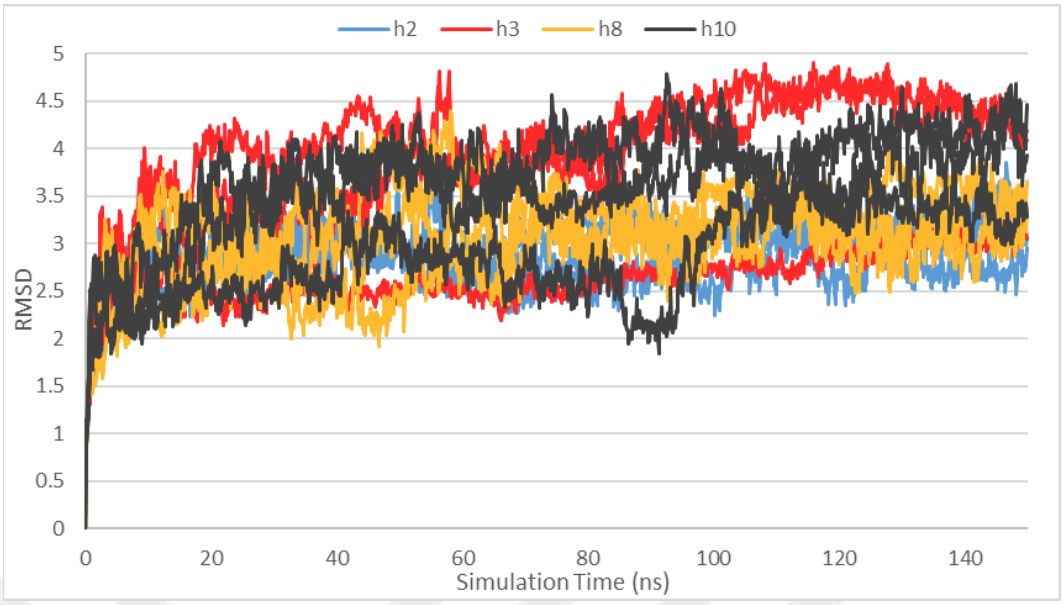



Figure 32. RMSD analysis of ERAP1 closed Hap2 (blue), Hap3 (red), Hap8 (orange) and Hap10 (grey) monomer structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0).

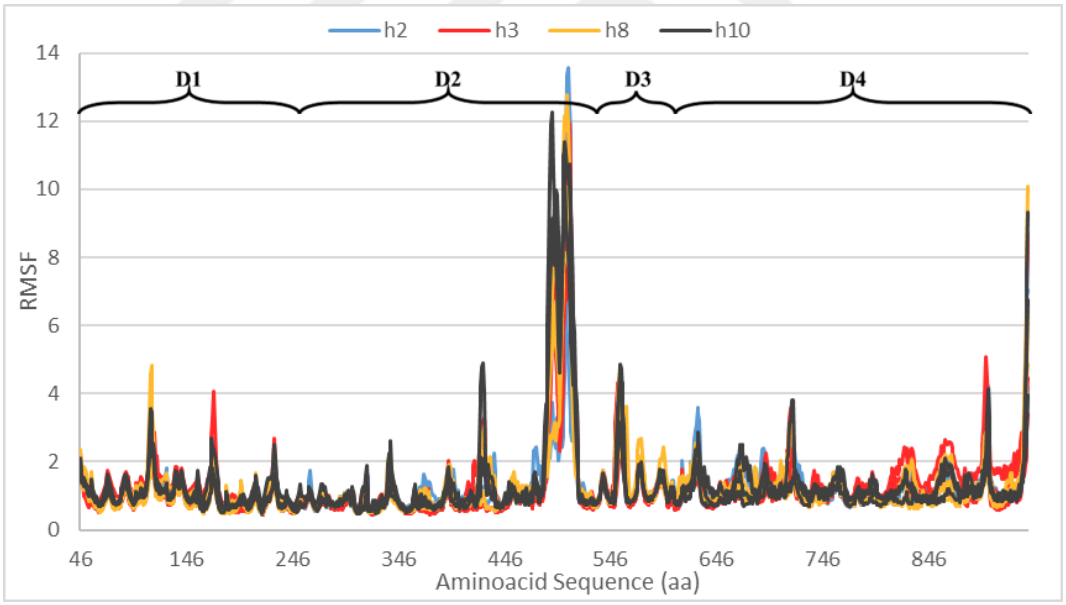



Figure 33. RMSF analysis of ERAP1 closed Hap2 (blue), Hap3 (red), Hap8 (orange) and Hap10 (grey) monomer structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0). Domains are separated by brackets (D: domain, 1–4: domain number).

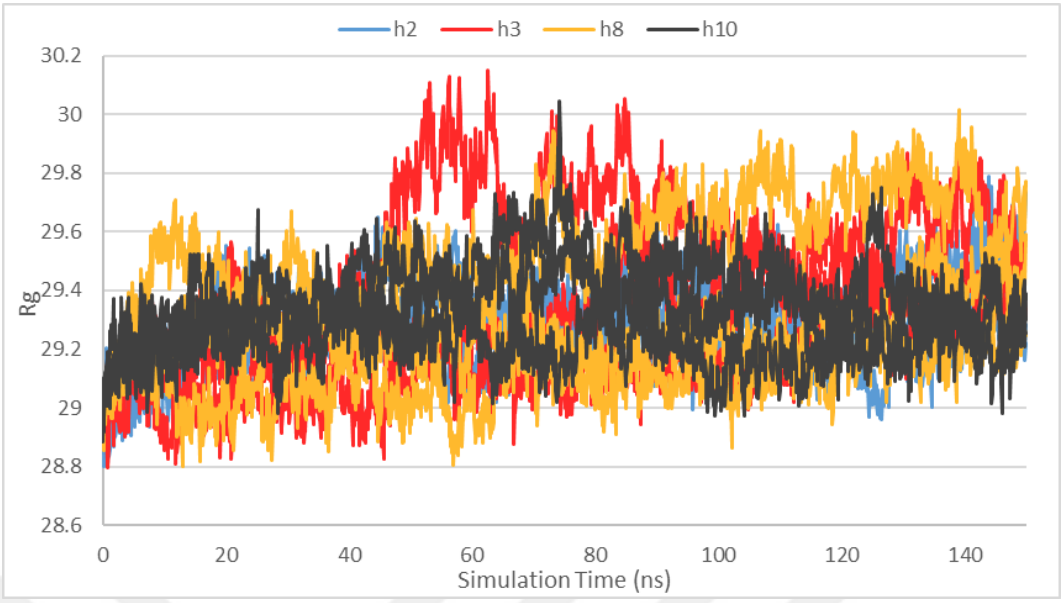



Figure 34. Radius of gyration in ERAP1 closed Hap2 (blue), Hap3 (red), Hap8 (orange) and Hap10 (grey) monomer structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0).

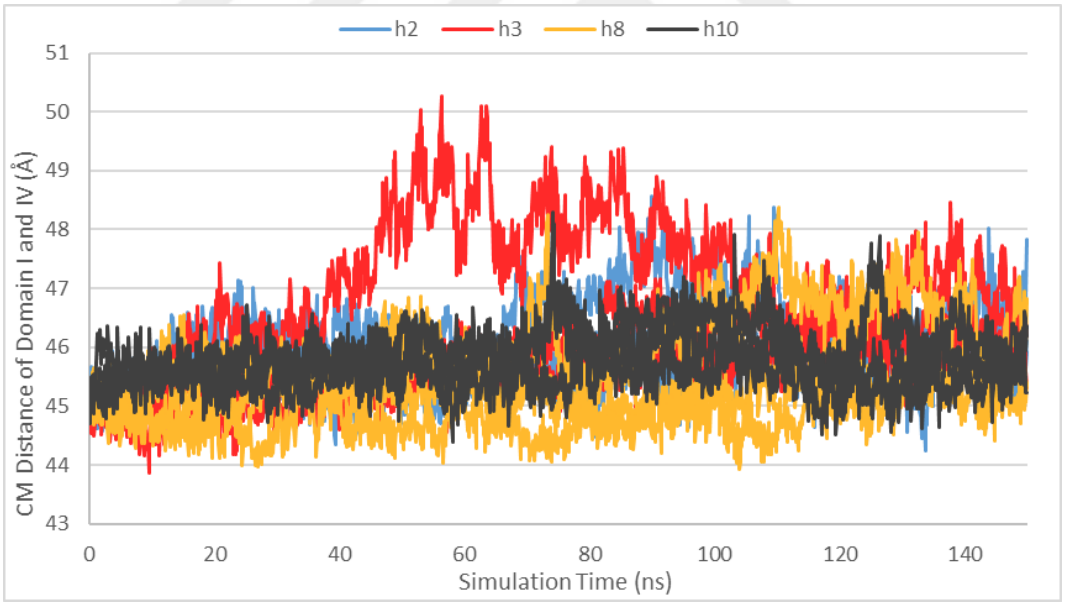



Figure 35. Domain I and IV CM distances of ERAP1 closed Hap2 (blue), Hap3 (red), Hap8 (orange) and Hap10 (grey) monomer structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0).

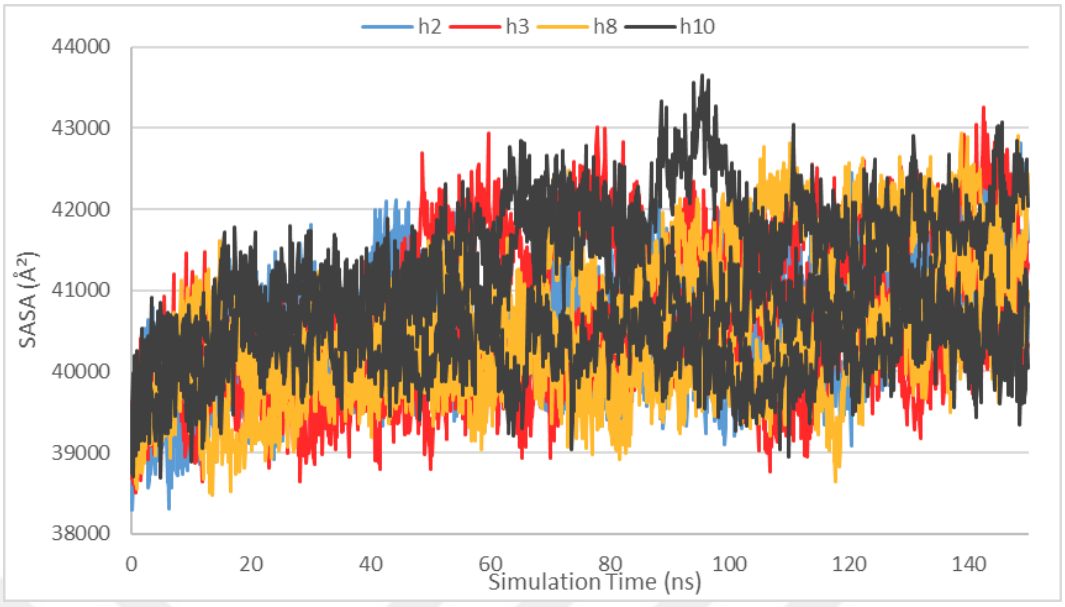



Figure 36. SASA analysis of ERAP1 closed Hap2 (blue), Hap3 (red), Hap8 (orange) and Hap10 (grey) monomer structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0).

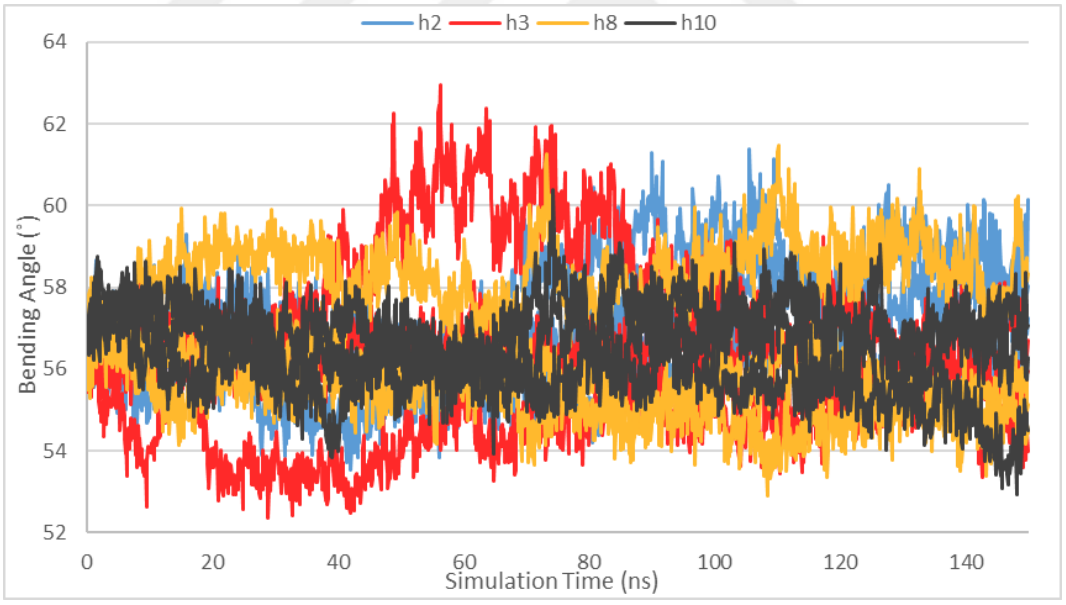



Figure 37. Theta angle of ERAP1 closed Hap2 (blue), Hap3 (red), Hap8 (orange) and Hap10 (grey) monomer structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0).

Analysis of ERAP2 closed N392 and K392 allotypes were observed along the simulation. In RMSD, stability values of all ERAP2 runs were in gradual increase throughout the simulation (Figure 38). Only a minor peak was observed in K392 run 3 near 80 ns unlike other simulations. These values were slightly lower than what was observed in ERAP1 allotypes. Results of RMSF analysis were quite similar and low among all simulations except K392 run 3 which showed an overall increase in domain I followed by domain IV (Figure 39). Another exception was ER retention loop with a significant peak in all ERAP2 runs but not any higher than ERAP1 allotypes. Results of RGYR in all simulations were almost in moderate trend (Figure 40). Likewise, CM distance of domain I and IV indicated similar trends (Figure 41). In both analyses, however, run 3 of K392 had a peak similarly sighted in RMSD in the exact time period. In SASA, values of ERAP2 allotypes gradually increased with no exception (Figure 42). Major changes were involved in run 1 of N392 and run 1 and 3 of K392. Opening angle of ERAP2 allotypes did not change as much as ERAP1 allotypes (Figure 43).

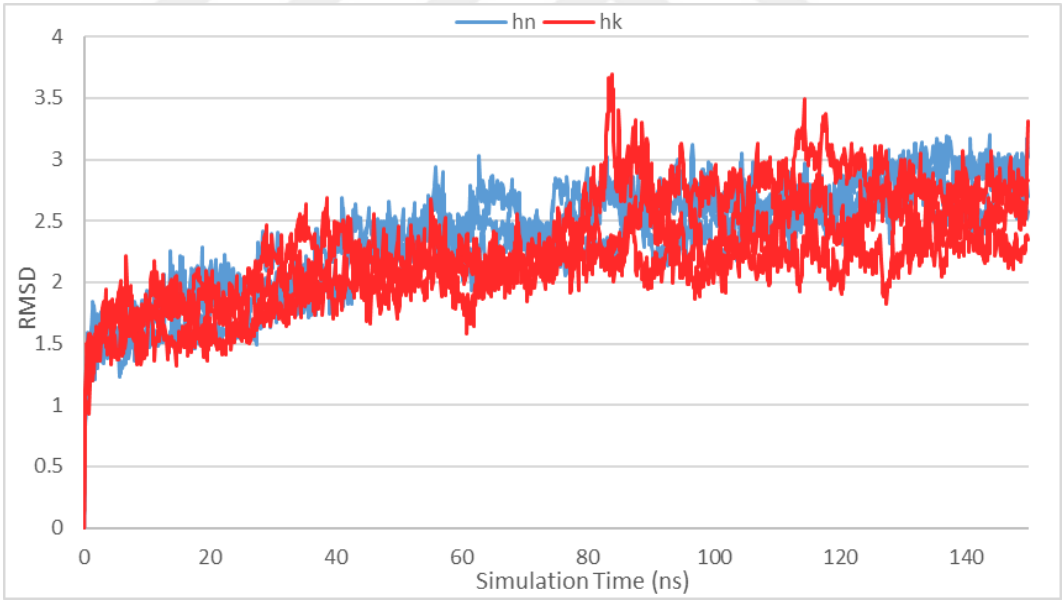



Figure 38. RMSD analysis of ERAP2 closed N392 (blue) and K392 (red) monomer structures throughout simulation. Reference structures were used for ERAP2 (PDB ID: 5AB0).

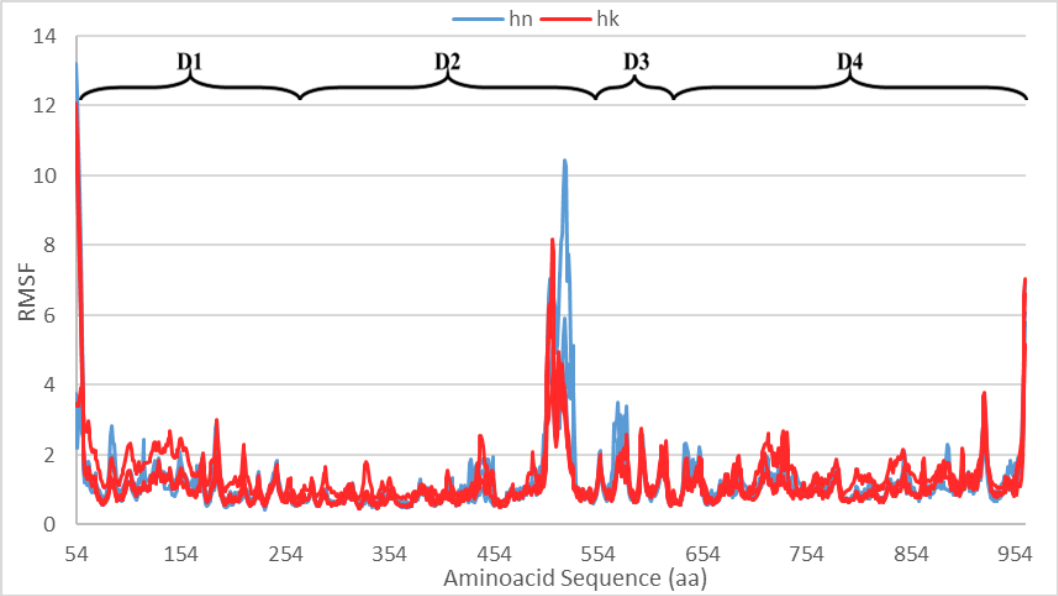



Figure 39. RMSF analysis of ERAP2 closed N392 (blue) and K392 (red) monomer structures throughout simulation (PDB ID: 5AB0). Domains are separated by brackets (D: domain, 1–4: domain number).

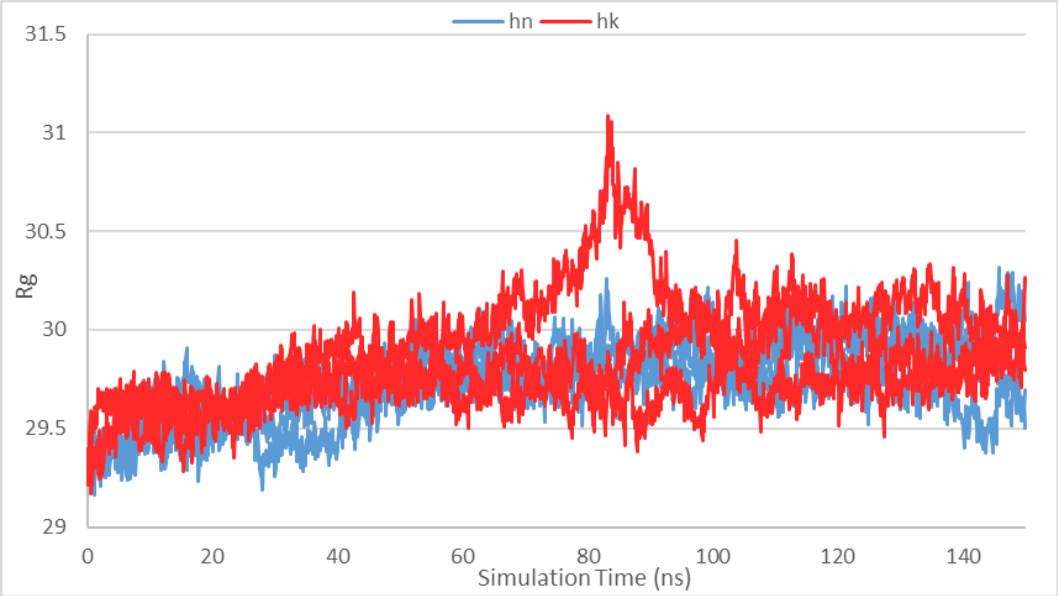



Figure 40. Radius of gyration in ERAP2 closed N392 (blue) and K392 (red) monomer structures throughout simulation. Reference structures were used for ERAP2 (PDB ID: 5AB0).

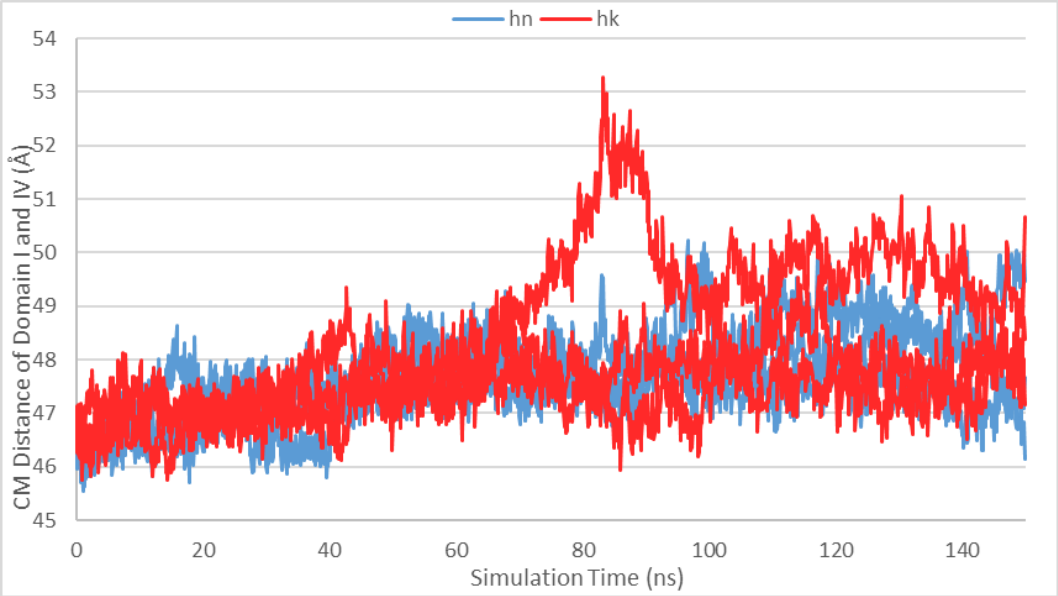



Figure 41. Domain I and IV CM distances of ERAP2 closed N392 (blue) and K392 (red) monomer structures throughout simulation. Reference structures were used for ERAP2 (PDB ID: 5AB0).

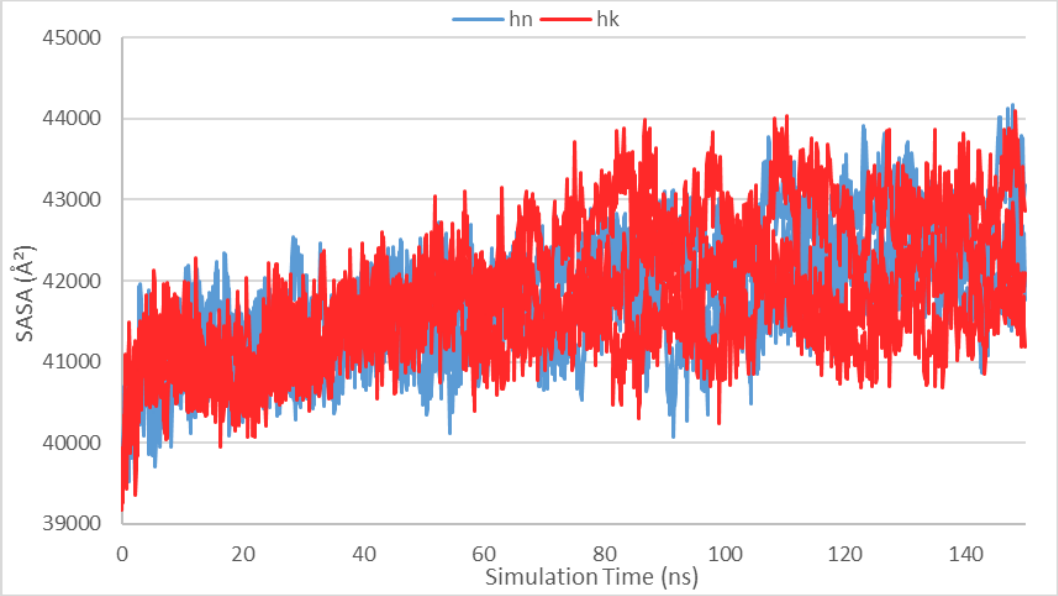



Figure 42. SASA analysis of ERAP2 closed N392 (blue) and K392 (red) monomer structures throughout simulation. Reference structures were used for ERAP2 (PDB ID: 5AB0).

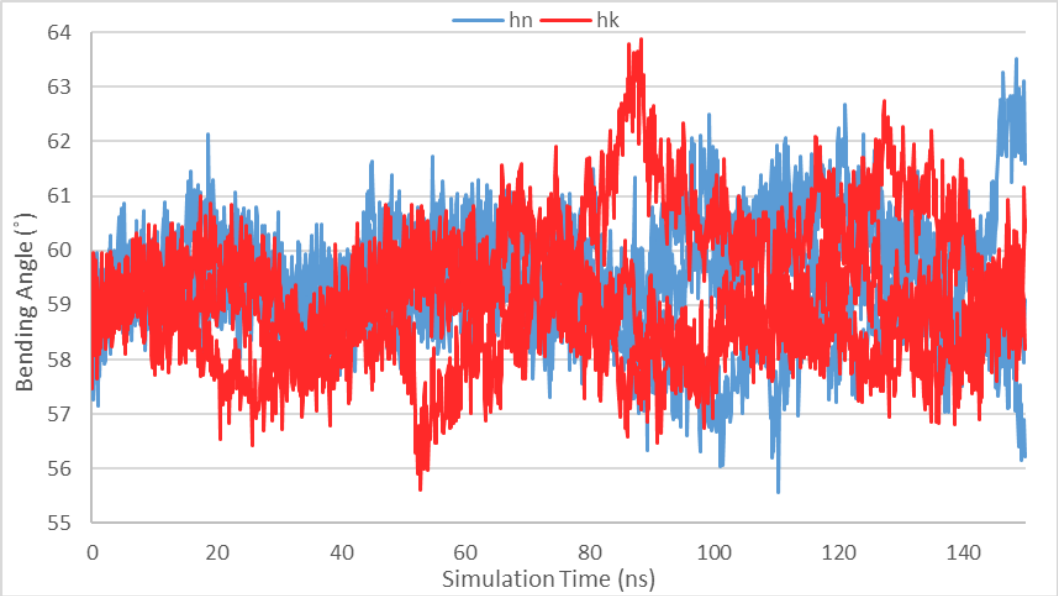



Figure 43. Theta angle of ERAP2 closed N392 (blue) and K392 (red) monomer structures throughout simulation. Reference structures were used for ERAP2 (PDB ID: 5AB0).

In the case of dimers, ERAP1 Hap2–Hap2 homodimer and ERAP1/2 Hap2–N392 heterodimer structures were compared together in the first place. In RMSD, Hap2–Hap2 values were quite balanced and almost similar in all five runs (Figure 44). Conversely, two out of five Hap2–N392 runs rose up significantly. Excluding the first and fifth runs, run 4 had a moderate increase with a stable line in the end and the rest of runs were as low as Hap2–Hap2 runs. The highest RMSF values observed in Hap2–Hap2 was in domain I of chain A in the first run followed by domain IV in the same run (Figure 45). Domain I of chain A in run 4 and chain A & B in run 5 were also in moderate increase. Nonetheless, they were minimal in comparison to the heterodimer counterpart. Interestingly, Hap2–N392 run 1 had a significant overall increase in the whole sequence of both chains with almost four times higher than any Hap2–Hap2 run. In run 5 of Hap2–N392, this much increase was also present but they were uneven. Hap2–N392 run 4 mostly had high values around domain I and IV regions only. In RGYR analysis, Hap2–N392 run 1 followed the same trend as in its RMSD results whereas run 5 was a tad lower than in its RMSD results (Figure 46). As usual, ER retention loop and near-terminal sequences exhibited high values in all runs. Results of Hap2–Hap2 did not change greatly in both RGYR and CM distance of domain IV regions (Figure 49). In heterodimers, Hap2–N392 showed similarity to

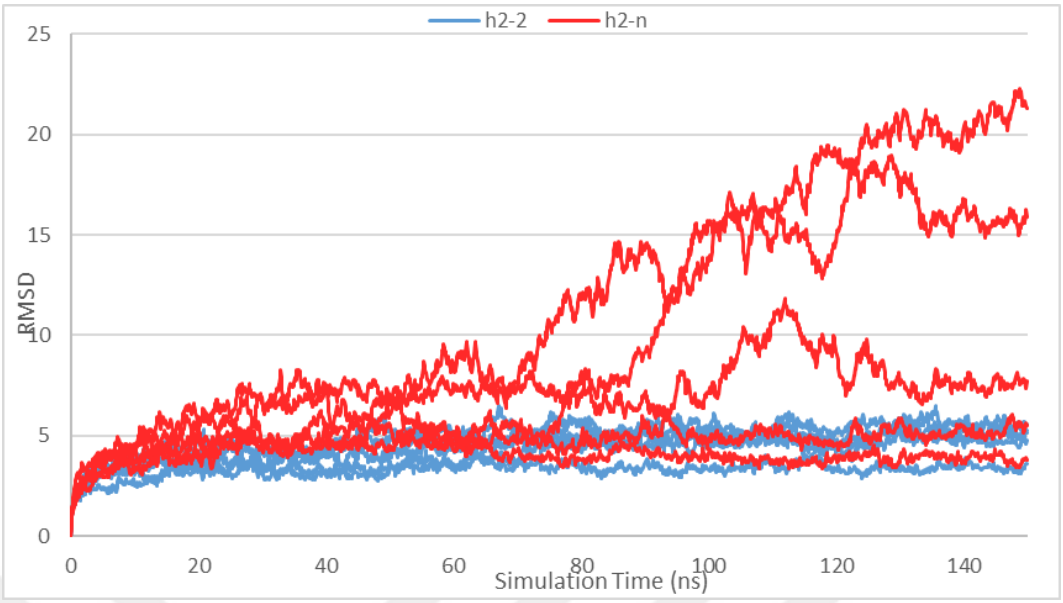



Figure 44. RMSD analysis of ERAP1 closed Hap2–Hap2 homodimer (blue) and ERAP1/2 closed Hap2–N392 heterodimer (red) structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

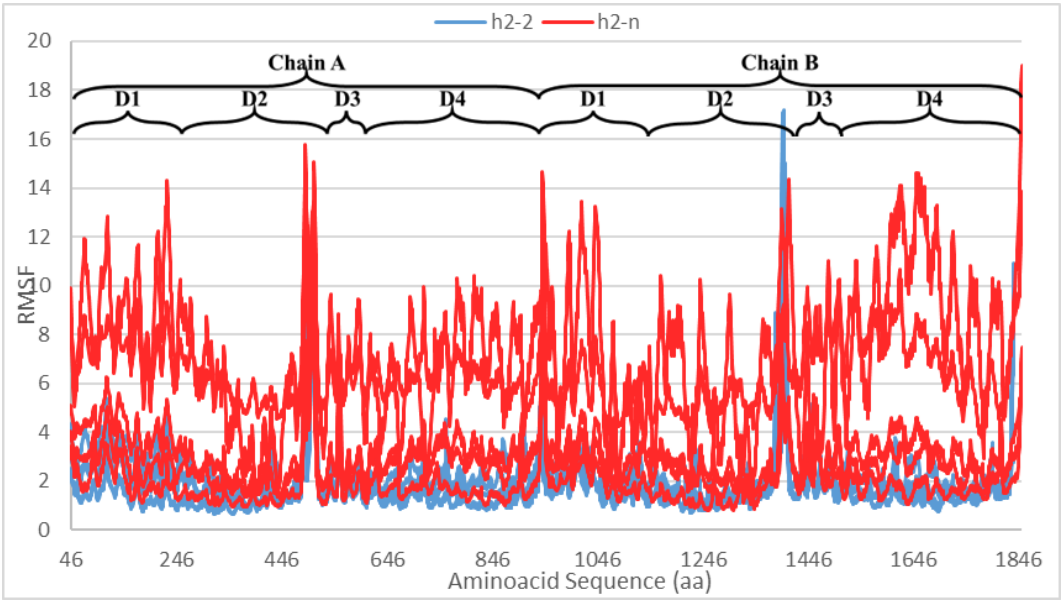



Figure 45. RMSF analysis of ERAP1 closed Hap2–Hap2 homodimer (blue) and ERAP1/2 closed Hap2–N392 heterodimer (red) structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0). Chains and domains are separated by brackets (D: domain, 1–4: domain number).

aforementioned results with only run 4 being as low as Hap2–Hap2 runs at the end of simulation. When CM distance was calculated between domain I and IV of individual

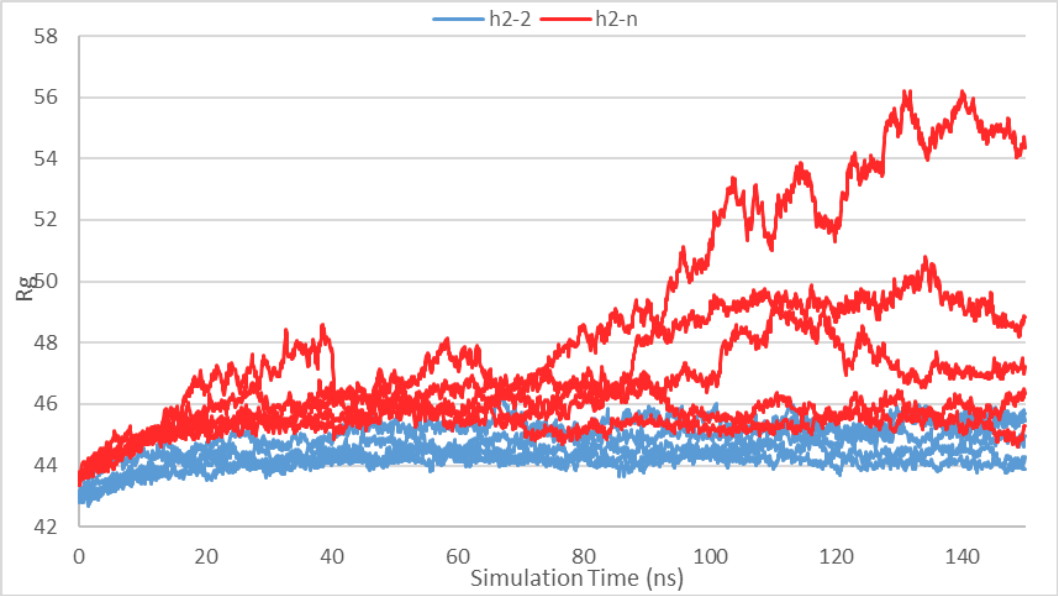



Figure 46. Radius of gyration in ERAP1 closed Hap2–Hap2 homodimer (blue) and ERAP1/2 closed Hap2–N392 heterodimer (red) structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

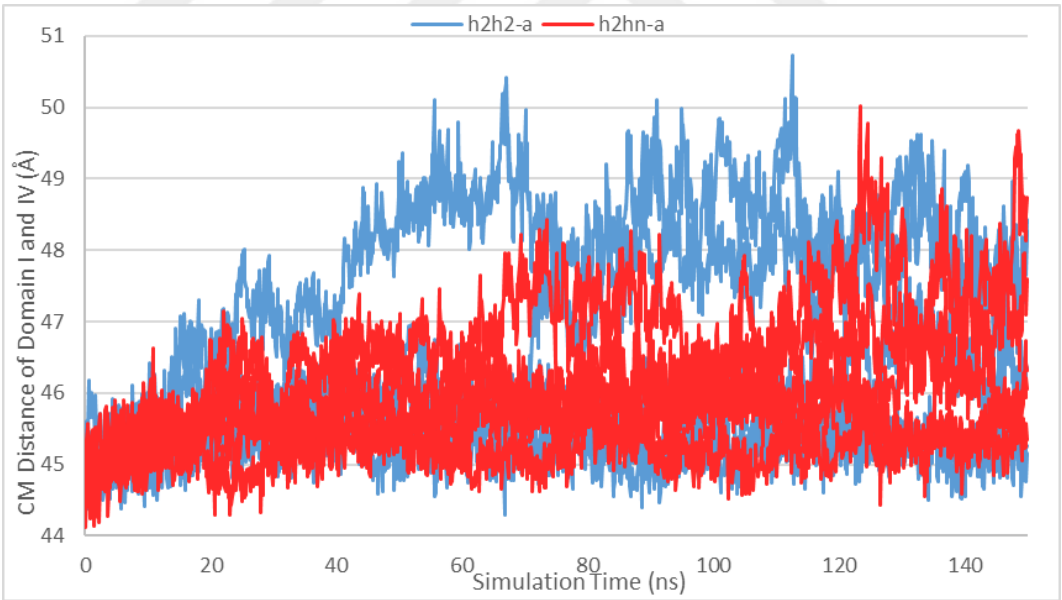



Figure 47. Domain I and IV CM distances of the first chain (chain A) in ERAP1 closed Hap2–Hap2 homodimer (blue) and ERAP1/2 closed Hap2–N392 heterodimer (red) structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

chains, the results were different. In regards to chain A, Hap2–Hap2 run 1 and run 5 had considerably increased values (Figure 47). This kind of increase for Hap2–N392

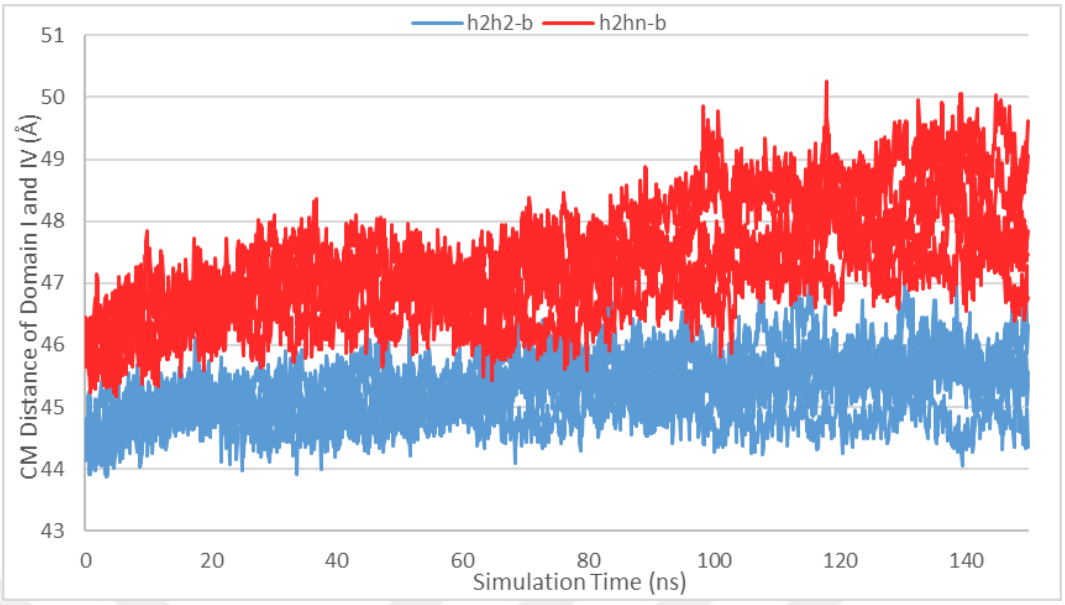



Figure 48. Domain I and IV CM distances of the second chain (chain B) in ERAP1 closed Hap2–Hap2 homodimer (blue) and ERAP1/2 closed Hap2–N392 heterodimer (red) structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

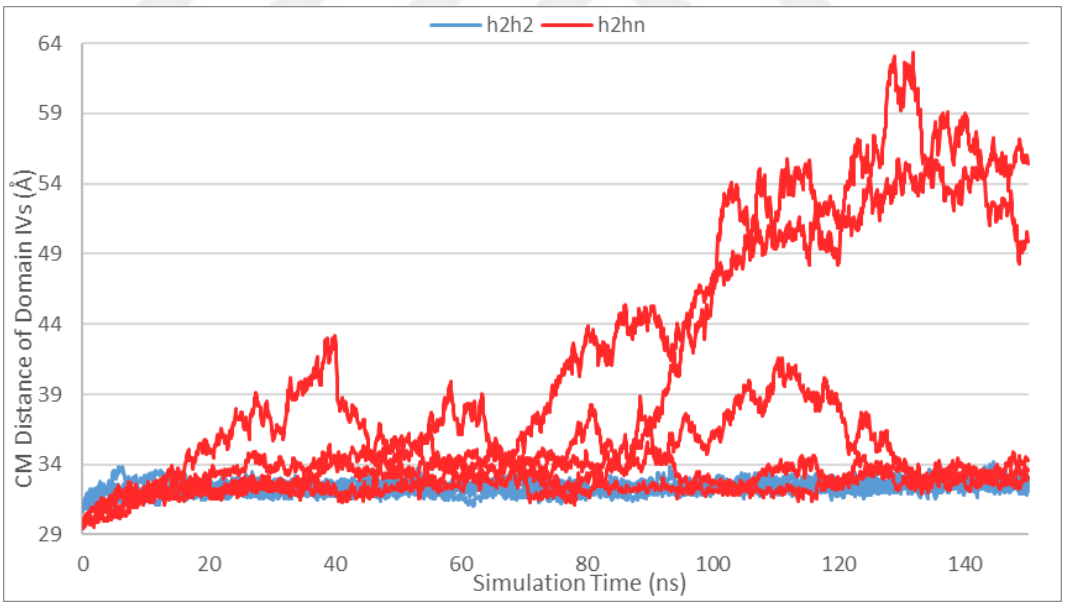



Figure 49. Domain IV CM distances of chains in ERAP1 closed Hap2–Hap2 homodimer (blue) and ERAP1/2 closed Hap2–N392 heterodimer (red) structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

was only observed in run 1 and only slightly in run 4. As for chain B, a clear division line between Hap2–Hap2 and Hap2–N392 was detected (Figure 48). Runs of Hap2–

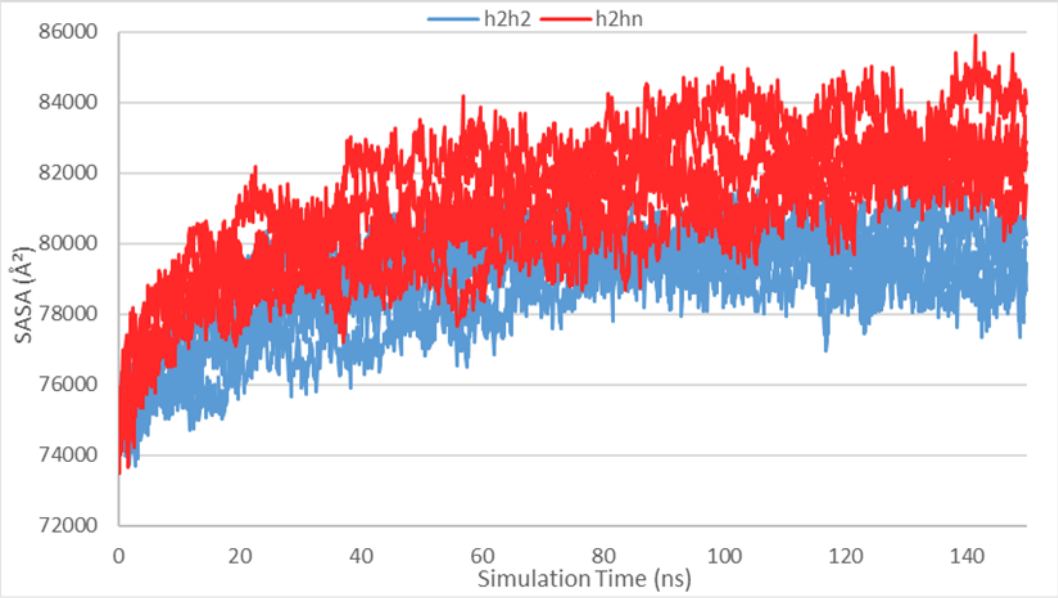



Figure 50. SASA analysis of ERAP1 closed Hap2–Hap2 homodimer (blue) and ERAP1/2 closed Hap2–N392 heterodimer (red) structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

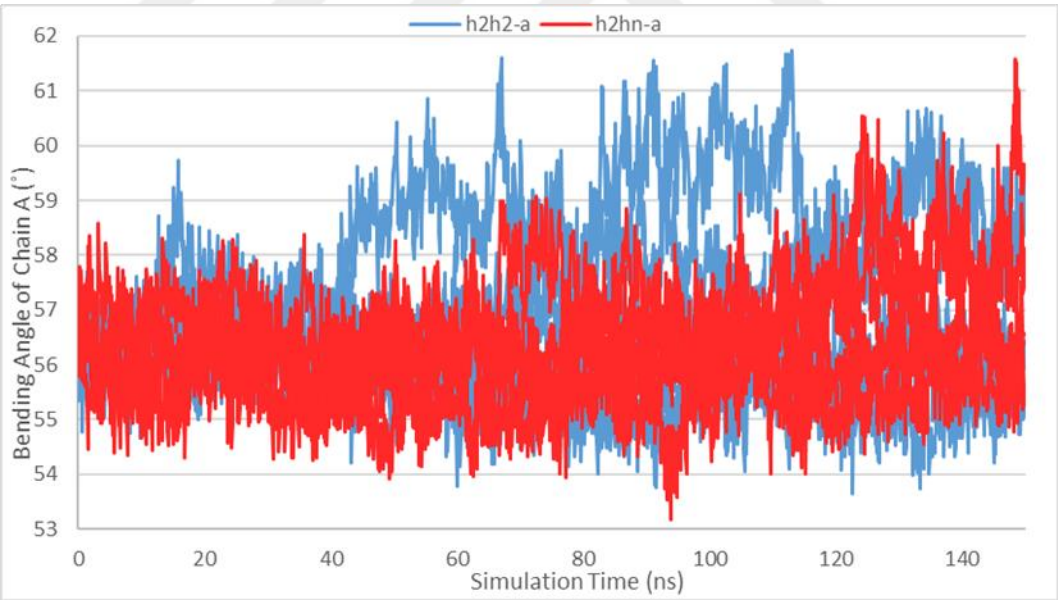



Figure 51. Theta angle of the first chain (chain A) in ERAP1 closed Hap2–Hap2 homodimer (blue) and ERAP1/2 closed Hap2–N392 heterodimer (red) structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

N392 had more increase over time. In SASA analysis, all runs were in marginal increase and all Hap2–N392 runs increased faster than Hap2–Hap2 runs (Figure 50).

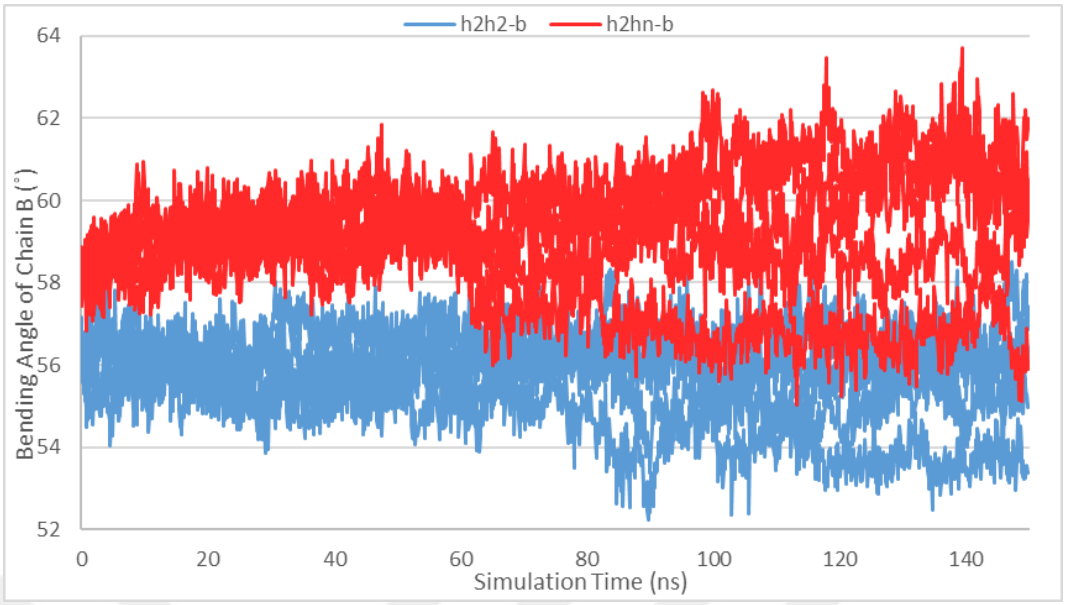



Figure 52. Theta angle of the second chain (chain B) in ERAP1 closed Hap2–Hap2 homodimer (blue) and ERAP1/2 closed Hap2–N392 heterodimer (red) structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

In the first run of Hap2–Hap2, a sudden increase in chain A angle took place in mid run and retained in theta angle analysis (Figure 51). Furthermore, a gradual increase was observed in the first run of Hap2–N392. Other than that, chain A angle of many runs were stable. Chain B angles almost divided into two sections (Figure 52). Angles of Hap2–N392 were higher in overall with run 3 exhibiting a smaller angle in second half of the simulation. Chain B angle of Hap2–Hap2 run 3 decreased whereas other parallel runs remained stable.

Combinations of ERAP1 Hap8 and ERAP2 allotypes were examined. A remarkable increase was observed in first two runs of Hap8–Hap8 runs unlike Hap2–Hap2 runs (Figure 53). In run 2, a sudden peak overtook the run with a significant imbalance before the simulation end. In run 1, RMSD values only equilibrated after a certain rise. However, in run 3, Hap8–Hap8 conserved its stability. All Hap8–K392 runs were in higher flexibility than other dimers. On the other hand, Hap8–N392 results were moderate and between Hap8–Hap8 homodimer and Hap8–K392 heterodimer. Run 1 of Hap8–K392 increased very early in the simulation unlike other parallel runs which were in similar levels of flexibility later on. Unsurprisingly, run 1

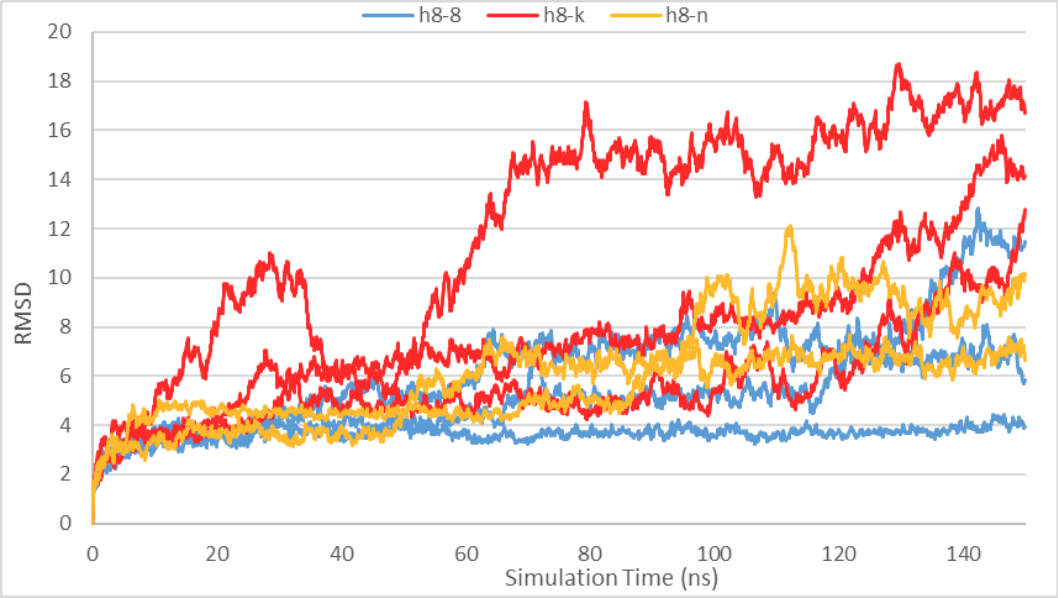



Figure 53. RMSD analysis of ERAP1 closed Hap8–Hap8 homodimer (blue), ERAP1/2 closed Hap8–K392 (red) and Hap8–N392 (orange) heterodimer structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

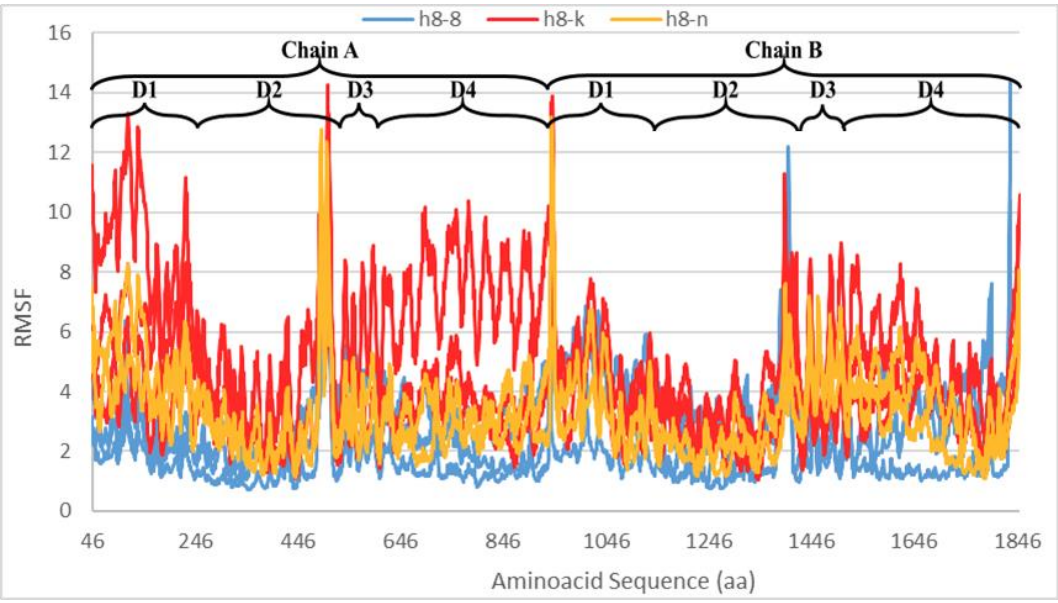



Figure 54. RMSF analysis of ERAP1 closed Hap8–Hap8 homodimer (blue), ERAP1/2 closed Hap8–K392 (red) and Hap8–N392 (orange) heterodimer structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0). Chains and domains are separated by brackets (D: domain, 1–4: domain number).

and 2 of Hap8–Hap8 had uneven values through entire sequence in RMSF analysis (Figure 54). Run 3 of Hap8–Hap8 had the most stable sequence. In heterodimers, only

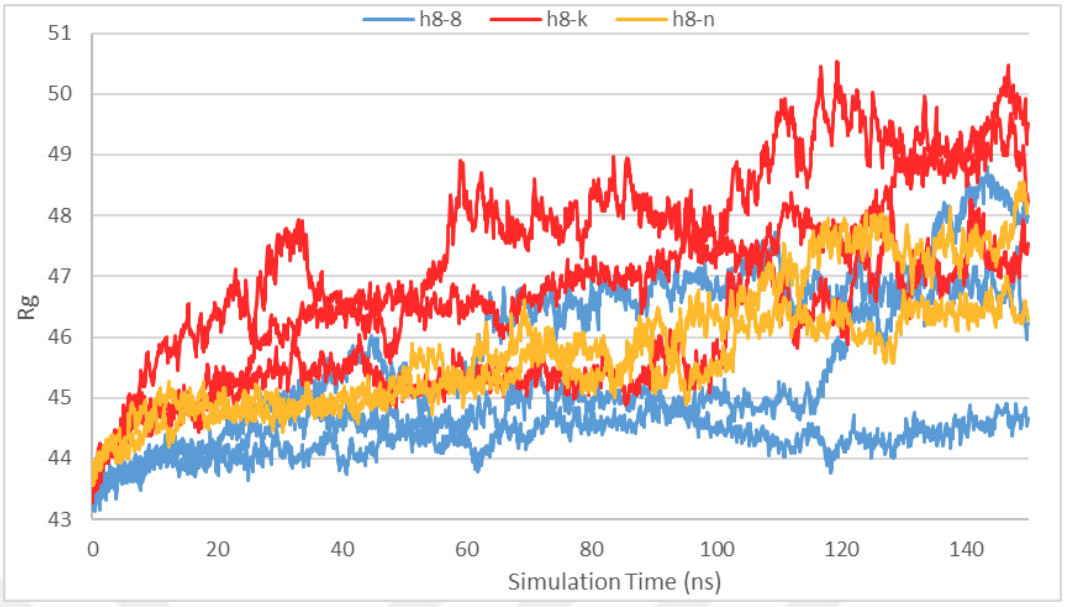



Figure 55. Radius of gyration in ERAP1 closed Hap8–Hap8 homodimer (blue), ERAP1/2 closed Hap8–K392 (red) and Hap8–N392 (orange) heterodimer structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

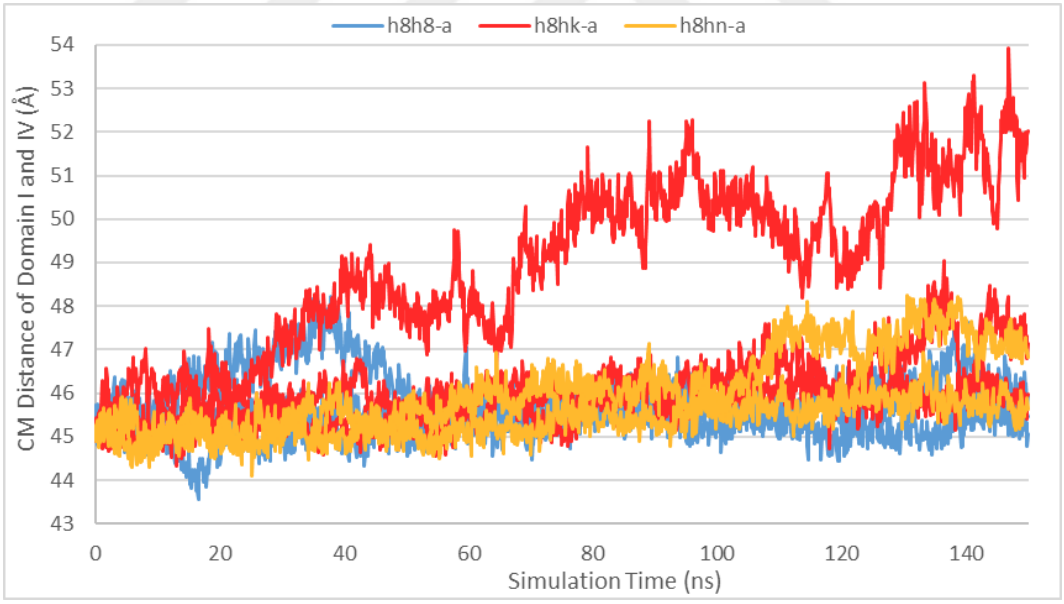



Figure 56. Domain I and IV CM distances of the first chain (chain A) in ERAP1 closed Hap8–Hap8 homodimer (blue), ERAP1/2 closed Hap8–K392 (red) and Hap8–N392 (orange) heterodimer structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

run 2 of Hap8–K392 and run 1 of Hap8–N392 were the most stable structures, albeit high. Domain I and IV of chain A and domain IV of chain B were significantly unstable

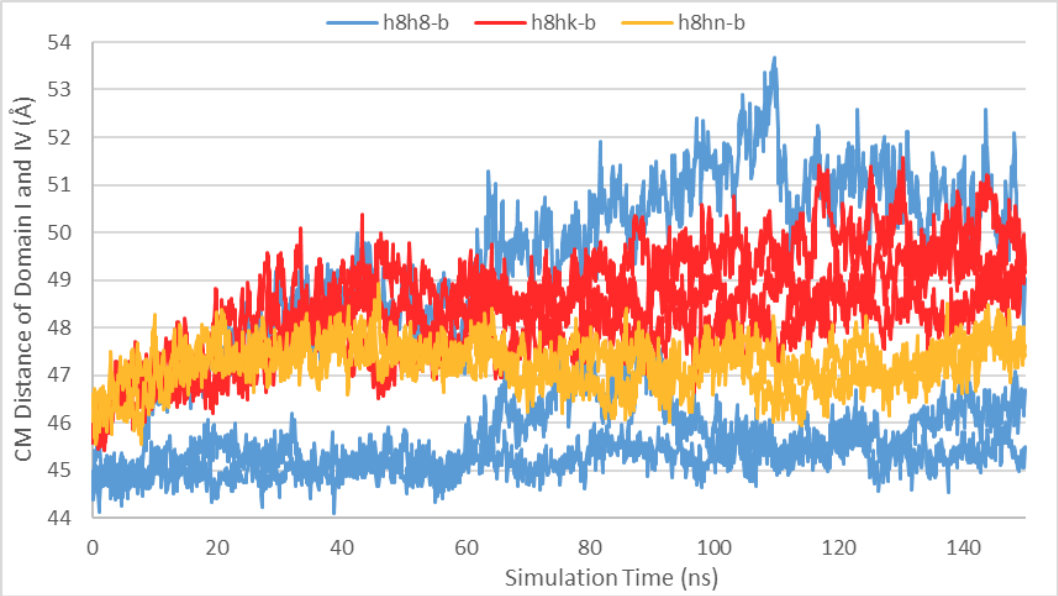



Figure 57. Domain I and IV CM distances of the second chain (chain B) in ERAP1 closed Hap8–Hap8 homodimer (blue), ERAP1/2 closed Hap8–K392 (red) and Hap8–N392 (orange) heterodimer structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

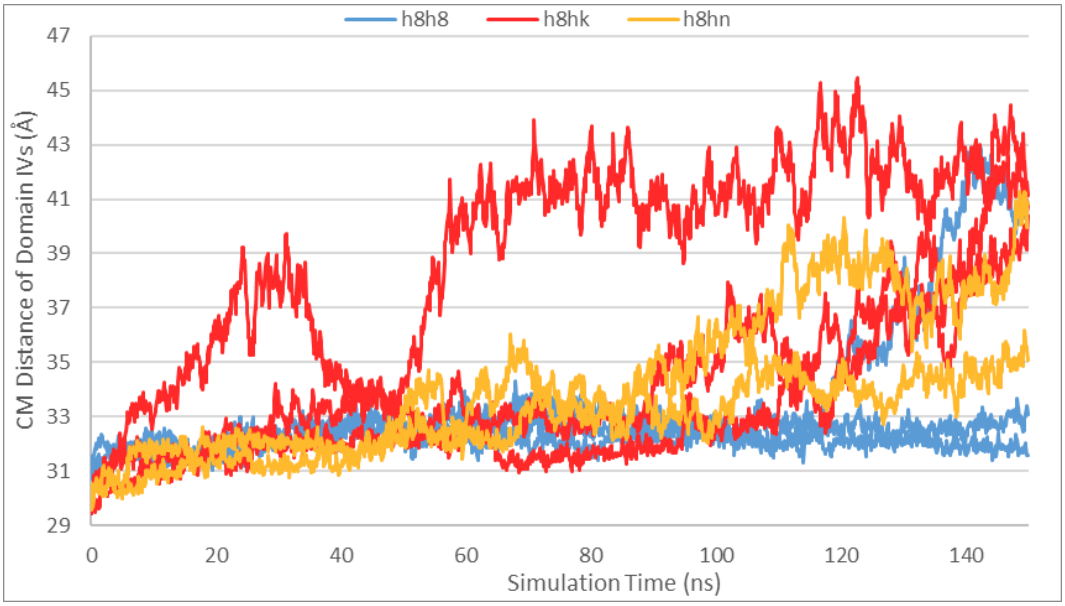



Figure 58. Domain IV CM distances of chains in ERAP1 closed Hap8–Hap8 homodimer (blue), ERAP1/2 closed Hap8–K392 (red) and Hap8–N392 (orange) heterodimer structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

in run 1 of Hap8–K392 in comparison to these runs and was followed by domain I of chain A in the third run of Hap8–K392 and domain I of chain A and domain IV of

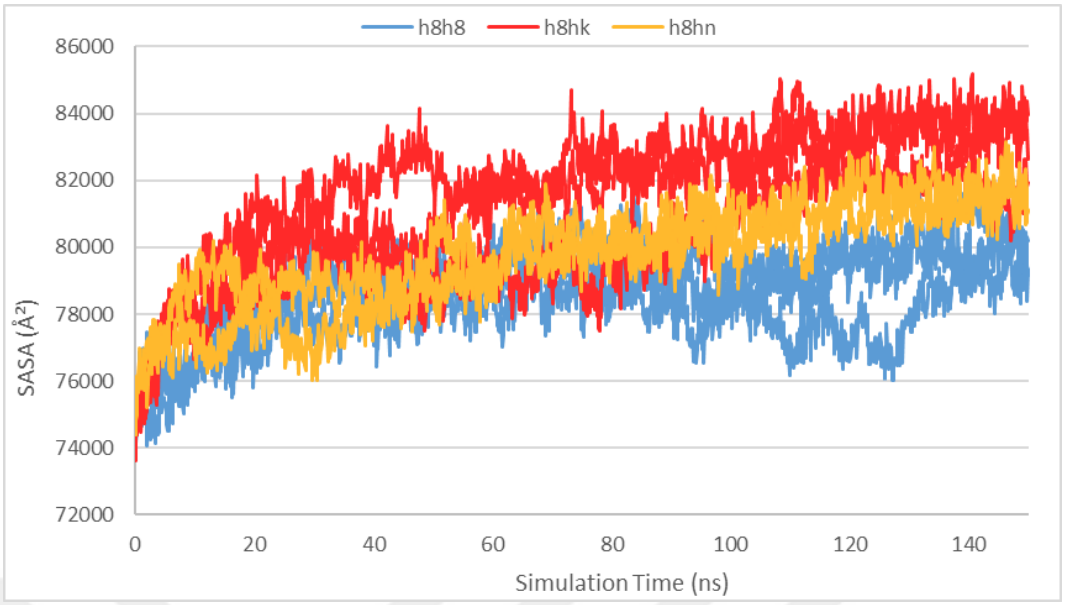



Figure 59. SASA analysis of ERAP1 closed Hap8–Hap8 homodimer (blue), ERAP1/2 closed Hap8–K392 (red) and Hap8–N392 (orange) heterodimer structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

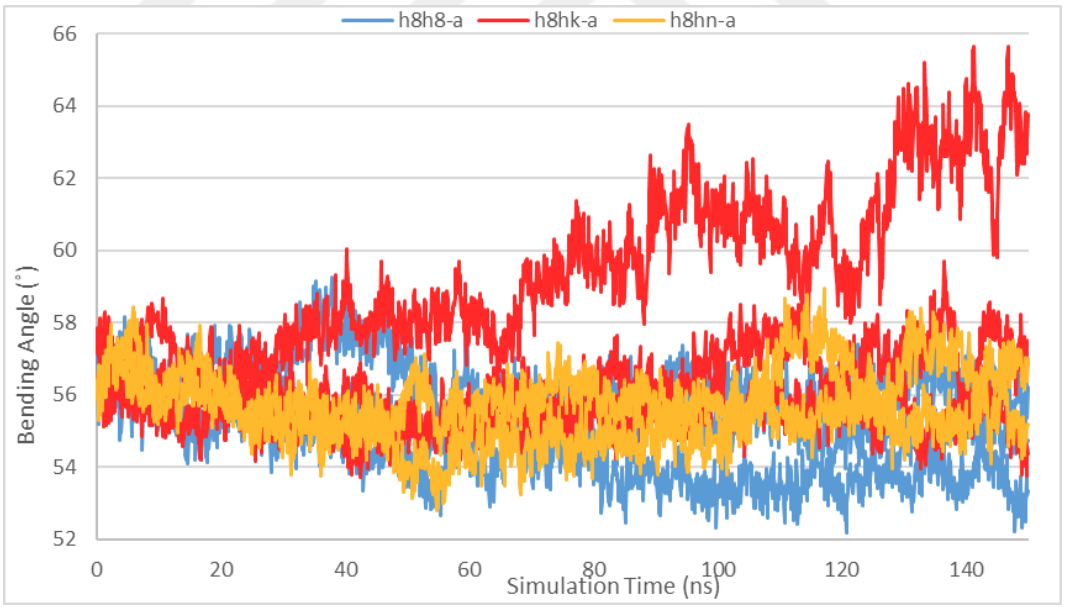



Figure 60. Theta angle of the first chain (chain A) in ERAP1 closed Hap8–Hap8 homodimer (blue), ERAP1/2 closed Hap8–K392 (red) and Hap8–N392 (orange) heterodimer structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

chain B in 2 nd run of Hap8–N392. Results of RGYR were not different from RMSD results (Figure 55). Only, run 1 of Hap8–K392 had closer values to other runs. Domain

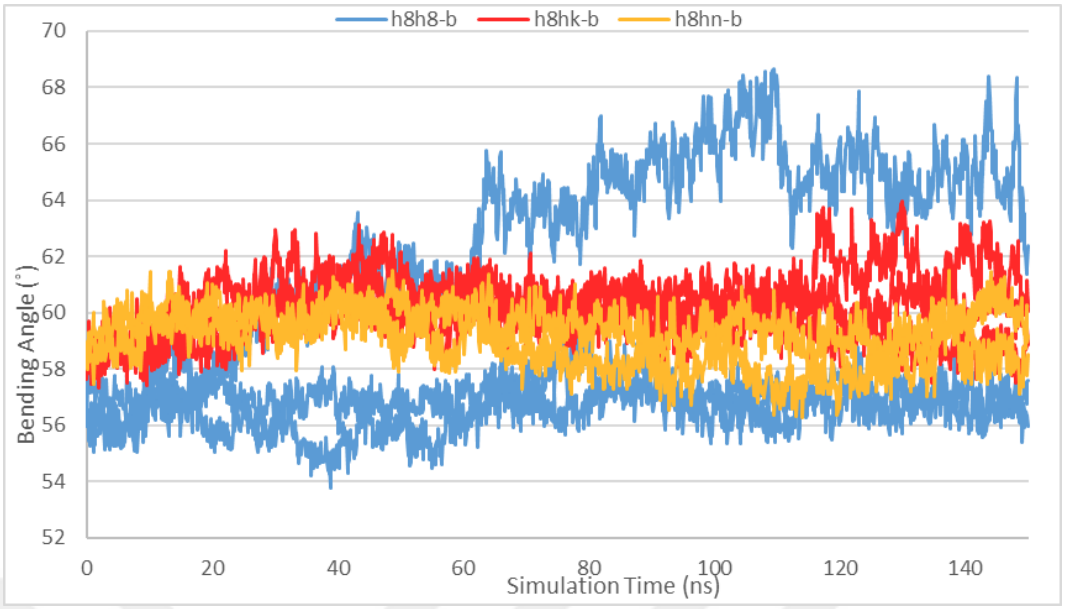



Figure 61. Theta angle of the second chain (chain B) in ERAP1 closed Hap8–Hap8 homodimer (blue), ERAP1/2 closed Hap8–K392 (red) and Hap8–N392 (orange) heterodimer structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

IV CM distances only indicated a notable increase in run 2 of Hap8–Hap8 comparable to Hap8–K392 heterodimers (Figure 58). Heterodimers showed similar patterning of RMSD and RGYR results. Domain I and IV CM distances were a different case. In chain A, intrachain distances of all Hap8–Hap8 runs were smooth (Figure 56). Chain A of Hap8–K392 run 1 was in lower peak and chain A of Hap8–N392 run 1 rose up more than its second parallel run. All other runs were steady. However, run 1 values in Hap8–Hap8 alone was very high in chain B unlike any other homodimer (Figure 57). This was comparable to Hap8–K392 runs which showed values as high. Runs of Hap8–N392 initially decreased then increased back. While SASA of dimer runs generally increased, run 3 of Hap8–Hap8 showed minor decrease at times (Figure 59). Interestingly, many runs demonstrated declining values only in the early-simulation. Changes in bending angle of chain A exhibited a small decreasing pattern in all runs of Hap8–Hap8 (Figure 60). This angle was significantly large in run 1 of Hap8–K392 while it was mostly stable in the rest of runs. On the contrary, chain B of run 1 in Hap8–Hap8 differed a lot (Figure 61). It had a significant increase overpassing the angle of heterodimers which were stable along the simulation.

PCA of all closed monomers including ERAP1 and ERAP2 allotypes were compared (Figure 62). Even though the motions in the second and the third runs of Hap3 were a tad higher, ERAP1 primary allotypes were stable in general. ER retention loop was the most flexible region in all runs. Besides that, small but common motions around the structures were observed in loop region between residues 550-560 and regions where conformational changes occur such as helix H4a and near-region (between residues 420-433) as well as domain I and IV.

ERAP2 allotypes in PCA results differed vastly from each other (Figure 62). While N392 runs were satisfyingly stable, high-degree motions were present in run 2 and 3 of K392. For instance, domain I in the third run of K392 was significantly flexible.

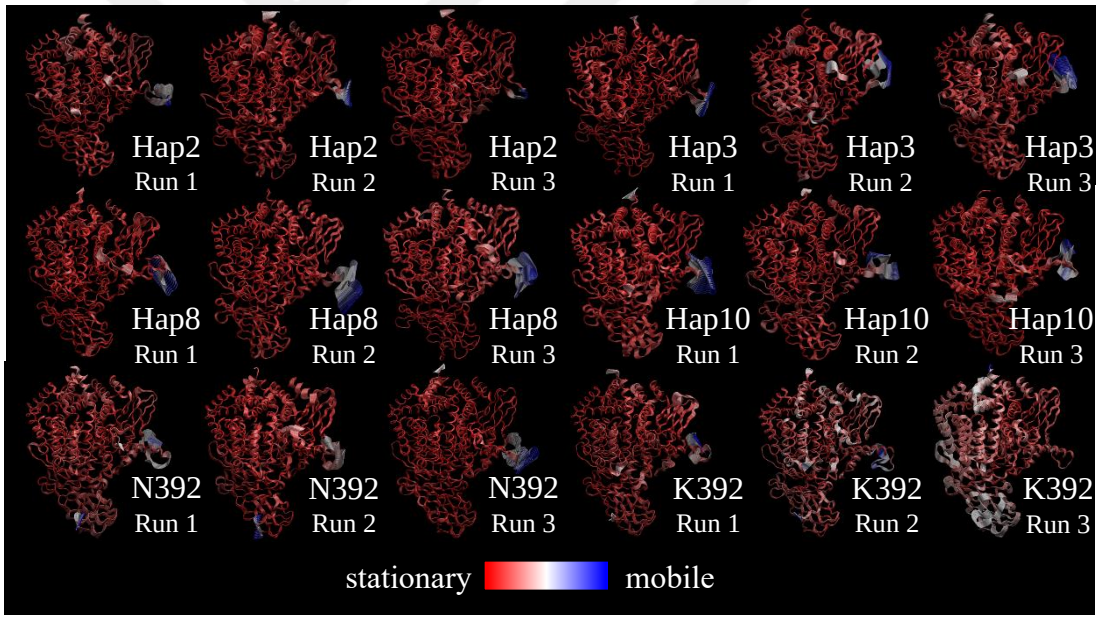



Figure 62. PCA of ERAP1 and ERAP2 closed monomer structures after 150 ns simulation. From red to blue, residual mobility increases. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

PCA results of dimer simulations were evaluated (Figure 63). Most of dimers demonstrated lower levels of stability compared to monomers. Only a few Hap2–Hap2 runs retained their stability in optimal levels. In contrast, all runs of Hap2–N392 except the third run, run 1 and 3 of Hap8–K392 and run 2 of Hap8–Hap8 were in extreme

motions. Among unstable runs, many heterodimers demonstrated separation of chains whereas many homodimers showcased changes in conformational state of individual chains.

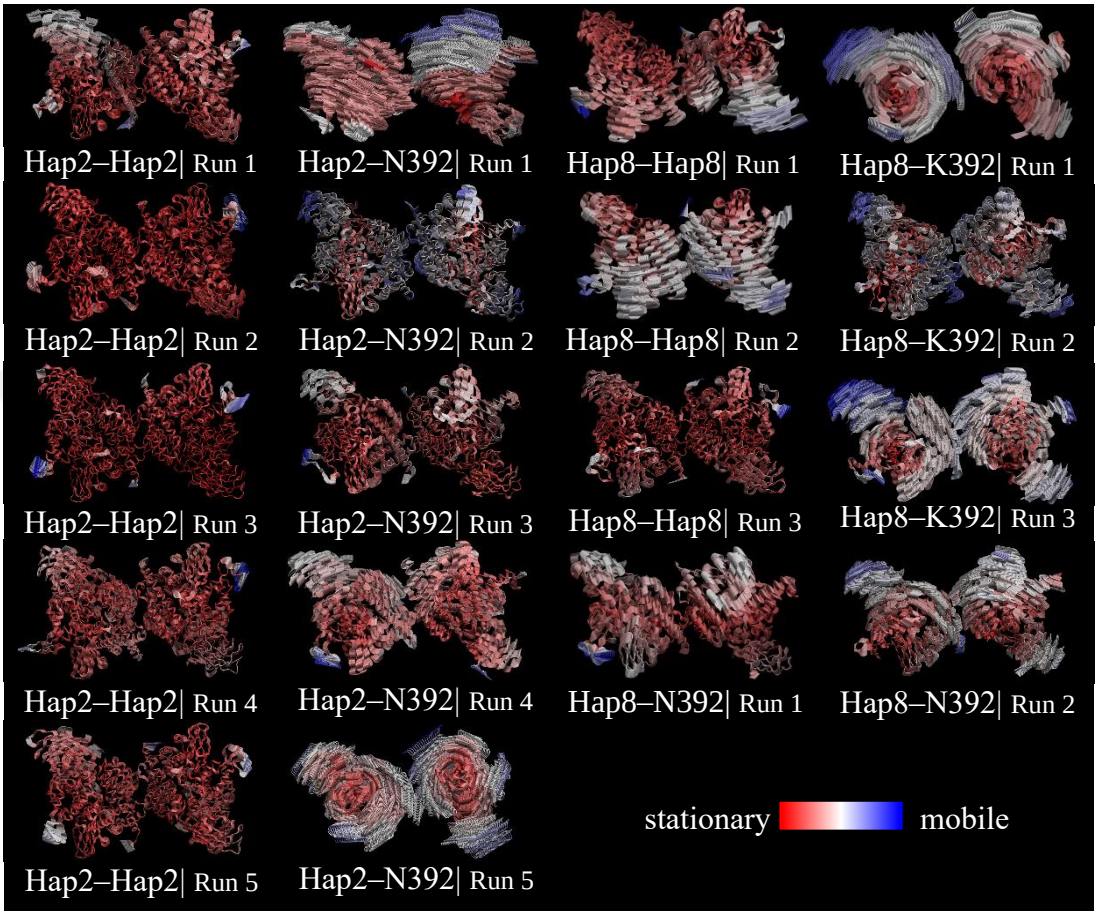



Figure 63. PCA of ERAP1 closed homodimer and ERAP1/2 closed heterodimer structures after 150 ns simulation. From red to blue, residual mobility increases. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

## 4.7 Catalytic Tyrosine in Active Site of ERAP1 and ERAP2 is Subject to Conformational Difference

Active site motions of simulations were captured in 50-ns time intervals. Beside active site residues (residues in GAMEN motif and Tyr438/Tyr455); Zn(II) ion, zinc binding residues (His353/His370, His357/His374, Glu376/Glu393) and Phe433/Phe450 were included for representation. Results of some ERAP1 primary

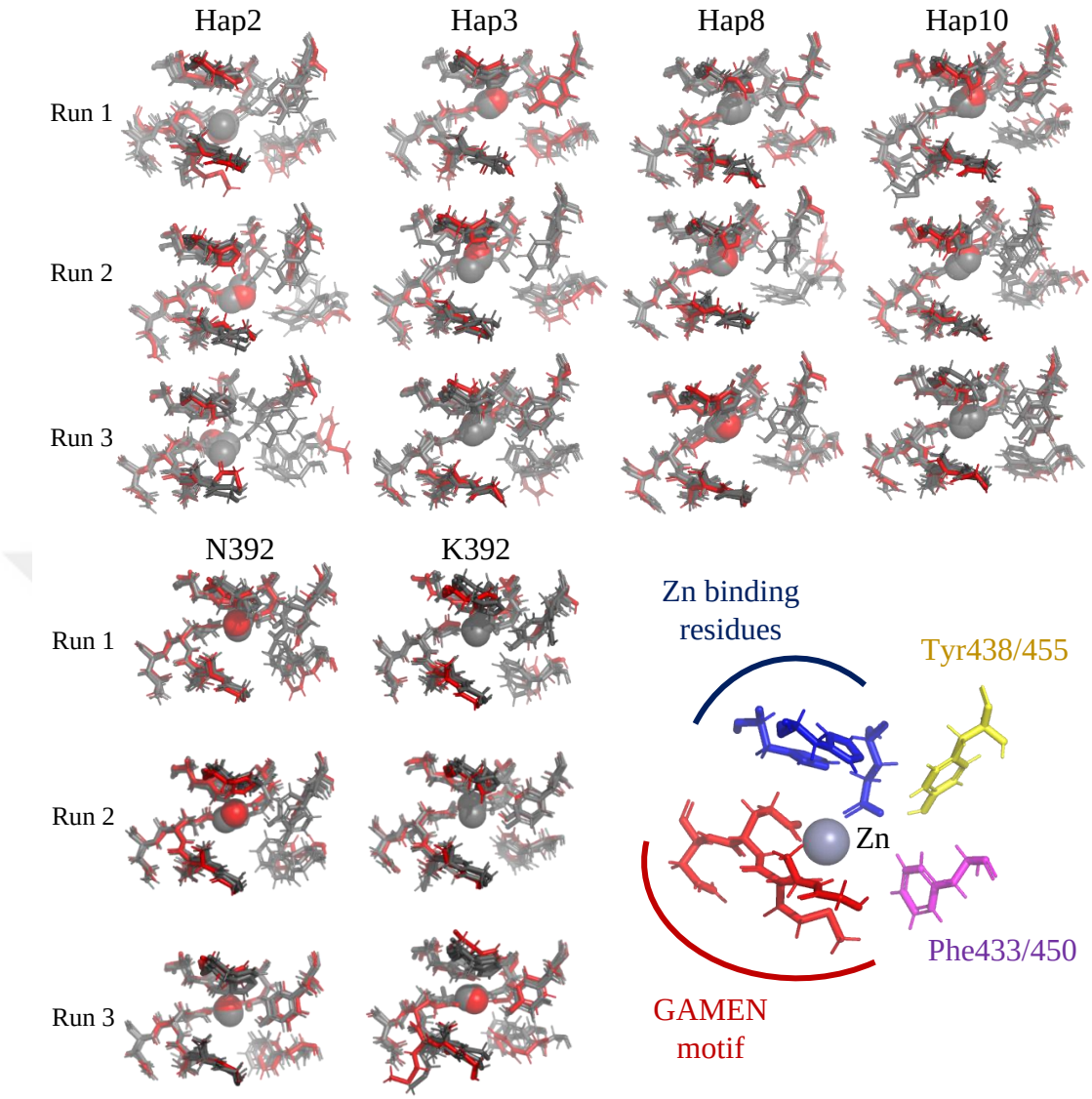



Figure 64. Active site positions of ERAP1 closed primary allotypes (top) and ERAP2 closed allotypes (bottom left) before simulation, at 50, 100 and 150 ns of simulation. Structures at 150 ns are represented in red while structures of other time frames are represented in grey. Locations of residues were demonstrated (bottom right). Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

allotypes showed reorientation and rotation of Phe433 and Tyr438 residues especially at 150 ns (Figure 64). Phe motions were significant in run 3 of Hap2 and run 2 of Hap8. Rotation of Tyr438 was clearly observed in all runs of Hap2, run 2 of Hap3, run 2 and 3 of Hap8 and run 2 of Hap10. In ERAP2 allotypes, only slight motions of Phe450 were prevalent while rotation in Tyr455 was caught in run 1 and 2 of N392 and run 1

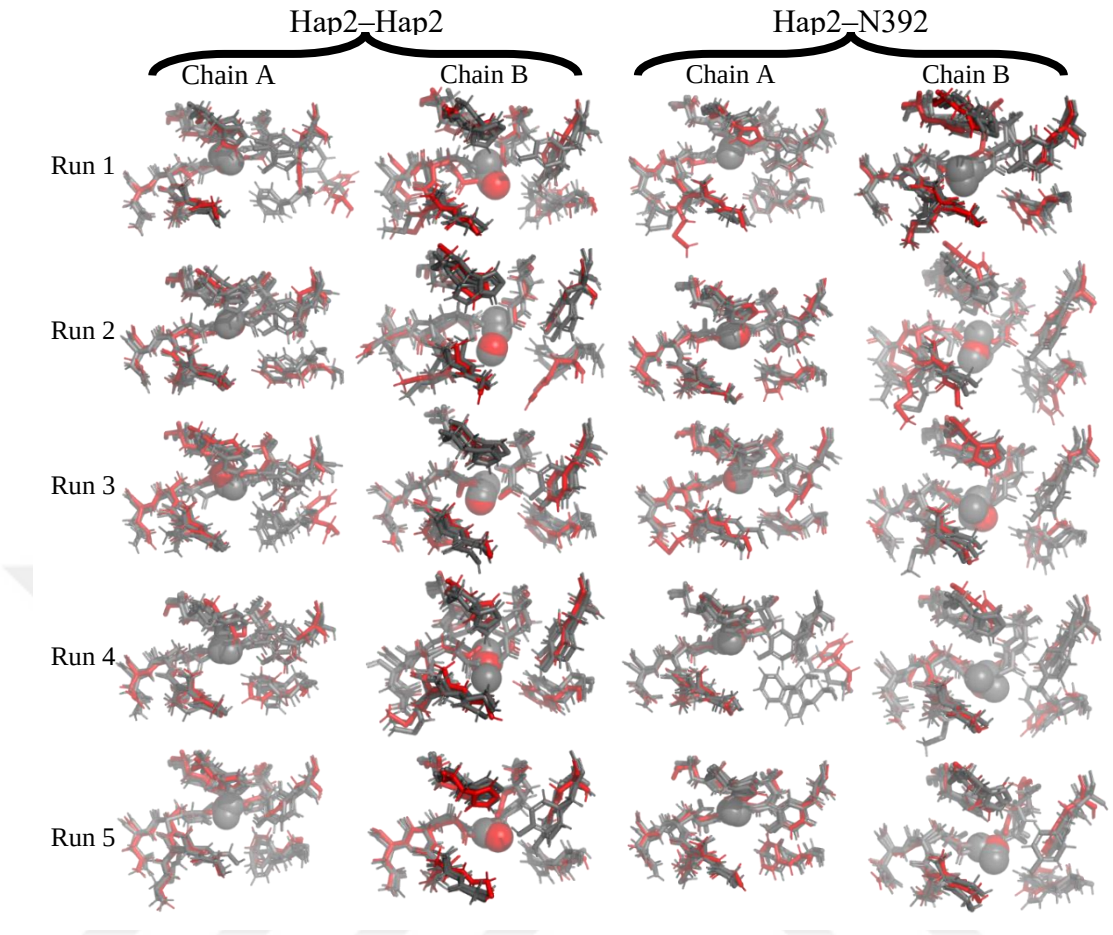



Figure 65. Active site positions of ERAP1 Hap2–Hap2 closed homodimer (left) and ERAP1/2 Hap2–N392 (right) closed heterodimer structures before simulation, at 50, 100 and 150 ns of simulation. Structures at 150 ns are represented in red while structures of other time frames are represented in grey. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0). Residue locations were demonstrated in Figure 64.

of K392 (Figure 64). No π-π interaction was observed between Phe433/450 and Tyr438/455 in any run.

Active site positions of ERAP1 Hap2–Hap2 closed homodimer and ERAP1/2 Hap2–N392 closed heterodimer structures were investigated (Figure 65). In chain A of Hap2–Hap2, Phe433 was in significant motion in run 1 and 3. Similar kind of motion was observed in chain B of Hap2–Hap2 in run 2. Rotation of catalytic residue Tyr438 was also significant in many runs of Hap2–Hap2. Overall changes in active sites of Hap2–N392 were not as significant as Hap2–Hap2. Nonetheless, a clear motion was observed in Phe450 from chain A of run 4 and rotation of Tyr450 in chain

A and B of run 3 as well as chain B of run 4 and 5. Interestingly, shifting of Phe450 in ERAP2 chains was fairly low.

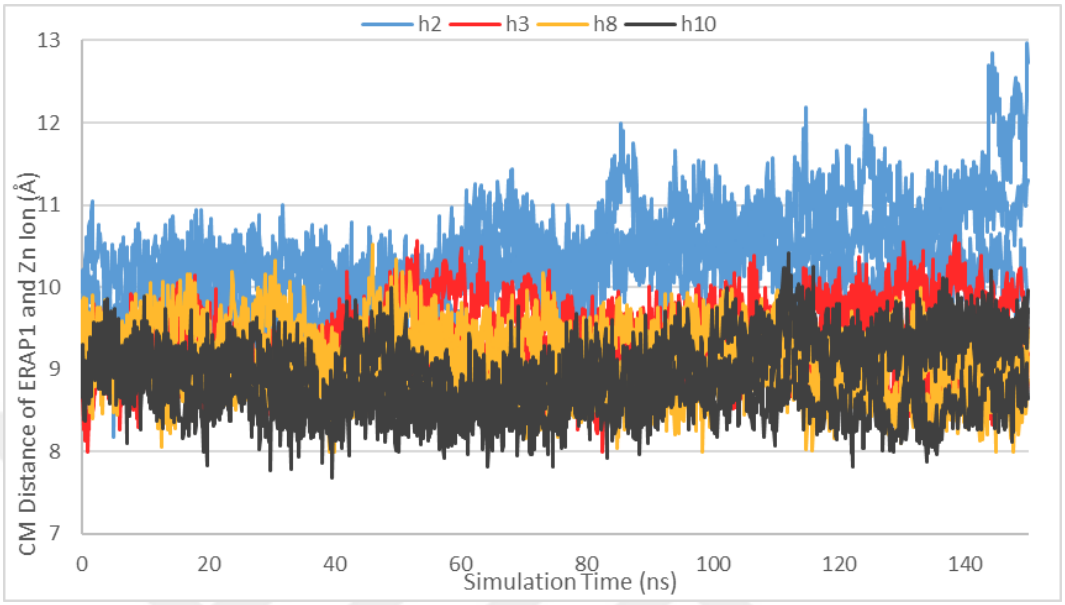



Figure 66. CM distance between protein and Zn(II) ion in ERAP1 closed Hap2 (blue), Hap3 (red), Hap8 (orange) and Hap10 (grey) monomer structures throughout simulation. Reference structure was used for ERAP1 (PDB ID: 2YD0).

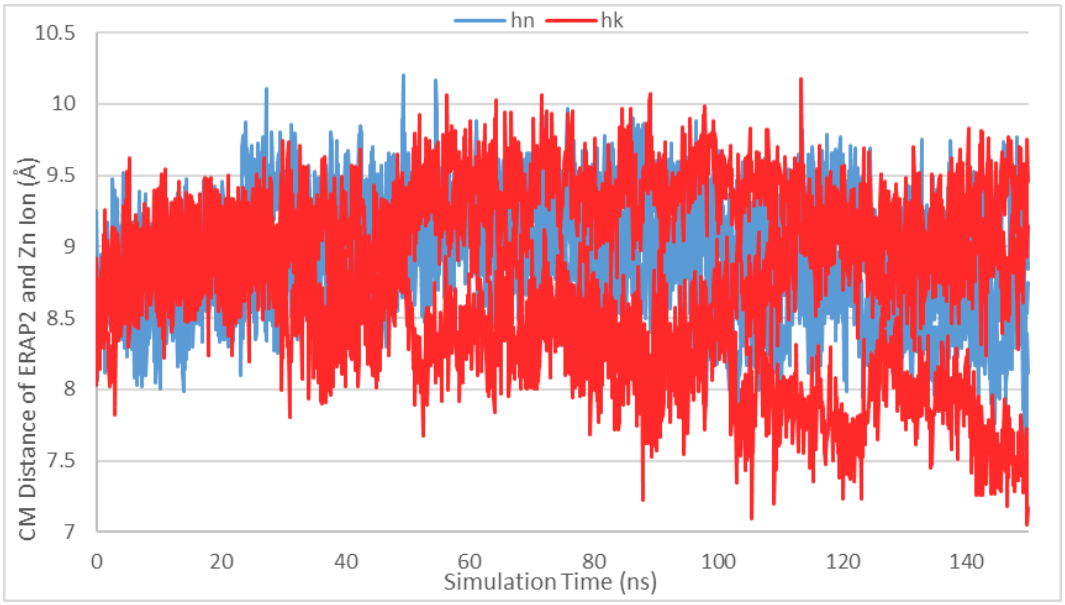



Figure 67. CM distance between protein and Zn(II) ion in ERAP2 closed N392 (blue) and K392 (red) monomer structures throughout simulation. Reference structure was used for ERAP2 (PDB ID: 5AB0).

Distance between CM of protein chain and Zn(II) was calculated for all closed monomers. In ERAP1 primary allotypes, Hap2 distances were marginally higher than other allotypes and distance of run 2 and 3 from the same allotype gradually increased, albeit slowly (Figure 66). Finally, ERAP2 allotypes were mostly in similar trends as stable runs of ERAP1 (Figure 67). Here, K392 run 3 distance decreased by the end of simulation.

Zn distance to CM of protein chain was analyzed in dimers Hap2–Hap2 and Hap2–N392. In overall, chain A distance was a little bit more steady than chain B distance. Chain A distance in run 5 of Hap2–Hap2 and run 4 of Hap2–N392 marginally increased (Figure 68). Chain B distance in run 2 of Hap2–Hap2 marginally increased while, in run 5 of Hap2–Hap2, it decreased (Figure 69). All runs except run 5 of Hap2–N392 demonstrated a significant increase in chain B distance as simulation time passed.

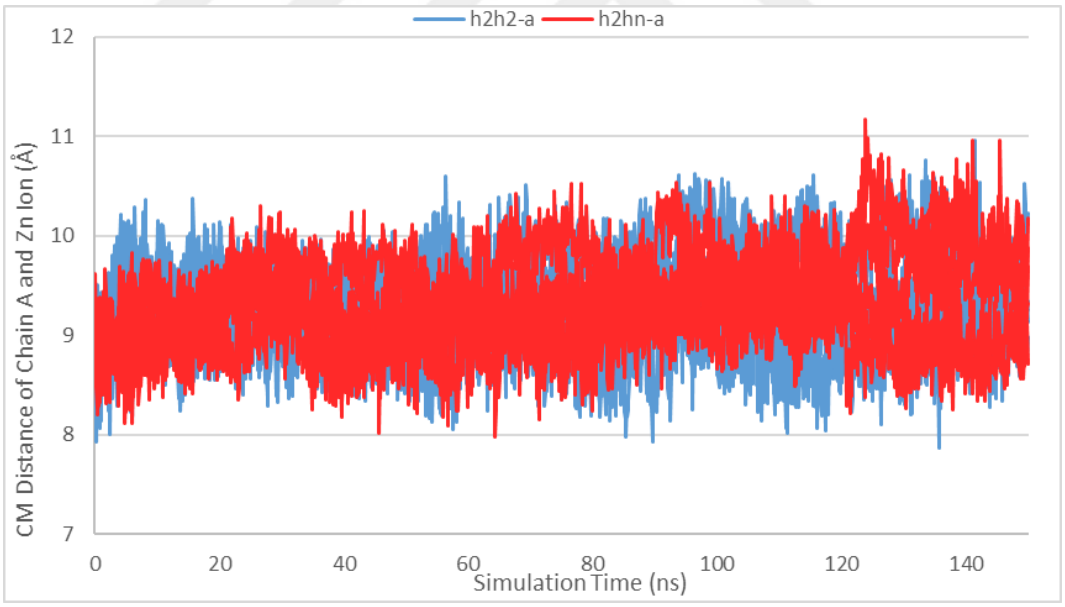



Figure 68. CM distance between protein and Zn(II) ion of the first chain (chain A) in ERAP1 closed Hap2–Hap2 homodimer (blue) and ERAP1/2 closed Hap2–N392 heterodimer (red) structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

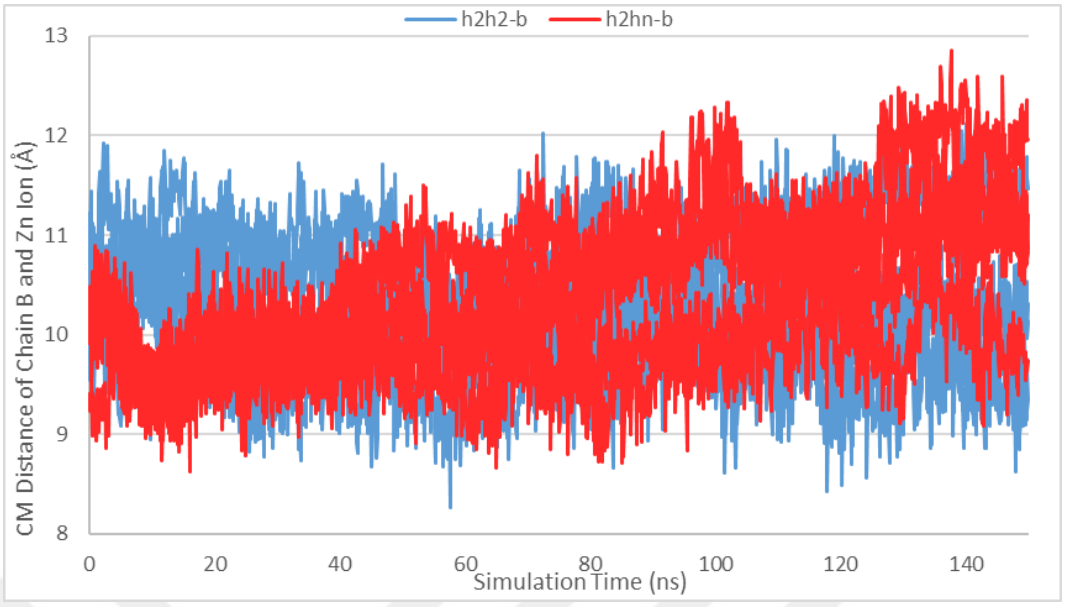



Figure 69. CM distance between protein and Zn(II) ion of the second chain (chain B) in ERAP1 closed Hap2–Hap2 homodimer (blue) and ERAP1/2 closed Hap2–N392 heterodimer (red) structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

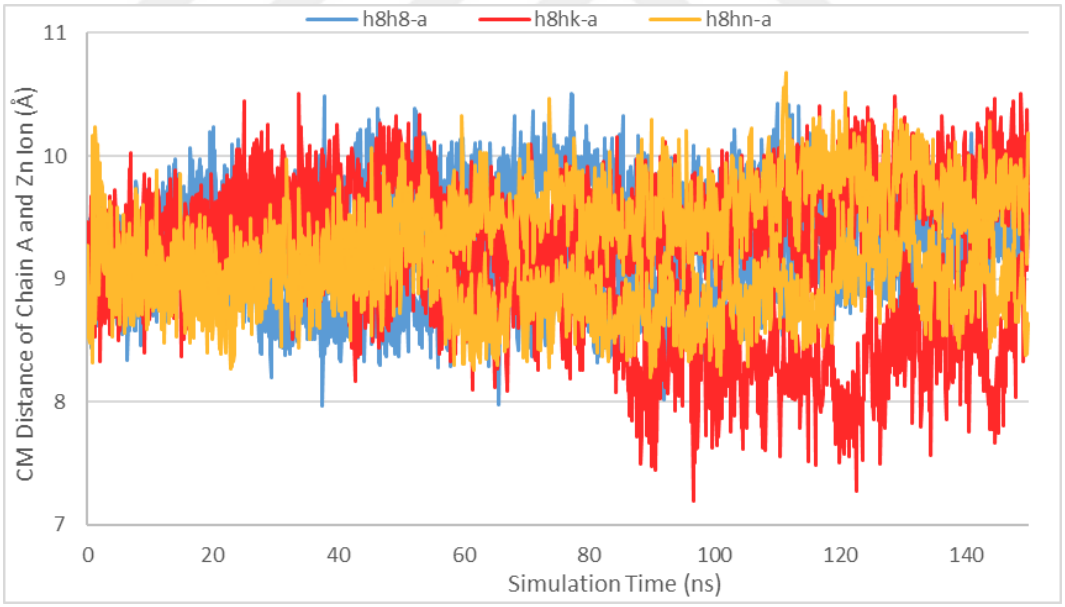



Figure 70. CM distance between protein and Zn(II) ion of the first chain (chain A) in ERAP1 closed Hap8–Hap8 homodimer (blue), ERAP1/2 closed Hap8–K392 (red) and Hap8–N392 (orange) heterodimer structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

Hap8 homo- and heterodimers were evaluated in terms of Zn(II) and protein CM distance. In chain A, changes were minimal in many Hap8–Hap8 runs with about 1

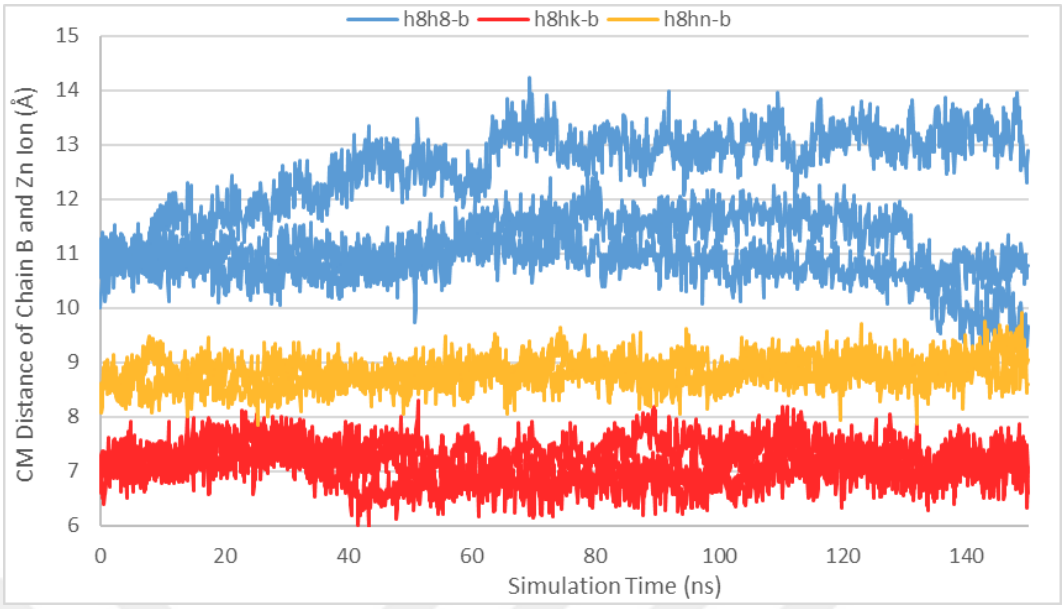



Figure 71. CM distance between protein and Zn(II) ion of the second chain (chain B) in ERAP1 closed Hap8–Hap8 homodimer (blue), ERAP1/2 closed Hap8–K392 (red) and Hap8–N392 (orange) heterodimer structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

angstrom difference (Figure 70). Even so, swinging patterns were present in some heterodimers such as Hap8–K392. Distinctively, run 3 of Hap8–K392 showed a temporary fall then regained its initial distance values just before the end of the simulation. Runs of Hap8–Hap8 only expressed significant changes in Chain B distances (Figure 71). Strangely enough, Hap8–K392 runs had the lowest distances in chain B followed by more stable Hap8–N392 runs. These distances were significant against homodimer distances in chain B.

C-alpha distance between Tyr438 and other active site residues was inspected in Hap2–Hap2 and Hap2–N392 simulations (Figure 72, 73, 74 and 75). In chain A of Hap2–Hap2 homodimer, the most notable change was found between residues Tyr438 and Gly317 with minimal varying distances among parallel runs. The same change applied to chain A of Hap2–N392 with conjugate residues Tyr455 and Gly334. Other residue pairs were more stable. Chain B active site distances were generally similar to chain A results in Hap2–Hap2. However, the second run of Hap2–Hap2 had a sharp decline after 90 ns in the simulation with the residue pairs 317-438 and 318-438. Additionally, distance of residue pair 321-438 in Hap2–Hap2 run 1 began to fall down.

As for Hap2–N392, the third run had the least consistent distance values during simulation and all runs displayed undeniable changes. This time, residue pairs besides 334-455 were diverse in terms of distance.

Distance of Tyr438/Tyr455 to other active site residues was investigated in Hap8–Hap8 homodimer, Hap8–K392 and Hap8–N392 heterodimers (Figure 76, 77, 78, 79, 80 and 81). Compared to dimers with Hap2, distance of Tyr438/Tyr455 to Gly317/Gly334 was more stable in chain A of Hap8–Hap8 and Hap8–N392 runs. Stability of chain A in Hap8–K392 runs was only marginally flexible after 40 ns. Changes in chain B were a lot more prevalent in Hap8–Hap8 homodimers than both heterodimers. Runs of Hap8–N392 exhibited almost straight distance values. Similarly, Hap8–K392 runs were stable in terms of distance.

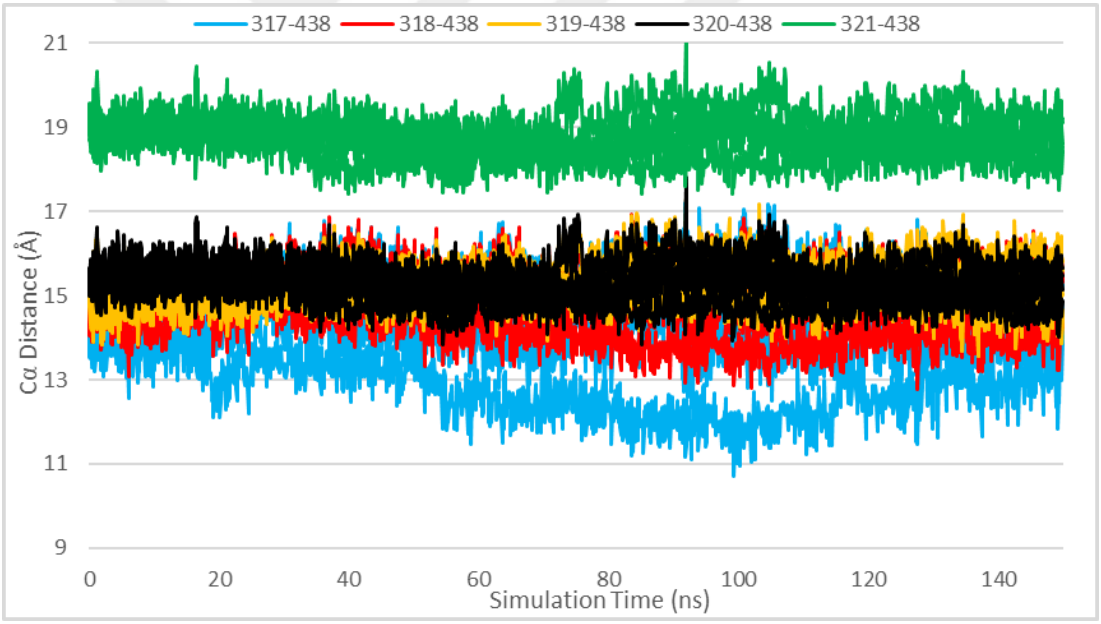



Figure 72. C-alpha distance of Tyr438 against other active site residues in the first chain (chain A) of ERAP1 closed Hap2–Hap2 homodimer structure throughout simulation. Tyr-Gly (cyan), Tyr-Ala (red), Tyr-Met (yellow), Tyr-Glu (black) and Tyr-Asn (green) distances are colored. Reference structure was used for ERAP1 (PDB ID: 2YD0).

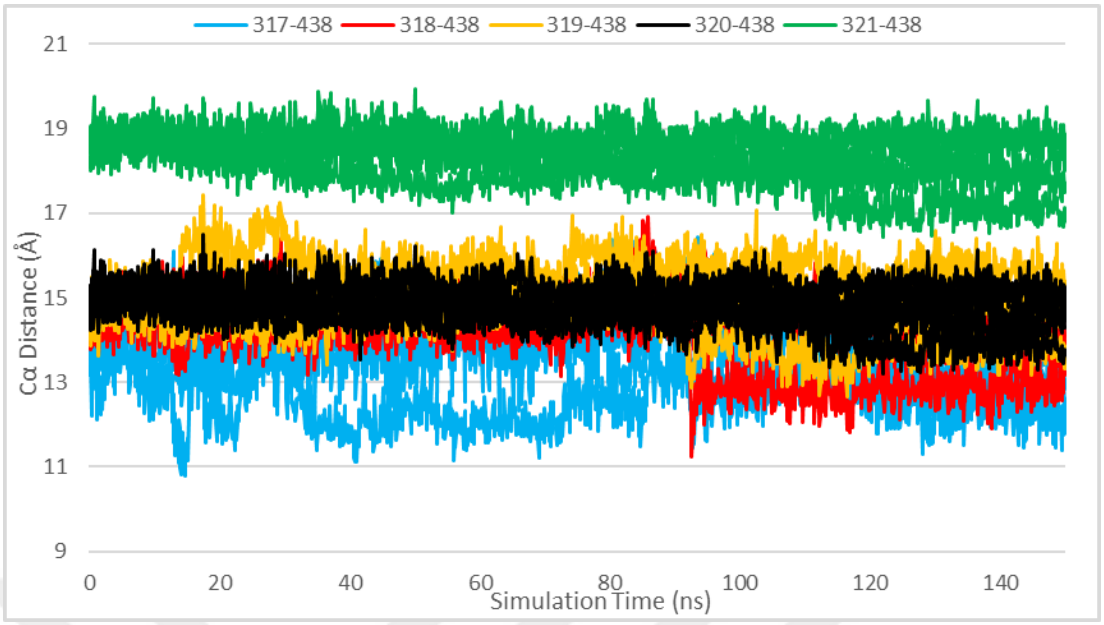



Figure 73. C-alpha distance of Tyr438 against other active site residues in the second chain (chain B) of ERAP1 closed Hap2–Hap2 homodimer structure throughout simulation. Tyr-Gly (cyan), Tyr-Ala (red), Tyr-Met (yellow), Tyr-Glu (black) and Tyr-Asn (green) distances are colored. Reference structure was used for ERAP1 (PDB ID: 2YD0).

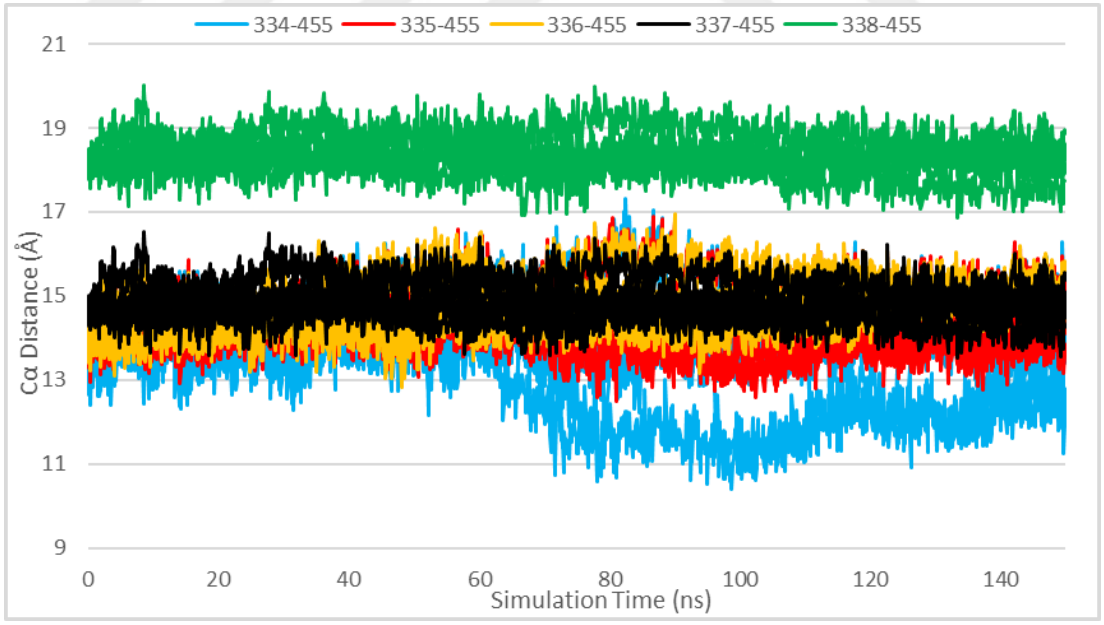



Figure 74. C-alpha distance of Tyr438 against other active site residues in the first chain (chain A) of ERAP1 closed Hap2–N392 heterodimer structure throughout simulation. Tyr-Gly (cyan), Tyr-Ala (red), Tyr-Met (yellow), Tyr-Glu (black) and Tyr-Asn (green) distances are colored. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

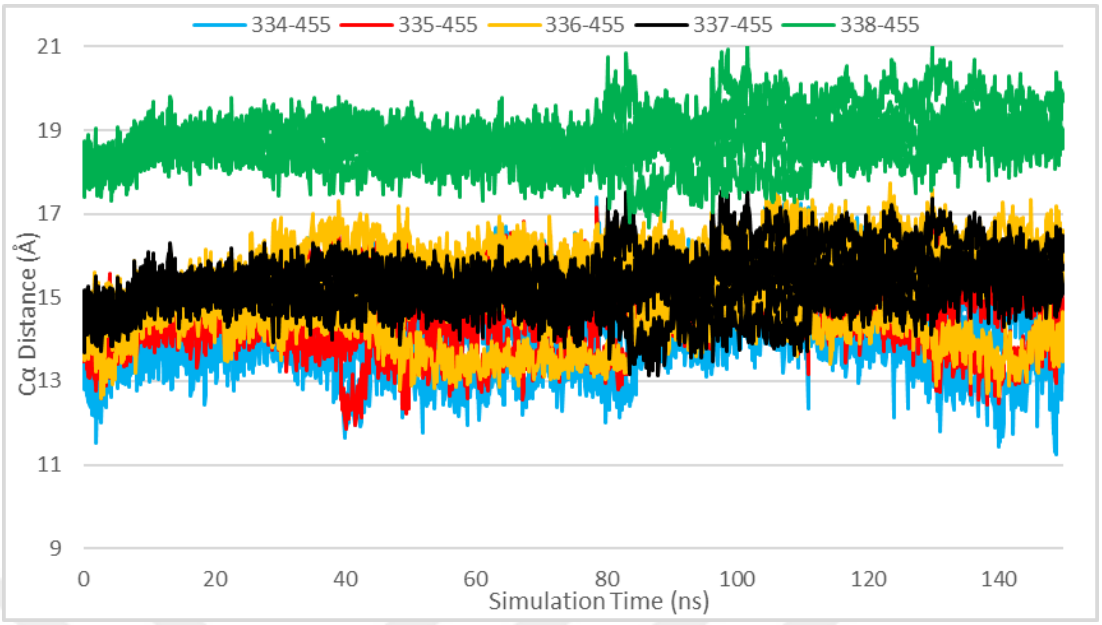



Figure 75. C-alpha distance of Tyr438 against other active site residues in the second chain (chain B) of ERAP1 closed Hap2–N392 heterodimer structure throughout simulation Tyr-Gly (cyan), Tyr-Ala (red), Tyr-Met (yellow), Tyr-Glu (black) and Tyr-Asn (green) distances are colored. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

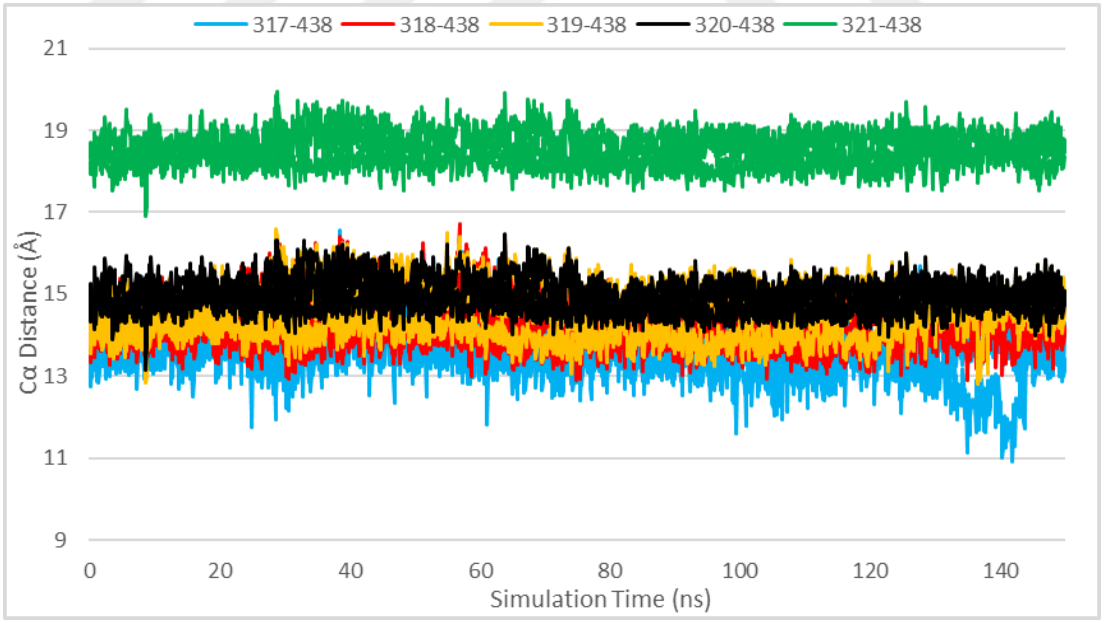



Figure 76. C-alpha distance of Tyr438 against other active site residues in the first chain (chain A) of ERAP1 closed Hap8–Hap8 homodimer structure throughout simulation. Tyr-Gly (cyan), Tyr-Ala (red), Tyr-Met (yellow), Tyr-Glu (black) and Tyr-Asn (green) distances are colored. Reference structure was used for ERAP1 (PDB ID: 2YD0).

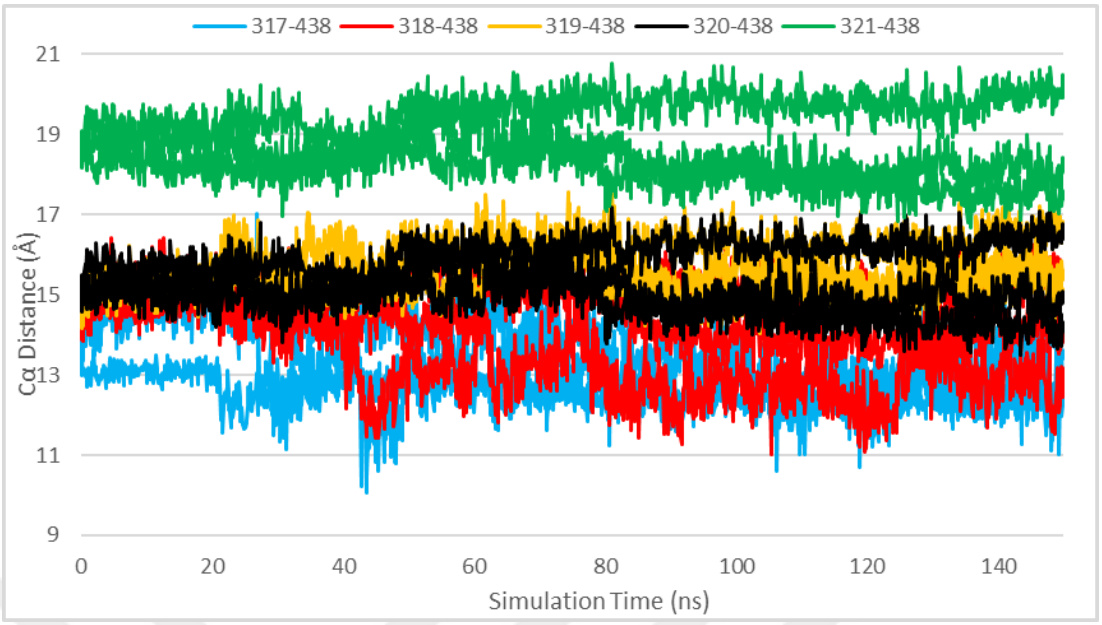



Figure 77. C-alpha distance of Tyr438 against other active site residues in the second chain (chain B) of ERAP1 closed Hap8–Hap8 homodimer structure throughout simulation. Tyr-Gly (cyan), Tyr-Ala (red), Tyr-Met (yellow), Tyr-Glu (black) and Tyr-Asn (green) distances are colored. Reference structure was used for ERAP1 (PDB ID: 2YD0).

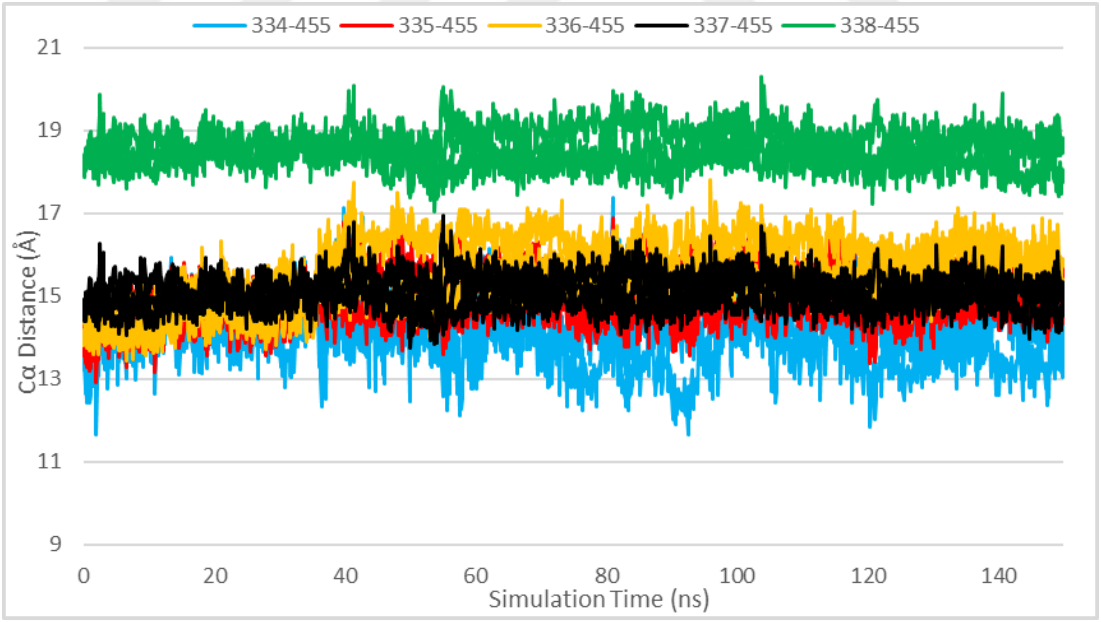



Figure 78. C-alpha distance of Tyr438 against other active site residues in the first chain (chain A) of ERAP1 closed Hap8–K392 heterodimer structure throughout simulation. Tyr-Gly (cyan), Tyr-Ala (red), Tyr-Met (yellow), Tyr-Glu (black) and Tyr-Asn (green) distances are colored. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

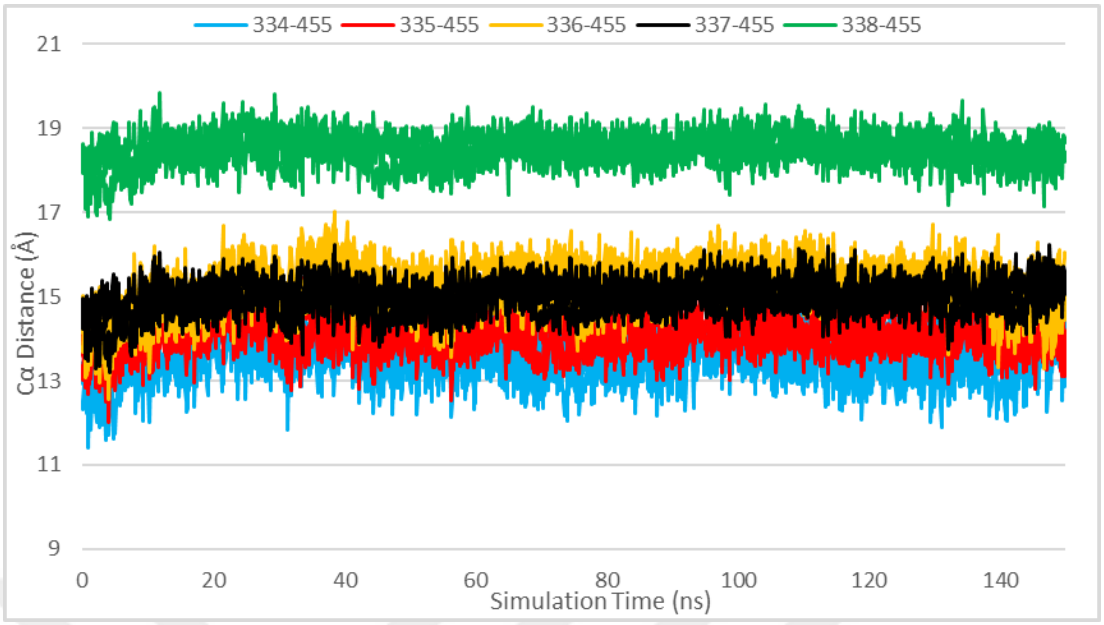



Figure 79. C-alpha distance of Tyr438 against other active site residues in the second chain (chain B) of ERAP1 closed Hap8–K392 heterodimer structure throughout simulation. Tyr-Gly (cyan), Tyr-Ala (red), Tyr-Met (yellow), Tyr-Glu (black) and Tyr-Asn (green) distances are colored. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

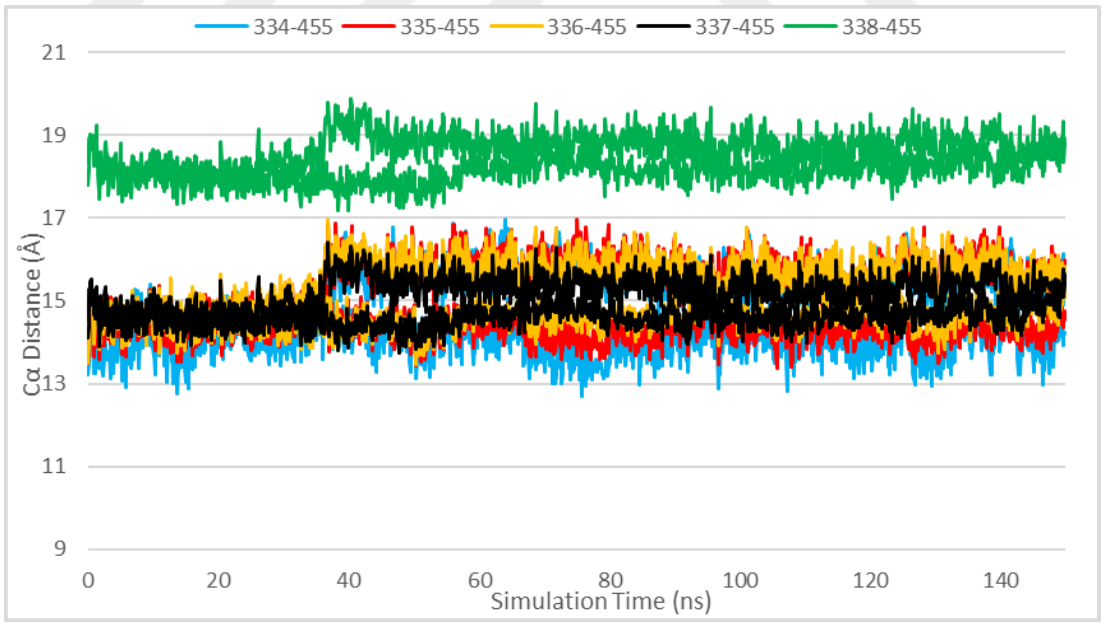



Figure 80. C-alpha distance of Tyr438 against other active site residues in the first chain (chain A) of ERAP1 closed Hap8–N392 heterodimer structure throughout simulation. Tyr-Gly (cyan), Tyr-Ala (red), Tyr-Met (yellow), Tyr-Glu (black) and Tyr-Asn (green) distances are colored. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

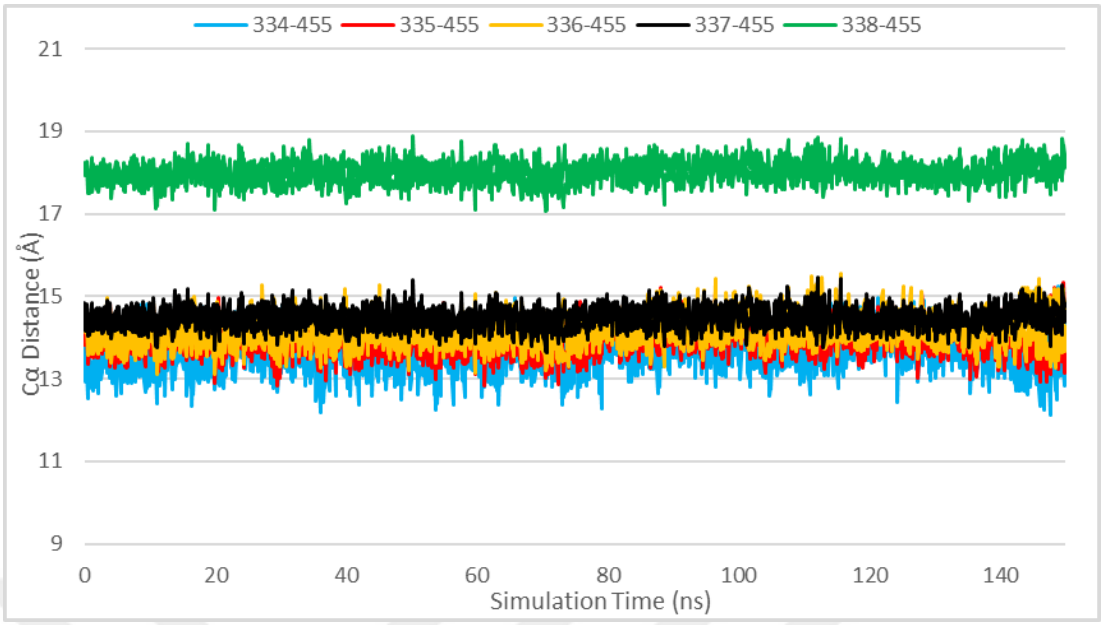



Figure 81. C-alpha distance of Tyr438 against other active site residues in the second chain (chain B) of ERAP1 closed Hap8–N392 heterodimer structure throughout simulation. Tyr-Gly (cyan), Tyr-Ala (red), Tyr-Met (yellow), Tyr-Glu (black) and Tyr-Asn (green) distances are colored. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

### 4.8 Interactions in Dimerization Site Ensure Chain Stability

To investigate structural changes more closely, possible interactions between individual chains of dimer were identified in ERAP1 Hap2–Hap2 homodimer and ERAP1/2 Hap2–N392 heterodimer runs. To this end, salt bridge analysis was performed. First eliminations were done between parallel runs of each dimer combination. Out of many residue pairs responsible of salt bridge interactions, 13 residue pairs were found common among five parallel runs of Hap2–Hap2 (Figure 82). In Hap2–N392, number of common residue pairs dropped to 10 with many parallel runs taken account (Figure 83). In spite of that, Hap2–N392 had more salt bridge pairs than Hap2–Hap2 in total and many of them found uniquely in different parallel runs.

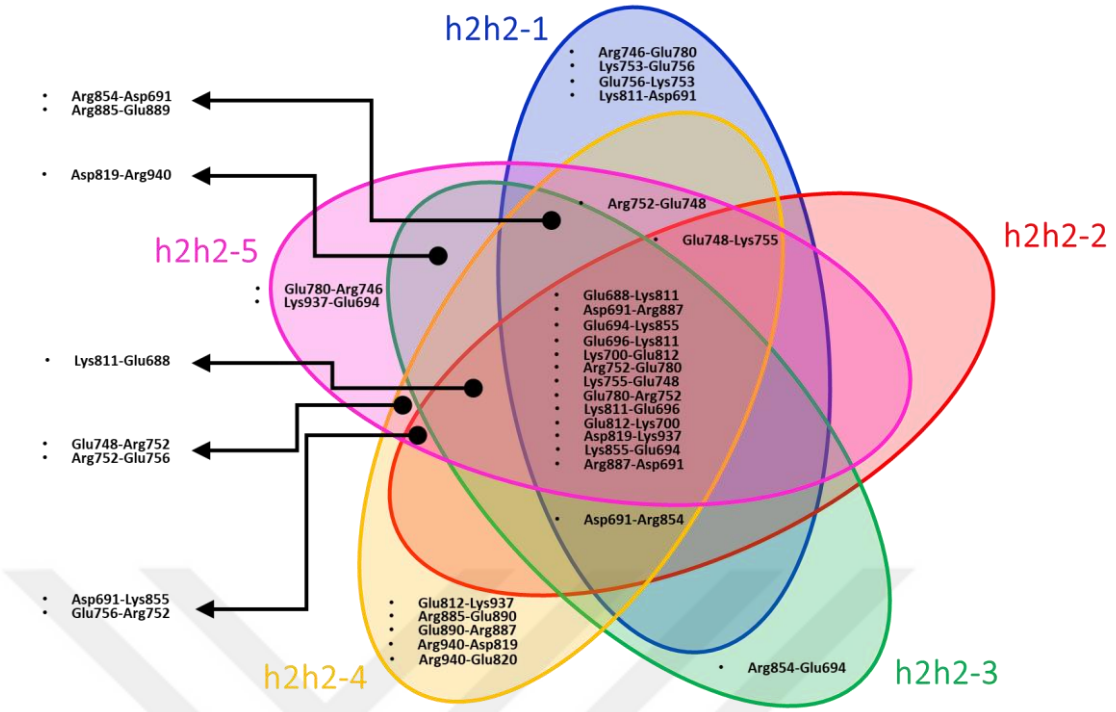



Figure 82. Residue pairs of ERAP1 closed Hap2–Hap2 homodimer structures forming salt bridges between domain IV dimerization sites. Sets are categorized by parallel runs. Reference structure was used for ERAP1 (PDB ID: 2YD0).

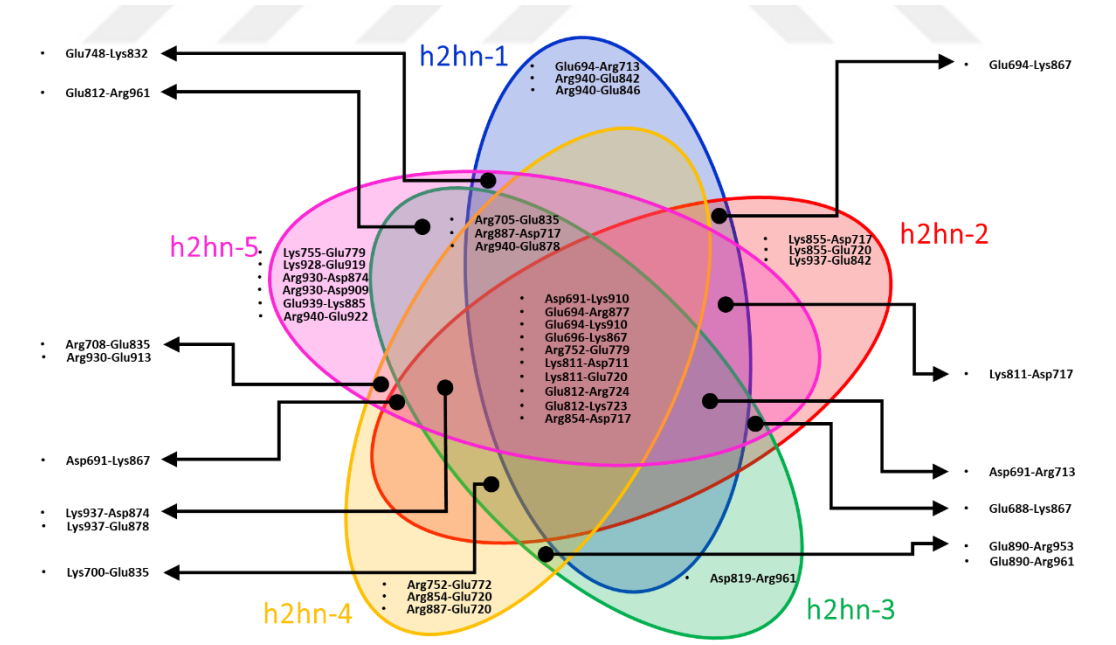



Figure 83. Residue pairs of ERAP1/2 closed Hap2–N392 heterodimer structures forming salt bridges between domain IV dimerization sites**.** Sets are categorized by parallel runs. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

After determination of common salt bridges among parallel runs, common pairs between Hap2–Hap2 and Hap2–N392 structures were investigated (Figure 84). As ERAP1 and ERAP2 sequences do not perfectly match, conjugate residues were identified through sequence and structural alignment. Interestingly, only two conjugate residue pairs were found common in all runs done for Hap2–Hap2 and Hap2–N392.

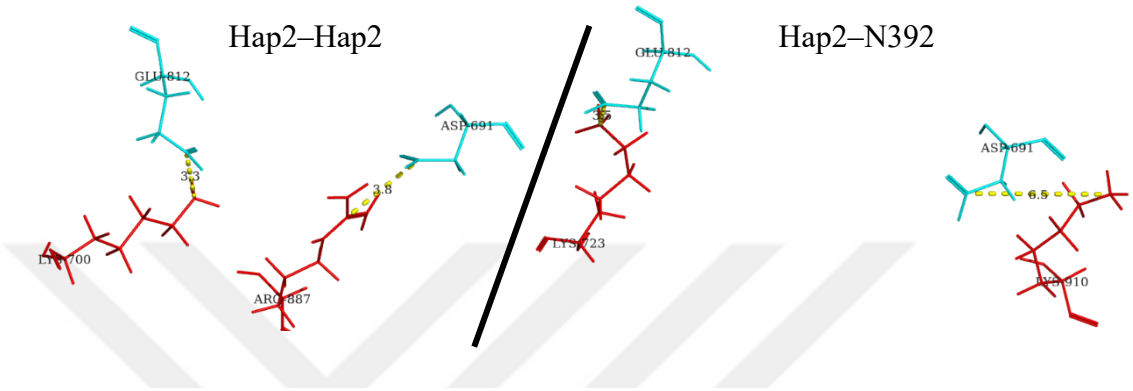



Figure 84. Common and conjugate interchain residue pairs present in both ERAP1 closed Hap2–Hap2 homodimer (left) and ERAP1/2 closed Hap2–N392 heterodimer (right) structures. Residues in chain A (cyan) and chain B (red) are shown. Structural pose and angstrom distances shown in yellow dashes represent only the first frame in the first runs of both structures.

Aside from salt bridges, possible contact residues between chains were identified within 3 Å distance. Average and standard deviation of all existing contacts throughout parallel runs were calculated and the most frequently occurring interactions were combed out. Top 50 residue pairs with great contacts were listed for both Hap2–Hap2 and Hap2–N392. Afterwards, these contacts were classified into either salt bridge or other interaction depending on the residue type. In Hap2–Hap2 results, best contacts were identified in salt bridges most of which was in close contact over 50% of total simulation time (Figure 85). Moreover, top 50 residue pairs were available for contact over 20% of total simulation time. However, results were substantially different in Hap2–N392 contact residues (Figure 86). Generally, top 50 residue pairs were only present under 50% of the total time. The best contacts were not dominated by salt bridges which were scattered around.

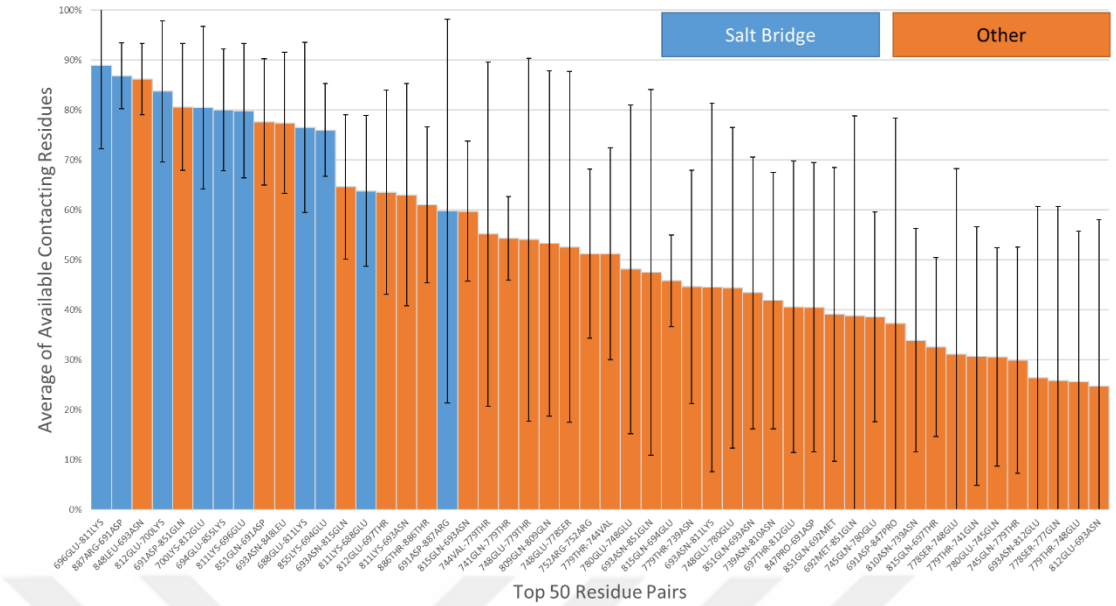



Figure 85. Top 50 interchain residue pairs of ERAP1 closed Hap2–Hap2 homodimer structures. Average of all 5 parallel runs show how likely residue pairs interact to each other. Salt bridge (blue) and other interactions (orange) are shown. Reference structure was used for ERAP1 (PDB ID: 2YD0).

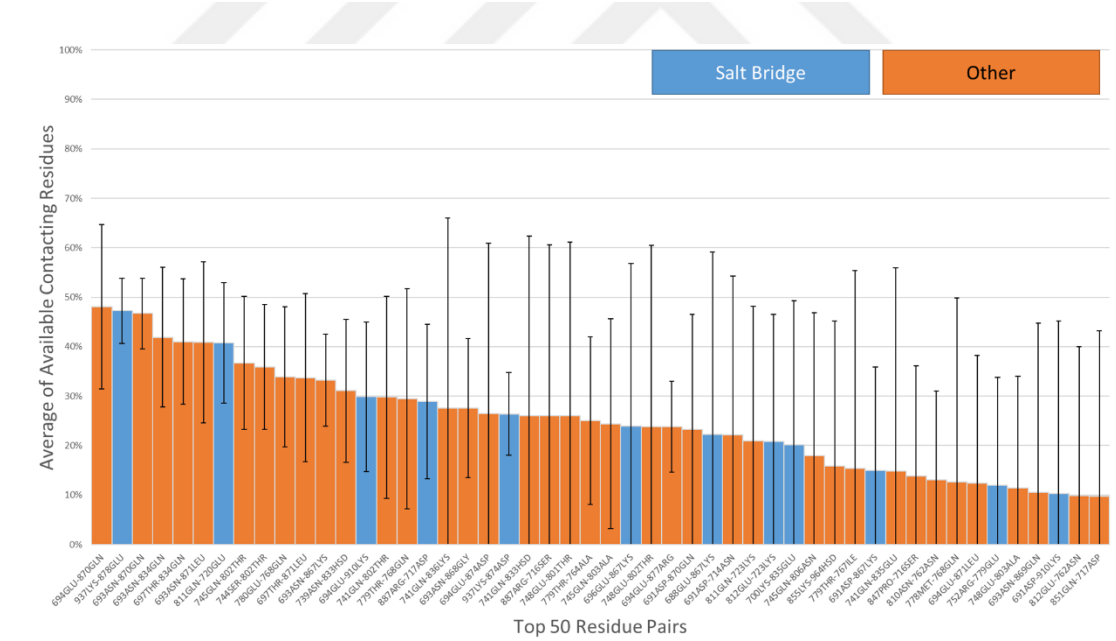



Figure 86. Top 50 interchain residue pairs of ERAP1/2 closed Hap2–N392 heterodimer structures. Average of all 5 parallel runs show how frequently residue pairs interact with each other. Salt bridge (blue) and other interactions (orange) are shown. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

A type of weak interaction called H-bond was also analyzed in Hap2–Hap2 and Hap2–N392 simulations (Figure 87). A clear difference was observed between Hap2–Hap2 and Hap2–N392 runs. Number of H-bonds in Hap2–Hap2 runs mostly remained stable through time whereas number of H-bonds in Hap2–N392 were much less stable causing either staggering in interaction and low number of bonds. While H-bond number in run 1 of Hap2–N392 completely diminished in the second half of the simulation, it dropped to near zero in run 5 of Hap2–N392. Taking standard deviation of all runs into consideration, an almost seamless division between the number of H-bonds in Hap2–Hap2 and Hap2–N392 was realized.

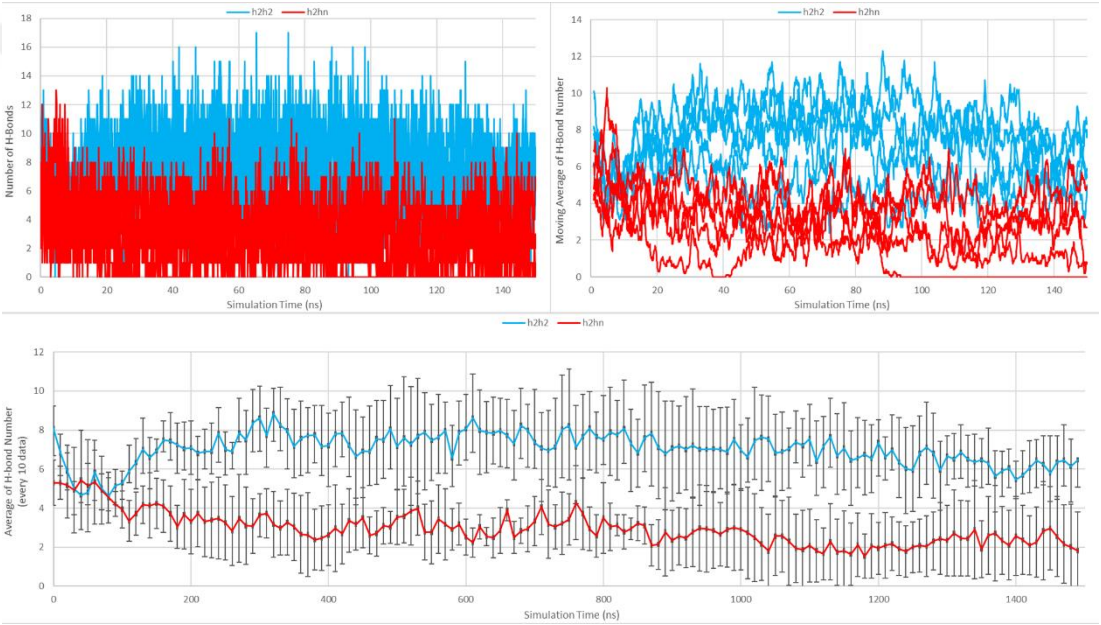



Figure 87. Number of H-bonds (top left), moving average of number of H-bonds (top right) and average of H-bond number (bottom) in ERAP1 closed Hap2–Hap2 homodimer (cyan) and ERAP1/2 closed Hap2–N392 heterodimer (red) structures throughout simulation. Reference structures were used for ERAP1 (PDB ID: 2YD0) and ERAP2 (PDB ID: 5AB0).

# 5 DISCUSSION

AS is a chronic inflammatory autoimmune disease causing inflammation in joints (1). While AS pathogenesis is unclear, ERAP1 protein was found as a risk factor of the disease (3). It was hypothesized that ERAP1 and ERAP2 affect MHC-I immunopeptidome due to polymorphic differences in their structure and alter catalytic activity and peptide trimming efficiency. Selected ERAP1 and ERAP2 allotypes in monomer, homodimer and heterodimer forms were simulated and analyzed in the aim of deciphering their possible contribution to AS mechanism.

## 5.1 Interpretation of ERAP Structural Differences

In experimental part of the project, Hap8, Hap3 and Hap2 allotypes were discovered, in this order, as top 3 abundant allotypes in Turkish AS patients (unpublished results). This is almost in line with the results of one study about the frequency of ERAP1 allotype combinations from “1000 Genomes Project” which showed Hap2–Hap2 and Hap2–Hap8 in high frequency (58). In previous studies, Hap2 and Hap3 were associated with AS as AS risk factors, Hap10 was associated with AS as AS protective factor and no AS association was made for Hap8 in GWAS (53). Furthermore, Hap10 was found as a destabilizing allotype in preliminary protein stability prediction. Due to these reasons, these four allotypes were selected for further analysis.

In the findings of this study, Hap2, Hap3, Hap8 and Hap10 ERAP1 closed monomers showed subtle changes in stability of the structures in most of the simulation time. These changes occurred with minor allotypic differences. In ERAP2 closed monomers, N392 and K392 allotypes demonstrated slightly more stable structures than ERAP1 monomers. Larger conformational changes of ERAP1 and ERAP2 monomers in between different allotypes were almost non-existent (or minimal in some Hap3 and K392 structures in domain I and IV) and these changes did not remain stable in the course of the simulations. Additionally, high motions observed

in monomer structures were not located close to individual SNPs. This may explain why structural differences were not diverse.

Exon 10 coding sequence has ER retention signal which retains ERAP1 in ER (23). Only the sequence encoded by exon 10 coding sequence (485–508) was very flexible in all monomers. This long sequence was one of few unstructured sequences in ERAP1 crystal models that were remodeled in this study. However, this sequence was present in ERAP2 crystal structure and had a partial α-helix secondary structure. While ERAP1 and ERAP2 sequences around this region were not highly conserved, remodeled ERAP1 exhibited a loop as a secondary structure in this sequence. This was further verified by I-TASSER generated models, showing the formation of an alpha-helical structure with the lowest probability. The grafted ERAP2 exon 10 loop structures on ERAP1 by Papakyriakou *et al.* exhibits a long alpha-helical region and their study bases the homo- and heterodimers mainly based on the presence of this helix as they used only existing crystal structures. Interestingly, in this thesis, ERAP2 α-helix in the same region bended in some structures during the simulation and showed more flexibility from original crystal structure.

Beside ERAP1 and ERAP2 monomer structures, ERAP1 and ERAP2 were shoen to be found in heterodimer structures *in vitro* and *in vivo* as well (19). In literature, ERAP1–ERAP2 heterodimer exhibited an increase in ERAP1 substrate affinity at the cost of loss in ERAP2 substrate affinity (49). However, no interaction site for both chains were known that could make a consistent ERAP1–ERAP2 heterodimer structure to analyze. This prompted to research proteins of the same family that form a similar dimer structure. In one study, crystal structure of APN homodimer was provided with possible dimerization sites indicated (25). As APN is another protein from M1 aminopeptidase family, its interaction sites were used as a reference for ERAP1 homodimer and ERAP1–ERAP2 heterodimer docking. This site is located on the surface of domain IV of the original APN structure and predicted to appear in domain IV of both ERAP1 and ERAP2 structures (22).

By the time this study was finalized, another research group completed a computational analysis of ERAP1/2 heterodimer using ERAP2 and IRAP homodimer models besides APN homodimer (51). They found out a unique interaction site involving exon 10 loops of chains in ERAP2 homodimer (specifically, between the chain in asymmetric unit and the symmetry-related chain). After substitution of homodimer chains with ERAP1 and ERAP2, heterodimer templates were created. While initial heterodimer structure referenced by ERAP2 homodimer showed lower binding affinity (higher estimated binding free energy, $\Delta G_{est}$) indicating weaker attraction, rearrangement of B2 structure (ERAP2 in chain A and ERAP1 in chain B) within the simulation caused a change in dimeric interaction, increasing buried surface area (BSA). On the other hand, heterodimer templates with binding poses of IRAP and APN homodimers demonstrated lower BSA along the simulation. However, chain interactions around exon 10 loop found in the crystal structure of ERAP2 homodimer is speculative as it may only form upon crystallization as a crystallization artifact. Moreover, the short glycan chains were included in the dynamics. Inclusion of glycans to these dynamics studies could form a basis for more realistic simulations of these proteins. However, structurally resolved ERAP1 and ERAP2 proteins were heterologously produced in insect hosts (*Trichoplusia ni* in ERAP1 2YD0 and ERAP2 5AB0). Glycan motifs could be different among species affecting interactions between ERAP proteins. Therefore, the true extent of the glycan structures were still debatable. Meanwhile, ERAP1 exon 10 loop was rebuilt by using the secondary structure of ERAP2 in the same study. This rebuilt region has an alpha-helical structure between 505–513 amino acids which is in contrast to secondary structure prediction results as they were predicted with no stable secondary structure in ERAP1 by multiple secondary structure prediction algorithms and modeling programs. One drawback in the aforementioned study is the direct grafting of crystal structure binding modes from ERAP2, IRAP and APN proteins on ERAP1-2 instead of predicting these potential binding modes through molecular docking methods. Since binding poses in the dimer interface of the crystals could be misleading as each crystallization requires a specific set of chemicals to achieve and/or enforce good diffracting crystal formation and individual orientations of the proteins forming these crystals could be far from their original states. Supportively, our docking results from five kinds of ERAP2 and

ERAP1 heterodimer structure demonstrated a better binding affinity when dimerized at domain IV region in opposition to exon 10 loop region. Beside crystal structures, ERAP2 and ERAP1 chains were linked by Fos/Jun zipper tags in the same study. As interactions from domain I of ERAP2 and domain III of ERAP1 were enforced by zippers, it does not allow for the real interactions in heterodimerization. In exon 10 loop interactions, domain I of ERAP2 and the hinge domain, domain III, of ERAP1 also joined the interaction in which conformational dynamics and substrate access to active site could be affected. Conformational changes via interactions at domain IV could be far less due to fewer restrictions in angular motions between domains and substrate binding regions. According to BSA results of APN templated ERAP1/2 heterodimer in this study and SASA results of our ERAP1/2 heterodimers in current study, there is a consistency between the results confirming decreased stability in both heterodimer structures with similar binding poses. Unfortunately, there were not enough analyses (e.g. RMSF and salt bridges) by this study to confirm these results further. While remaining computational studies were performed for only ERAP1 monomers in synergism with experimental methods, they used conformational analyses such as radius of gyration and theta angle to consolidate their results (31,91). As computational methods can only predict dimeric interactions so much, experimental techniques such as cryogenic electron microscopy (cryo-EM) are more reasonable to discover heterodimerization of ERAP1 and ERAP2 structures.

For the observation of impact by homo- and heterodimerization, Hap2 and Hap8 allotypes of ERAP1 closed structures as well as N392 and K392 allotypes of ERAP2 closed structures were assigned in dimer formation. At first, Hap2–Hap2 and Hap2–N392 structures were analyzed. Significantly, Hap2–N392 heterodimer structures were much more flexible than Hap2–Hap2 structures. In many Hap2–N392 heterodimer structures, great loss of interaction reported between ERAP1 and ERAP2 chains. Some heterodimers had displaced chains which caused different contacting residues than from the beginning. Meanwhile, Hap2–Hap2 homodimer structures were quite intact in terms of chain interaction. These observations were quite different in the case of Hap8. Intriguingly, Hap8–Hap8 homodimers showed a lot of movements in comparison to their Hap2–Hap2 counterpart. Structure of Hap8–K392 heterodimer

was slightly more flexible than Hap8–N392 heterodimer structure and as flexible as Hap2–N392 structure. Interactions between heterodimer chains largely vanished. It is important to note that Hap2 homodimers were all fairly stable despite the fact that Hap2 dimers were run 5 times and Hap8 dimers were run only 3 times (except Hap8–N392 which was run twice due to insufficient time) in this study. This suggests Hap2 and Hap8 may impact stability in different ways in homo- and heterodimerization. One expectation was that dimerization site would stabilize domain IV of both homodimer and heterodimer structures. However, this was in contrary to the results.

Conformational changes in ERAP1 proteins usually include motions around domain I and IV region. In transition to ERAP1 open conformation, domain IV opens up and inclines away from the active site that it exposes the active site to the solvent (20). Although only a few chain interactions were lost, one of the chains in several Hap2–Hap2 structures was prone to a more opened up conformation. This was relevant to motions around domain I and IV regions. Conformational changes in Hap2–N392 were less-occurring and ERAP2 chain of heterodimers had lower flexibility. Conformational changes were more common in Hap8 dimer structures. One chains of many structures were close to semi-open state. This was even more prevalent in Hap8 heterodimer structures. These results suggest that conformational differences are more likely to be affected by ERAP1 polymorphisms.

Active site motions were analyzed in detail. According to the results, the catalytic Tyr residue (Tyr438 in ERAP1, Tyr455 in ERAP2) usually demonstrated significant rotation. As for Phe residue, not only rotation was observed but also rearrangement of the same residue in some structures. According to one study, catalytic Tyr residue was coupled with Phe residue near catalytic site (Phe433 in ERAP1, Phe450 in ERAP2, located in helix H4a) which is rearranged by domain opening/closure movements (31). However no π-π interaction was observed. These results may imply changes in active site as it is known that Tyr438 side chain rotates towards the active site and stabilizes an intermediate during reaction (20).

Differences in active site motions were observed between different structures. Monomers of ERAP1 and ERAP2 were mostly stable in active site region while the catalytic Tyr in Hap2–Hap2 homodimers were in slight rotations after a certain time. On the other hand, rotation of the catalytic Tyr was not as common in Hap2–N392 heterodimers. Rearrangement of Phe near catalytic site was also prevalent in some structures, especially in Hap2–Hap2 homodimers. These findings suggest that allotypic differences and dimerization of ERAP1 proteins impact active site positions.

As a last measure, intermolecular interactions of ERAP1 Hap2–Hap2 homodimers and ERAP1/2 Hap2–N392 heterodimers were analyzed. Structures of Hap2–Hap2 demonstrated many common contacts within the interaction distance. Number of salt bridges and H-bonds were also higher and several salt bridges were formed by common residues in Hap2–Hap2 homodimer structures indicating strong interactions between chains. In contrast, Hap2–N392 heterodimer had significant loss in interactions between chains which had less defined residue pairs. Number of salt bridges were not as common and H-bonds were very low and sometimes absent due to complete separation of chains. According to these results, Hap2–Hap2 homodimer chain interactions are highly conserved and structures are in a much better stability than Hap2–N392 heterodimer.

## 5.2 Limitations of the Study

Structural studies of ERAP1 and ERAP2 structures were limited. Many studies experimenting with ERAP1 and ERAP2 focused on either functional effects of allotypic differences or structural features of single variant. Hence, comparison of the results to other experiments was not ideal. Nevertheless, this study was novel in identifying structural differences of allotypes in monomer and dimer forms at different time periods using computational techniques.

Because many ERAP1 common allotypes were present, many kinds of dimer combinations were possible. Due to the limited time, however, only selected allotypes were experimented in monomeric and dimeric forms.

While many studies involved in finding inhibitors to bind ERAP1 protein (30,31,34,35), study on hunting up the substrates of ERAP1/2 proteins was lacking. For the investigation of their interaction and their effect on enzymes, arthritogenic peptides were decided to be determined in a future experimental study of the current project to analyze interactions of protein–substrate complex.

ERAP1 and ERAP2 structures have not been crystallized in their heterodimer form. Although computational methods such as molecular docking were useful in predicting interactions of ERAP1 and ERAP2 structures in molecular level, they were, at least, not precise at demonstrating dimeric interactions as much as experimental methods used for identification.

As ERAP1/2 proteins have transmembrane region which is tethered to ER at the N-terminal end, their simulation was originally determined to be on a lipid bilayer instead of in a pure solvent box (22). However, the terminal end was only partially structured in crystal structures. Since many ERAP studies were done in solution, this limitation was disregarded.

# 6 CONCLUSION

ERAP1 and ERAP2 monomers were computationally analyzed to give insights to their contributions in AS pathogenesis. In the analyses, ERAP1 and ERAP2 closed monomers were stable in presence of allotypic differences. Nevertheless, allotypic differences impacted ERAP homo- and heterodimers significantly. These changes in simulated homodimers usually affected active site motions and conformational state of enzyme whereas changes in simulated heterodimers mostly affected interaction of chains. In active site of ERAP1 chains, rotation of catalytic Tyr residue and rearrangement of Phe residue located near catalytic site were monitored which were in lower frequencies in active site of ERAP2 chains. Structure of Hap2–Hap2 homodimers had stronger salt bridge and H-bonding in dimer interface causing stronger chain interactions than Hap2–N392 heterodimer structures. Additionally, dimer combinations of Hap8 were more destabilized than Hap2 dimer combinations which could alter peptide trimming efficiency and MHC-I immunopeptidome profiles causing development of AS. Future work on discovering the most immunogenic peptides and determining structural differences of arthritogenic peptide-bound ERAP1 and ERAP2 is necessary to bridge the gap of complex AS mechanism in the way to drug development.

# 7 REFERENCES


1. Braun J, Sieper J. Ankylosing spondylitis. The Lancet. 2007 Apr 21;369(9570):1379–90.
2. Zambrano-Zaragoza JF, Agraz-Cibrian JM, González-Reyes C, Durán-Avelar M de J, Vibanco-Pérez N. Ankylosing spondylitis: from cells to genes. Int J Inflamm. 2013;2013:501653.
3. Ranganathan V, Gracey E, Brown MA, Inman RD, Haroon N. Pathogenesis of ankylosing spondylitis - recent advances and future directions. Nat Rev Rheumatol. 2017 Jun;13(6):359–67.
4. Alvarez-Navarro C, López de Castro JA. ERAP1 structure, function and pathogenetic role in ankylosing spondylitis and other MHC-associated diseases. Mol Immunol. 2014 Jan;57(1):12–21.
5. Agrawal N, Brown MA. Genetic associations and functional characterization of M1 aminopeptidases and immune-mediated diseases. Genes Immun. 2014 Dec;15(8):521–7.
6. Smith JA. Update on ankylosing spondylitis: current concepts in pathogenesis. Curr Allergy Asthma Rep. 2015 Jan;15(1):489.
7. Reeves E, James E. The role of polymorphic ERAP1 in autoinflammatory disease. Biosci Rep. 2018 Aug 29;38(4):BSR20171503.
8. Haroon N, Inman RD. Endoplasmic reticulum aminopeptidases: biology and pathogenic potential. Nat Rev Rheumatol. 2010 Aug;6(8):461–7.
9. Evans DM, Spencer CCA, Pointon JJ, Su Z, Harvey D, Kochan G, et al. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nat Genet. 2011 Aug;43(8):761–7.
10. Vitulano C, Tedeschi V, Paladini F, Sorrentino R, Fiorillo MT. The interplay between HLA-B27 and ERAP1/ERAP2 aminopeptidases: from anti-viral protection to spondyloarthritis. Clin Exp Immunol. 2017 Dec;190(3):281–90.
11. Zhu W, He X, Cheng K, Zhang L, Chen D, Wang X, et al. Ankylosing spondylitis: etiology, pathogenesis, and treatments. Bone Res. 2019 Aug 5;7(1):1–16.
12. Blum JS, Wearsch PA, Cresswell P. Pathways of Antigen Processing. Annu Rev Immunol. 2013;31:443–73.
13. Chen B, Li J, He C, Li D, Tong W, Zou Y, et al. Role of HLA-B27 in the pathogenesis of ankylosing spondylitis (Review). Mol Med Rep. 2017 Apr;15(4):1943–51.
14. Drinkwater N, Lee J, Yang W, Malcolm TR, McGowan S. M1 aminopeptidases as drug targets: broad applications or therapeutic niche? FEBS J. 2017;284(10):1473–88.
15. Mpakali A, Maben Z, Stern LJ, Stratikos E. Molecular pathways for antigenic peptide generation by ER aminopeptidase 1. Mol Immunol. 2019;113:50–7.
16. Hattori A, Tsujimoto M. Endoplasmic reticulum aminopeptidases: biochemistry, physiology and pathology. J Biochem (Tokyo). 2013 Sep;154(3):219–28.

17. Admon A. ERAP1 shapes just part of the immunopeptidome. Hum Immunol. 2019 May;80(5):296–301.
18. Lorente E, Barriga A, Johnstone C, Mir C, Jiménez M, López D. Concerted In Vitro Trimming of Viral HLA-B27-Restricted Ligands by Human ERAP1 and ERAP2 Aminopeptidases. PLOS ONE. 2013 Nov 1;8(11):e79596.
19. Saveanu L, Carroll O, Lindo V, Del Val M, Lopez D, Lepelletier Y, et al. Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. Nat Immunol. 2005 Jul;6(7):689–97.
20. Nguyen T, Chang SC, Evnouchidou I, York I, Zikos C, Rock K, et al. Structural Basis For Antigenic Peptide Precursor Processing by the Endoplasmic Reticulum Aminopeptidase ERAP1. Nat Struct Mol Biol. 2011 May 1;18:604–13.
21. Kochan G, Krojer T, Harvey D, Fischer R, Chen L, Vollmar M, et al. Crystal structures of the endoplasmic reticulum aminopeptidase-1 (ERAP1) reveal the molecular basis for N-terminal peptide trimming. Proc Natl Acad Sci U S A. 2011 May 10;108(19):7745–50.
22. Peer WA. The role of multifunctional M1 metallopeptidases in cell cycle progression. Ann Bot. 2011 May 1;107(7):1171–81.
23. Hattori A, Goto Y, Tsujimoto M. Exon 10 coding sequence is important for endoplasmic reticulum retention of endoplasmic reticulum aminopeptidase 1. Biol Pharm Bull. 2012;35(4):601–5.
24. Chang SC, Momburg F, Bhutani N, Goldberg AL. The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a ‘molecular ruler’ mechanism. Proc Natl Acad Sci U S A. 2005 Nov 22;102(47):17107–12.
25. Wong AHM, Zhou D, Rini JM. The X-ray Crystal Structure of Human Aminopeptidase N Reveals a Novel Dimer and the Basis for Peptide Processing. J Biol Chem. 2012 Oct;287(44):36804–13.
26. Stratikos E, Stern LJ. Antigenic peptide trimming by ER aminopeptidases—Insights from structural studies. Mol Immunol. 2013 Oct 1;55(3):212–9.
27. Ito K, Nakajima Y, Onohara Y, Takeo M, Nakashima K, Matsubara F, et al. Crystal structure of aminopeptidase N (proteobacteria alanyl aminopeptidase) from Escherichia coli and conformational change of methionine 260 involved in substrate recognition. J Biol Chem. 2006 Nov 3;281(44):33664–76.
28. Thunnissen MMGM, Nordlund P, Haeggström JZ. Crystal structure of human leukotriene A4 hydrolase, a bifunctional enzyme in inflammation. Nat Struct Biol. 2001 Feb;8(2):131–5.
29. McGowan PO, Sasaki A, D’Alessio AC, Dymov S, Labonté B, Szyf M, et al. Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. Nat Neurosci. 2009 Mar;12(3):342–8.
30. Giastas P, Neu M, Rowland P, Stratikos E. High-Resolution Crystal Structure of Endoplasmic Reticulum Aminopeptidase 1 with Bound Phosphinic Transition-State Analogue Inhibitor. ACS Med Chem Lett. 2019 May 9;10(5):708–13.

31. Maben Z, Arya R, Georgiadis D, Stratikos E, Stern LJ. Conformational dynamics linked to domain closure and substrate binding explain the ERAP1 allosteric regulation mechanism. Nat Commun. 2021 Dec;12(1):5302.
32. Sui L, Guo HC. ERAP1 binds peptide C-termini of different sequences and/or lengths by a common recognition mechanism. Immunobiology. 2021 Jul 1;226(4):152112.
33. Sui L, Gandhi A, Guo HC. Crystal structure of a polypeptide's C-terminus in complex with the regulatory domain of ER aminopeptidase 1. Mol Immunol. 2016 Dec 1;80:41–9.
34. Liddle J, Hutchinson JP, Kitchen S, Rowland P, Neu M, Cecconie T, et al. Targeting the Regulatory Site of ER Aminopeptidase 1 Leads to the Discovery of a Natural Product Modulator of Antigen Presentation. J Med Chem. 2020 Mar 26;63(6):3348–58.
35. Gandhi A, Lakshminarasimhan D, Sun Y, Guo HC. Structural insights into the molecular ruler mechanism of the endoplasmic reticulum aminopeptidase ERAP1. Sci Rep. 2011 Dec 13;1(1):186.
36. Tholander F, Muroya A, Roques BP, Fournié-Zaluski MC, Thunnissen MMGM, Haeggström JZ. Structure-Based Dissection of the Active Site Chemistry of Leukotriene A4 Hydrolase: Implications for M1 Aminopeptidases and Inhibitor Design. Chem Biol. 2008 Sep;15(9):920–9.
37. Jones PM, Robinson MW, Dalton JP, George AM. The Plasmodium falciparum Malaria M1 Alanyl Aminopeptidase (PfA-M1): Insights of Catalytic Mechanism and Function from MD Simulations. PLOS ONE. 2011 Dec 21;6(12):e28589.
38. Reeves E, Islam Y, James E. ERAP1: a potential therapeutic target for a myriad of diseases. Expert Opin Ther Targets. 2020 Jun 2;24(6):535–44.
39. York IA, Chang SC, Saric T, Keys JA, Favreau JM, Goldberg AL, et al. The ER aminopeptidase ERAP1 enhances or limits antigen presentation by trimming epitopes to 8–9 residues. Nat Immunol. 2002 Dec;3(12):1177–84.
40. Mavridis G, Arya R, Domnick A, Zoidakis J, Makridakis M, Vlahou A, et al. A systematic re-examination of processing of MHCI-bound antigenic peptide precursors by endoplasmic reticulum aminopeptidase 1. J Biol Chem. 2020 May 22;295(21):7193–210.
41. Papakyriakou A, Reeves E, Beton M, Mikolajek H, Douglas L, Cooper G, et al. The partial dissociation of MHC class I bound peptides exposes their N terminus to trimming by endoplasmic reticulum aminopeptidase 1. J Biol Chem. 2018 Mar 29;jbc.RA117.000313.
42. Chen H, Li L, Weimershaus M, Evnouchidou I, van Endert P, Bouvier M. ERAP1-ERAP2 dimers trim MHC I-bound precursor peptides; implications for understanding peptide editing. Sci Rep. 2016 Aug 12;6(1):28902.
43. Kanaseki T, Blanchard N, Hammer GE, Gonzalez F, Shastri N. ERAAP Synergizes with MHC Class I Molecules to Make the Final Cut in the Antigenic Peptide Precursors in the Endoplasmic Reticulum. Immunity. 2006 Nov 1;25(5):795–806.

44. Mavridis G, Mpakali A, Zoidakis J, Makridakis M, Vlahou A, Kaloumenou E, et al. The ERAP1 active site cannot productively access the N-terminus of antigenic peptide precursors stably bound onto MHC class I. Sci Rep. 2021 Aug 13;11(1):16475.
45. Colbert JD, Rock KL. How a tailor achieves the perfect fit. J Biol Chem. 2020 May 22;295(21):7211–2.
46. Saric T, Chang SC, Hattori A, York IA, Markant S, Rock KL, et al. An IFN-γ–induced aminopeptidase in the ER, ERAP1, trims precursors to MHC class I–presented peptides. Nat Immunol. 2002 Dec;3(12):1169–76.
47. Tran TM, Colbert RA. Endoplasmic reticulum aminopeptidase 1 and rheumatic disease: functional variation. Curr Opin Rheumatol. 2015 Jul;27(4):357–63.
48. Evnouchidou I, Momburg F, Papakyriakou A, Chroni A, Leondiadis L, Chang SC, et al. The Internal Sequence of the Peptide-Substrate Determines Its N-Terminus Trimming by ERAP1. PLOS ONE. 2008 Nov 6;3(11):e3658.
49. Evnouchidou I, Weimershaus M, Saveanu L, Endert P van. ERAP1–ERAP2 Dimerization Increases Peptide-Trimming Efficiency. J Immunol. 2014 Jul 15;193(2):901–8.
50. Martín-Esteban A, Sanz-Bravo A, Guasp P, Barnea E, Admon A, López de Castro JA. Separate effects of the ankylosing spondylitis associated ERAP1 and ERAP2 aminopeptidases determine the influence of their combined phenotype on the HLA-B*27 peptidome. J Autoimmun. 2017 May;79:28–38.
51. Papakyriakou A, Mpakali A, Stratikos E. Can ERAP1 and ERAP2 Form Functional Heterodimers? A Structural Dynamics Investigation. Front Immunol [Internet]. 2022 [cited 2022 Jun 13];13. Available from: https://www.frontiersin.org/article/10.3389/fimmu.2022.863529
52. Mattorre B, Caristi S, Donato S, Volpe E, Faiella M, Paiardini A, et al. A Short ERAP2 That Binds IRAP Is Expressed in Macrophages Independently of Gene Variation. Int J Mol Sci. 2022 Jan;23(9):4961.
53. Ombrello MJ, Kastner DL, Remmers EF. Endoplasmic reticulum-associated amino-peptidase 1 and rheumatic disease: genetics. Curr Opin Rheumatol. 2015 Jul;27(4):349.
54. Stamogiannos A, Koumantou D, Papakyriakou A, Stratikos E. Effects of polymorphic variation on the mechanism of Endoplasmic Reticulum Aminopeptidase 1. Mol Immunol. 2015 Oct;67(2 Pt B):426–35.
55. Stratikos E, Stamogiannos A, Zervoudi E, Fruci D. A Role for Naturally Occurring Alleles of Endoplasmic Reticulum Aminopeptidases in Tumor Immunity and Cancer Pre-Disposition. Front Oncol. 2014;4:363.
56. Evnouchidou I, Birtley J, Seregin S, Papakyriakou A, Zervoudi E, Samiotaki M, et al. A Common Single Nucleotide Polymorphism in Endoplasmic Reticulum Aminopeptidase 2 Induces a Specificity Switch That Leads to Altered Antigen Processing. J Immunol. 2012 Sep 1;189(5):2383–92.

57. Reeves E, Colebatch-Bourn A, Elliott T, Edwards CJ, James E. Functionally distinct ERAP1 allotype combinations distinguish individuals with Ankylosing Spondylitis. Proc Natl Acad Sci U S A. 2014 Dec 9;111(49):17594–9.
58. Hutchinson JP, Temponeras I, Kuiper J, Cortes A, Korczynska J, Kitchen S, et al. Common allotypes of ER aminopeptidase 1 have substrate-dependent and highly variable enzymatic properties. J Biol Chem [Internet]. 2021 Jan 1 [cited 2022 May 2];296. Available from: https://www.jbc.org/article/S0021-9258(21)00216-7/abstract
59. Wellcome Trust Case Control Consortium, Australo-Anglo-American Spondylitis Consortium (TASC), Burton PR, Clayton DG, Cardon LR, Craddock N, et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet. 2007 Nov;39(11):1329–37.
60. Harvey D, Pointon JJ, Evans DM, Karaderi T, Farrar C, Appleton LH, et al. Investigating the genetic association between ERAP1 and ankylosing spondylitis. Hum Mol Genet. 2009 Nov 1;18(21):4204–12.
61. Maksymowych WP, Inman RD, Gladman DD, Reeve JP, Pope A, Rahman P. Association of a specific ERAP1/ARTS1 haplotype with disease susceptibility in ankylosing spondylitis. Arthritis Rheum. 2009 May;60(5):1317–23.
62. Kadi A, Izac B, Said-Nahal R, Leboime A, Praet LV, Vlam K de, et al. Investigating the genetic association between ERAP1 and spondyloarthritis. Ann Rheum Dis. 2013 Apr 1;72(4):608–13.
63. Yao Y, Liu N, Zhou Z, Shi L. Influence of ERAP1 and ERAP2 gene polymorphisms on disease susceptibility in different populations. Hum Immunol. 2019 May;80(5):325–34.
64. Andrés AM, Dennis MY, Kretzschmar WW, Cannons JL, Lee-Lin SQ, Hurle B, et al. Balancing Selection Maintains a Form of ERAP2 that Undergoes Nonsense-Mediated Decay and Affects Antigen Presentation. PLOS Genet. 2010 Oct 14;6(10):e1001157.
65. Wiśniewski A, Kasprzyk S, Majorczyk E, Nowak I, Wilczyńska K, Chlebicki A, et al. ERAP1-ERAP2 haplotypes are associated with ankylosing spondylitis in Polish patients. Hum Immunol. 2019 May 1;80(5):339–43.
66. López de Castro JA. How ERAP1 and ERAP2 Shape the Peptidomes of Disease-Associated MHC-I Proteins. Front Immunol [Internet]. 2018 [cited 2022 Apr 30];9. Available from: https://www.frontiersin.org/article/10.3389/fimmu.2018.02463
67. López de Castro JA, Alvarez-Navarro C, Brito A, Guasp P, Martín-Esteban A, Sanz-Bravo A. Molecular and pathogenic effects of endoplasmic reticulum aminopeptidases ERAP1 and ERAP2 in MHC-I-associated inflammatory disorders: Towards a unifying view. Mol Immunol. 2016 Sep 1;77:193–204.
68. Goto Y, Ogawa K, Hattori A, Tsujimoto M. Secretion of Endoplasmic Reticulum Aminopeptidase 1 Is Involved in the Activation of Macrophages Induced by Lipopolysaccharide and Interferon-γ. J Biol Chem. 2011 Jun 17;286(24):21906–14.

69. UniProt Consortium T. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2018 Mar 16;46(5):2699–2699.
70. Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Di Costanzo L, et al. RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. Nucleic Acids Res. 2019 Jan 8;47(D1):D464–74.
71. Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM. PDBsum: Structural summaries of PDB entries. Protein Sci. 2018;27(1):129–34.
72. Rigsby RE, Parker AB. Using the PyMOL application to reinforce visual understanding of protein structure. Biochem Mol Biol Educ. 2016;44(5):433–7.
73. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J Mol Graph. 1996 Feb 1;14(1):33–8.
74. Brooks BR, Brooks CL, MacKerell AD, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: The Biomolecular Simulation Program. J Comput Chem. 2009 Jul 30;30(10):1545–614.
75. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Bioinforma Ed Board Andreas Baxevanis Al. 2016 Jun 20;54:5.6.1-5.6.37.
76. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. Nat Methods. 2015 Jan;12(1):7–8.
77. de Vries SJ, van Dijk M, Bonvin AMJJ. The HADDOCK web server for data-driven biomolecular docking. Nat Protoc. 2010 May;5(5):883–97.
78. Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, et al. Improvements to the APBS biomolecular solvation software suite. Protein Sci Publ Protein Soc. 2018 Jan;27(1):112–28.
79. color_h.py [Internet]. [cited 2022 Jun 9]. Available from: http://www.protein.osaka-u.ac.jp/rcsfp/supracryst/suzuki/jpxtal/Katsutani/color_h.py
80. Eisenberg D, Schwarz E, Komaromy M, Wall R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J Mol Biol. 1984 Oct 15;179(1):125–42.
81. Drozdetskiy A, Cole C, Procter J, Barton GJ. JPred4: a protein secondary structure prediction server. Nucleic Acids Res. 2015 Jul 1;43(Web Server issue):W389–94.
82. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics. 2000 Apr 1;16(4):404–5.
83. Wang S, Li W, Liu S, Xu J. RaptorX-Property: a web server for protein structure property prediction. Nucleic Acids Res. 2016 Jul 8;44(Web Server issue):W430–5.
84. Hansen JE, Lund O, Tolstrup N, Gooley AA, Williams KL, Brunak S. NetOglyc: Prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. Glycoconj J. 1998 Feb 1;15(2):115–30.
85. Ganesan P, Varadharaju N, Kuo-Chen C, Govindaraju A. Nglyc: A Random Forest Method for Prediction of N-Glycosylation Sites in Eukaryotic Protein Sequence. Protein Pept Lett. 2020 Feb 29;27(3):178–86.

86. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. Nucleic Acids Res. 2005 Jul 1;33(Web Server issue):W382-388.
87. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods. 2017 Jan;14(1):71–3.
88. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable Molecular Dynamics with NAMD. J Comput Chem. 2005 Dec;26(16):1781–802.
89. Koukos PI, Glykos NM. Grcarma: A fully automated task-oriented interface for the analysis of molecular dynamics trajectories. J Comput Chem. 2013 Oct 5;34(26):2310–2.
90. Mohamed shehata. RMSF analysis of trajectory (DCD file) using VMD [Internet]. 2019 [cited 2022 Jun 9]. Available from: https://www.youtube.com/watch?v=CkCREhk1SL4
91. Stamogiannos A, Maben Z, Papakyriakou A, Mpakali A, Kokkala P, Georgiadis D, et al. Critical Role of Interdomain Interactions in the Conformational Change and Catalytic Mechanism of Endoplasmic Reticulum Aminopeptidase 1. Biochemistry. 2017 Mar 14;56(10):1546–58.
92. sasa.tcl [Internet]. [cited 2022 Jun 9]. Available from: https://www.ks.uiuc.edu/Research/vmd/mailing_list/vmd-l/att-18670/sasa.tcl
93. Chiang H, Robinson LC, Brame CJ, Messina TC. Molecular mechanics and dynamics characterization of an in silico mutated protein: A stand-alone lab module or support activity for in vivo and in vitro analyses of targeted proteins. Biochem Mol Biol Educ. 2013;41(6):402–8.
94. Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. Bio3d: an R package for the comparative analysis of protein structures. Bioinforma Oxf Engl. 2006 Nov 1;22(21):2695–6.
95. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011 Jan;7(1):539.
96. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009 May 1;25(9):1189–91.
97. Heinig M, Frishman D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. Nucleic Acids Res. 2004 Jul 1;32(Web Server issue):W500–2.

# 8 APPENDIX

## APPENDIX 1

**Residues defined in “active residues” parameter of HADDOCK.**

**For ERAP1 structures at domain IV region:**

680, 681, 683, 684, 685, 687, 688, 690, 693, 694, 695, 696, 697, 698, 699, 701, 702, 703, 704, 705, 706, 734, 737, 738, 776, 777, 778, 779, 782, 805, 806, 807, 808, 809, 811, 812, 813, 814, 815, 816, 817, 836, 839, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 855, 856, 857, 858, 859, 860, 878, 879, 882, 883, 884, 885, 887, 888, 889, 891, 892, 893, 894, 895, 896, 912, 919, 933

**For ERAP2 structures at domain IV region:**

703, 707, 711, 713, 716, 717, 718, 719, 720, 721, 722, 724, 725, 726, 727, 728, 729, 734, 757, 758, 760, 761, 762, 763, 799, 800, 801, 802, 803, 805, 828, 829, 830, 831, 832, 834, 835, 836, 837, 838, 839, 840, 865, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 878, 879, 880, 881, 882, 883, 897, 901, 902, 905, 906, 907, 908, 910, 911, 912, 914, 915, 916, 917, 918, 919, 920, 921, 935, 938, 942, 956

**For ERAP1 structures at exon 10 loop region:**

371, 373, 374, 418, 420, 440, 444, 445, 448, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 538, 539, 540, 541, 542, 546, 561, 562, 563, 564, 565, 566, 570, 573, 574, 575, 576, 577, 578, 581, 582, 583, 589, 590, 591, 592, 593, 594

**For ERAP2 structures at exon 10 loop region:**

64, 65, 66, 67, 68, 72, 100, 101, 102, 103, 104, 105, 145, 147, 157, 158, 159, 160, 217, 258, 311, 314, 315, 316, 317, 318, 376, 380, 381, 382, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 498, 501, 502, 503, 506, 507, 508, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 966, 969

## APPENDIX 2

### Hydrophobicity scale measurement by “color_h.py” script.

```
# color_h
# -------

# PyMOL command to color protein molecules according to the Eisenberg
hydrophobicity scale

#
# Source: http://us.expasy.org/tools/pscale/Hphob.Eisenberg.html
# Amino acid scale: Normalized consensus hydrophobicity scale
# Author(s): Eisenberg D., Schwarz E., Komarony M., Wall R.
# Reference: J. Mol. Biol. 179:125-142 (1984)
#
# Amino acid scale values:
#
# Ala:  0.620
# Arg: -2.530
# Asn: -0.780
# Asp: -0.900
# Cys:  0.290
# Gln: -0.850
# Glu: -0.740
# Gly:  0.480
# His: -0.400
# Ile:  1.380
# Leu:  1.060
# Lys: -1.500
# Met:  0.640
# Phe:  1.190
# Pro:  0.120
# Ser: -0.180
# Thr: -0.050
# Trp:  0.810
# Tyr:  0.260
# Val:  1.080
#
# Usage:
# color_h (selection)
#
# Update:
# Color gradient legend added

from pymol import cmd

def color_h(selection='all'):
        s = str(selection)
        print(s)
        cmd.set_color('color_ile',[0.996,0.062,0.062])
        cmd.set_color('color_phe',[0.996,0.109,0.109])
        cmd.set_color('color_val',[0.992,0.156,0.156])
```

```
        cmd.set_color('color_leu',[0.992,0.207,0.207])
        cmd.set_color('color_trp',[0.992,0.254,0.254])
        cmd.set_color('color_met',[0.988,0.301,0.301])
        cmd.set_color('color_ala',[0.988,0.348,0.348])
        cmd.set_color('color_gly',[0.984,0.394,0.394])
        cmd.set_color('color_cys',[0.984,0.445,0.445])
        cmd.set_color('color_tyr',[0.984,0.492,0.492])
        cmd.set_color('color_pro',[0.980,0.539,0.539])
        cmd.set_color('color_thr',[0.980,0.586,0.586])
        cmd.set_color('color_ser',[0.980,0.637,0.637])
        cmd.set_color('color_his',[0.977,0.684,0.684])
        cmd.set_color('color_glu',[0.977,0.730,0.730])
        cmd.set_color('color_asn',[0.973,0.777,0.777])
        cmd.set_color('color_gln',[0.973,0.824,0.824])
        cmd.set_color('color_asp',[0.973,0.875,0.875])
        cmd.set_color('color_lys',[0.899,0.922,0.922])
        cmd.set_color('color_arg',[0.899,0.969,0.969])
        cmd.color("color_ile","("+s+" and resn ile)")
        cmd.color("color_phe","("+s+" and resn phe)")
        cmd.color("color_val","("+s+" and resn val)")
        cmd.color("color_leu","("+s+" and resn leu)")
        cmd.color("color_trp","("+s+" and resn trp)")
        cmd.color("color_met","("+s+" and resn met)")
        cmd.color("color_ala","("+s+" and resn ala)")
        cmd.color("color_gly","("+s+" and resn gly)")
        cmd.color("color_cys","("+s+" and resn cys)")
        cmd.color("color_tyr","("+s+" and resn tyr)")
        cmd.color("color_pro","("+s+" and resn pro)")
        cmd.color("color_thr","("+s+" and resn thr)")
        cmd.color("color_ser","("+s+" and resn ser)")
        cmd.color("color_his","("+s+" and resn his)")
        cmd.color("color_glu","("+s+" and resn glu)")
        cmd.color("color_asn","("+s+" and resn asn)")
        cmd.color("color_gln","("+s+" and resn gln)")
        cmd.color("color_asp","("+s+" and resn asp)")
        cmd.color("color_lys","("+s+" and resn lys)")
        cmd.color("color_arg","("+s+" and resn arg)")
cmd.extend('color_h',color_h)

def color_h2(selection='all'):
        s = str(selection)
        print(s)
        cmd.set_color("color_ile2",[0.938,1,0.938])
        cmd.set_color("color_phe2",[0.891,1,0.891])
        cmd.set_color("color_val2",[0.844,1,0.844])
        cmd.set_color("color_leu2",[0.793,1,0.793])
        cmd.set_color("color_trp2",[0.746,1,0.746])
        cmd.set_color("color_met2",[0.699,1,0.699])
        cmd.set_color("color_ala2",[0.652,1,0.652])
        cmd.set_color("color_gly2",[0.606,1,0.606])
        cmd.set_color("color_cys2",[0.555,1,0.555])
        cmd.set_color("color_tyr2",[0.508,1,0.508])
        cmd.set_color("color_pro2",[0.461,1,0.461])
        cmd.set_color("color_thr2",[0.414,1,0.414])
        cmd.set_color("color_ser2",[0.363,1,0.363])
        cmd.set_color("color_his2",[0.316,1,0.316])
        cmd.set_color("color_glu2",[0.27,1,0.27])
```

```
        cmd.set_color("color_asn2",[0.223,1,0.223])
        cmd.set_color("color_gln2",[0.176,1,0.176])
        cmd.set_color("color_asp2",[0.125,1,0.125])
        cmd.set_color("color_lys2",[0.078,1,0.078])
        cmd.set_color("color_arg2",[0.031,1,0.031])
        cmd.color("color_ile2","("+s+" and resn ile)")
        cmd.color("color_phe2","("+s+" and resn phe)")
        cmd.color("color_val2","("+s+" and resn val)")
        cmd.color("color_leu2","("+s+" and resn leu)")
        cmd.color("color_leu2","("+s+" and resn leu)")
        cmd.color("color_trp2","("+s+" and resn trp)")
        cmd.color("color_met2","("+s+" and resn met)")
        cmd.color("color_ala2","("+s+" and resn ala)")
        cmd.color("color_gly2","("+s+" and resn gly)")
        cmd.color("color_cys2","("+s+" and resn cys)")
        cmd.color("color_tyr2","("+s+" and resn tyr)")
        cmd.color("color_pro2","("+s+" and resn pro)")
        cmd.color("color_thr2","("+s+" and resn thr)")
        cmd.color("color_ser2","("+s+" and resn ser)")
        cmd.color("color_his2","("+s+" and resn his)")
        cmd.color("color_glu2","("+s+" and resn glu)")
        cmd.color("color_asn2","("+s+" and resn asn)")
        cmd.color("color_gln2","("+s+" and resn gln)")
        cmd.color("color_asp2","("+s+" and resn asp)")
        cmd.color("color_lys2","("+s+" and resn lys)")
        cmd.color("color_arg2","("+s+" and resn arg)")
cmd.extend('color_h2',color_h2)

ramp_new h_color, _enter_the_name_of_your_structure_here_, [1.380, -
2.530], [red, white]
```

## APPENDIX 3

### RMSF calculation by “rmsf.tcl” script.

```
# USAGE:
#
#            source rmsf.tcl
#
#
set reference [atomselect top "protein" frame 1]
# the frame being compared
set compare [atomselect top "protein"]
set num_steps [molinfo top get numframes]

for {set frame 0} {$frame < $num_steps} {incr frame} {
  # get the correct frame
  $compare frame $frame
  # compute the transformation
  set trans_mat [measure fit $compare $reference]
  # do the alignment
  $compare move $trans_mat
}

set outfile [open RMSF.txt w]
set sel [atomselect top "name CA"]
#puts $outfile "[measure rmsf $sel first 0 last -1 step 1]"
set rmsf [measure rmsf $sel first 0 last -1 step 1]

for {set i 0} {$i < [$sel num]} {incr i} {
  puts $outfile "[expr {$i+1}] [lindex $rmsf $i]"
}

close $outfile
```

## APPENDIX 4

### Theta angle (θ) calculation by "hinge_angle.tcl" script.

```
# USAGE:
#
#      source hinge_angle.tcl
#      hinge_angle (sel1) (sel2) (sel_hinge) _outputfile_.txt
#
#

proc hinge_angle {seltext1 seltext2 hinge ha_out } {

set sel1 [atomselect top "$seltext1"]
set sel2 [atomselect top "$seltext2"]
set selh [atomselect top "$hinge"]
set pi [expr {atan(1) * 4}]

set nf [molinfo top get numframes]

set outfile [open $ha_out w]

for {set i 0} {$i < $nf} {incr i} {
       puts "frame $i of $nf"
       $sel1 frame $i
       $sel2 frame $i
       $selh frame $i
       set com1 [measure center $sel1 weight mass]
       set com2 [measure center $sel2 weight mass]
       set comh [measure center $selh weight mass]

       set dh_1 [vecsub $com1 $comh]
       set dh_2 [vecsub $com2 $comh]

       set angle_top [vecdot $dh_1 $dh_2]

       set dh_1_len [veclength $dh_1]
       set dh_2_len [veclength $dh_2]

       set angle_bottom [vecdot $dh_1_len $dh_2_len]
       set angle_cos [expr $angle_top/$angle_bottom]
       set angle_rad [expr acos($angle_cos)]
       set angle($i.r) [expr $angle_rad*(180/$pi)]

       puts $outfile "$i , $angle($i.r)"

}

close $outfile

}
```

## APPENDIX 5

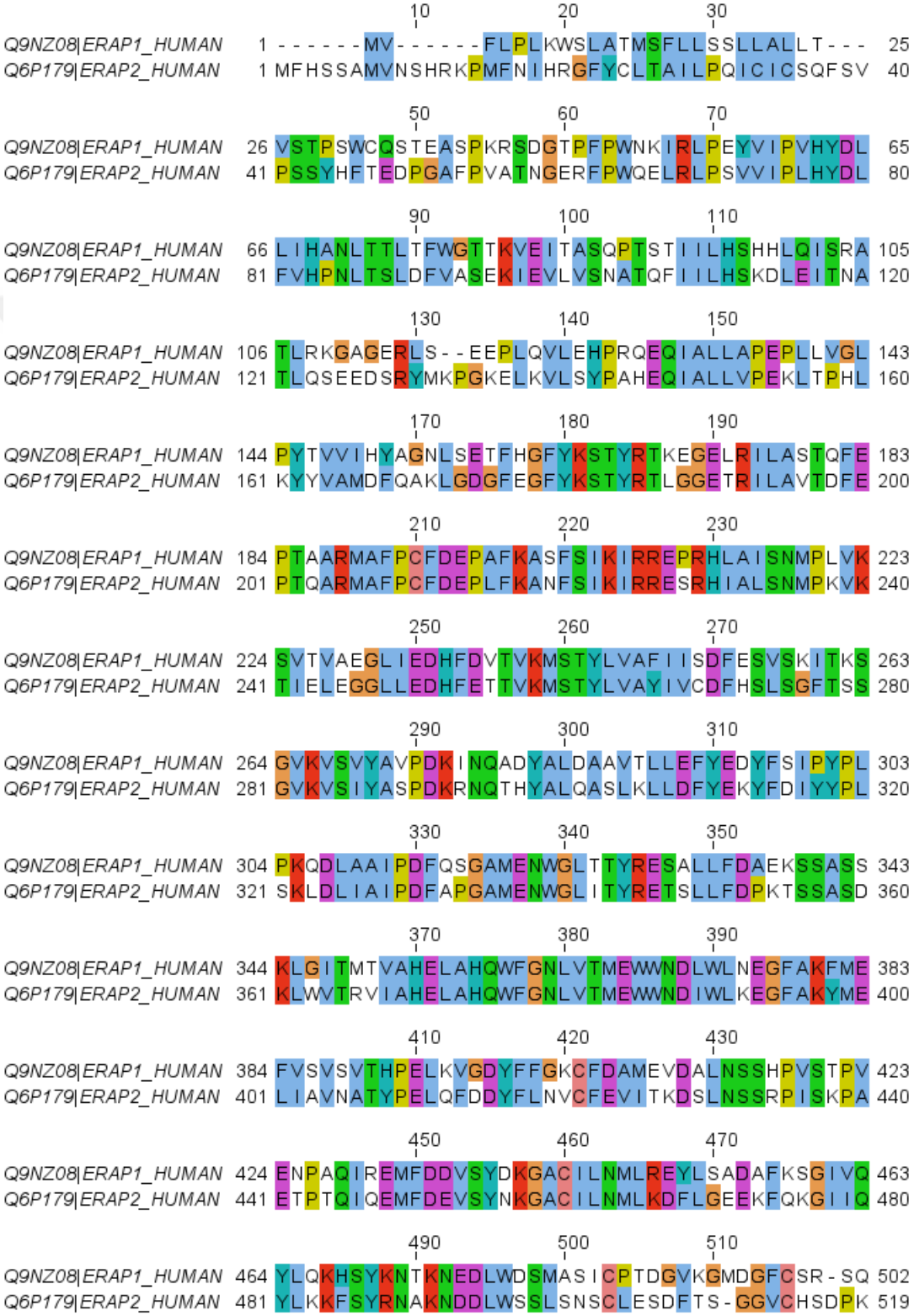
### SASA calculation by “sasa.tcl” script.

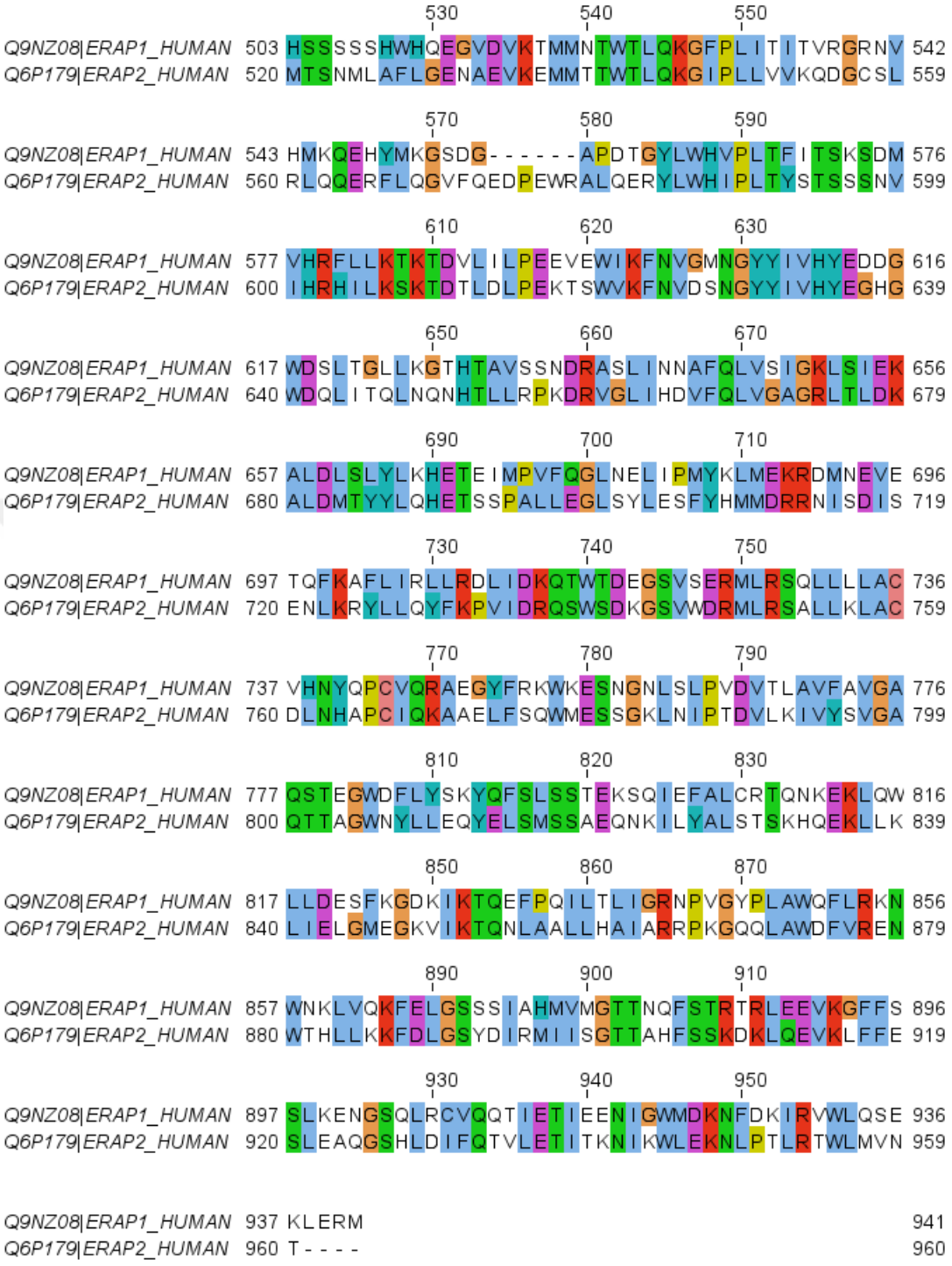
```
#####################################################################
# sasa.tcl                                                          #
# DESCRIPTION:                                                      #
#    This script is quick and easy to provide procedure             #
# for computing the Solvent Accessible Surface Area (SASA)          #
# of Protein and allows Users to select regions of protein.         #
#                                                                   #
# EXAMPLE USAGE:                                                    #
#         source sasa.tcl                                           #
#         Selection: chain A and resid 1                            #
#                                                                   #
#   AUTHORS:                                                        #
#      Sajad Falsafi (sajad.falsafi@yahoo.com)                    #
#       Zahra Karimi                                                #
#       3 Sep 2011                                                  #
#####################################################################

puts -nonewline "\n \t \t Selection: "
gets stdin selmode
# selection
set sel [atomselect top "$selmode"]
set protein [atomselect top "protein"]
set n [molinfo top get numframes]
set output [open "SASA_$selmode.dat" w]
# sasa calculation loop
for {set i 0} {$i < $n} {incr i} {
      molinfo top set frame $i
      set sasa [measure sasa 1.4 $protein -restrict $sel]
      puts "\t \t progress: $i/$n"
      puts $output "$sasa"
}
puts "\t \t progress: $n/$n"
puts "Done."
puts "output file: SASA_$selmode.dat"
close $output
```

## APPENDIX 6

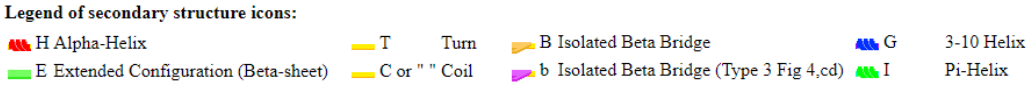
### Multiple sequence alignment of ERAP1 & ERAP2 sequences in Clustal Ω.



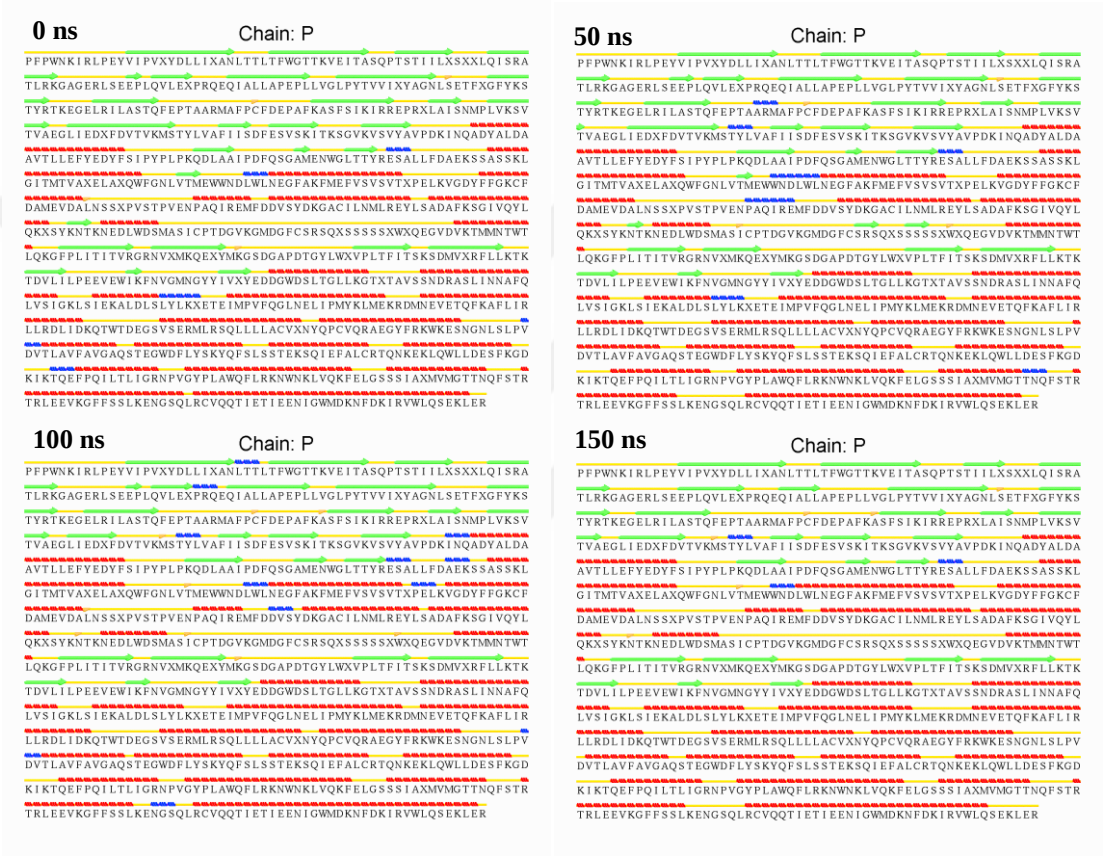

530 540 550
Q9NZ08|ERAP1_HUMAN 503 HSSSSSHWHQEGVDVKTMMNTWTLQKGFPLITITVRGRNV 542
Q6P179|ERAP2_HUMAN 520 MTSNMLAFLGENAEVKEMMTTWTLQKGIPLLVVKQDGCSL 559
570 580 590
Q9NZ08|ERAP1_HUMAN 543 HMKQEHYMKGSDG------APDTGYLWHVPLTFITSKSDM 576
Q6P179|ERAP2_HUMAN 560 RLQQERFLQGVFQEDPEWRALQERYLWHIPLTYSTSSSNV 599
610 620 630
Q9NZ08|ERAP1_HUMAN 577 VHRFLLKTKTDVLILPEEVEWIKFNVGMNGYYIVHYEDDG 616
Q6P179|ERAP2_HUMAN 600 IHRHILKSKTDTLDLPEKTSWVKFNVDSNGYYIVHYEGHG 639
650 660 670
Q9NZ08|ERAP1_HUMAN 617 WDSLTGLLKGTHTAVSSNDRASLINNAFQLVSIGKLSIEK 656
Q6P179|ERAP2_HUMAN 640 WDQLITQLNQNHTLLRPKDRVGLIHDVFQLVGAGRLTLDK 679
690 700 710
Q9NZ08|ERAP1_HUMAN 657 ALDLSLYLKHETEIMPVFQGLNELIPMYKLMEKRDMNEVE 696
Q6P179|ERAP2_HUMAN 680 ALDMTYYLQHETSSPALLEGLSYLESFYHMMDRRNISDIS 719
730 740 750
Q9NZ08|ERAP1_HUMAN 697 TQFKAFLIRLLRDLIDKQTWTDEGSVSERMLRSQLLLLAC 736
Q6P179|ERAP2_HUMAN 720 ENLKRYLLQYFKPVIDRQSWSDKGSVWDRMLRSALLKLAC 759
770 780 790
Q9NZ08|ERAP1_HUMAN 737 VHNYQPCVQRAEGYFRKWKESNGNLSLPVDVTLAVFAVGA 776
Q6P179|ERAP2_HUMAN 760 DLNHAPCIQKAAELFSQWMESSGKLNIPTDVLKIVYSVGA 799
810 820 830
Q9NZ08|ERAP1_HUMAN 777 QSTEGWDFLYSKYQFSLSSTEKSQIEFALCRTQNKEKLQW 816
Q6P179|ERAP2_HUMAN 800 QTTAGWNYLLEQYELSMSSAEQNKILYALSTSKHQEKLLK 839
850 860 870
Q9NZ08|ERAP1_HUMAN 817 LLDESFKGDKIKTQEFPQILTLIGRNPVGYPLAWQFLRKN 856
Q6P179|ERAP2_HUMAN 840 LIELGMEGKVIKTQNLAALLHAIARRPKGQQLAWDFVREN 879
890 900 910
Q9NZ08|ERAP1_HUMAN 857 WNKLVQKFELGSSSIAHMVMGTTNQFSTRTRLEEVKGFFS 896
Q6P179|ERAP2_HUMAN 880 WTHLLKKFDLGSYDIRMIISGTTAHFSSKDKLQEVKLFFE 919
930 940 950
Q9NZ08|ERAP1_HUMAN 897 SLKENGSQLRCVQQTIETIEENIGWMDKNFDKIRVWLQSE 936
Q6P179|ERAP2_HUMAN 920 SLEAQGSHLDIFQTVLETITKNIKWLEKNLPTLRTWLMVN 959
Q9NZ08|ERAP1_HUMAN 937 KLERM 941
Q6P179|ERAP2_HUMAN 960 T---- 960

## APPENDIX 7

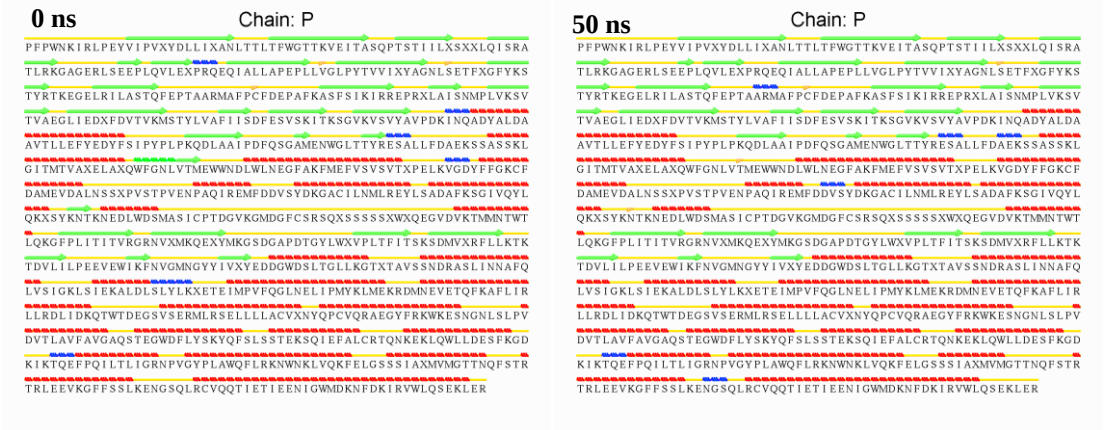
## STRIDE secondary structure assignment at 50 ns intervals of simulations.

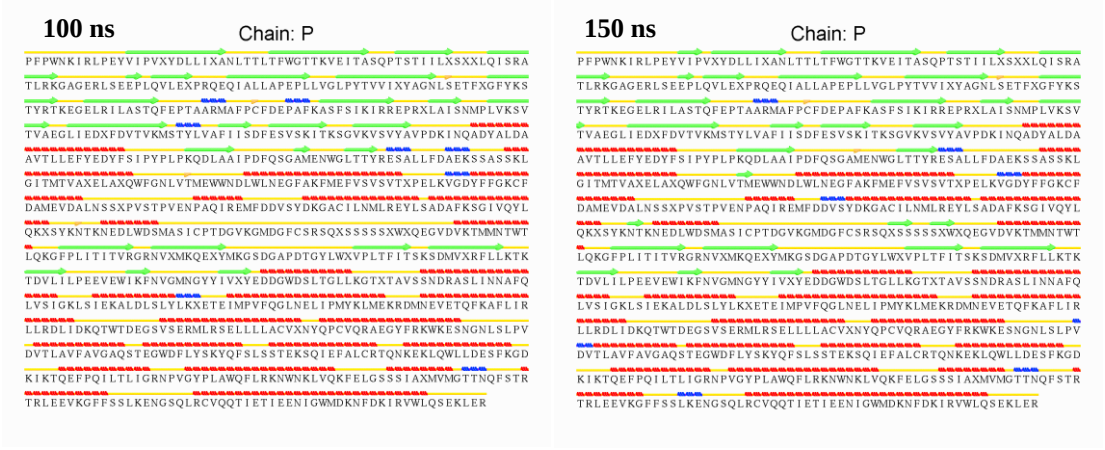



## Appendix 7a. STRIDE results of Hap2 monomer (first run).

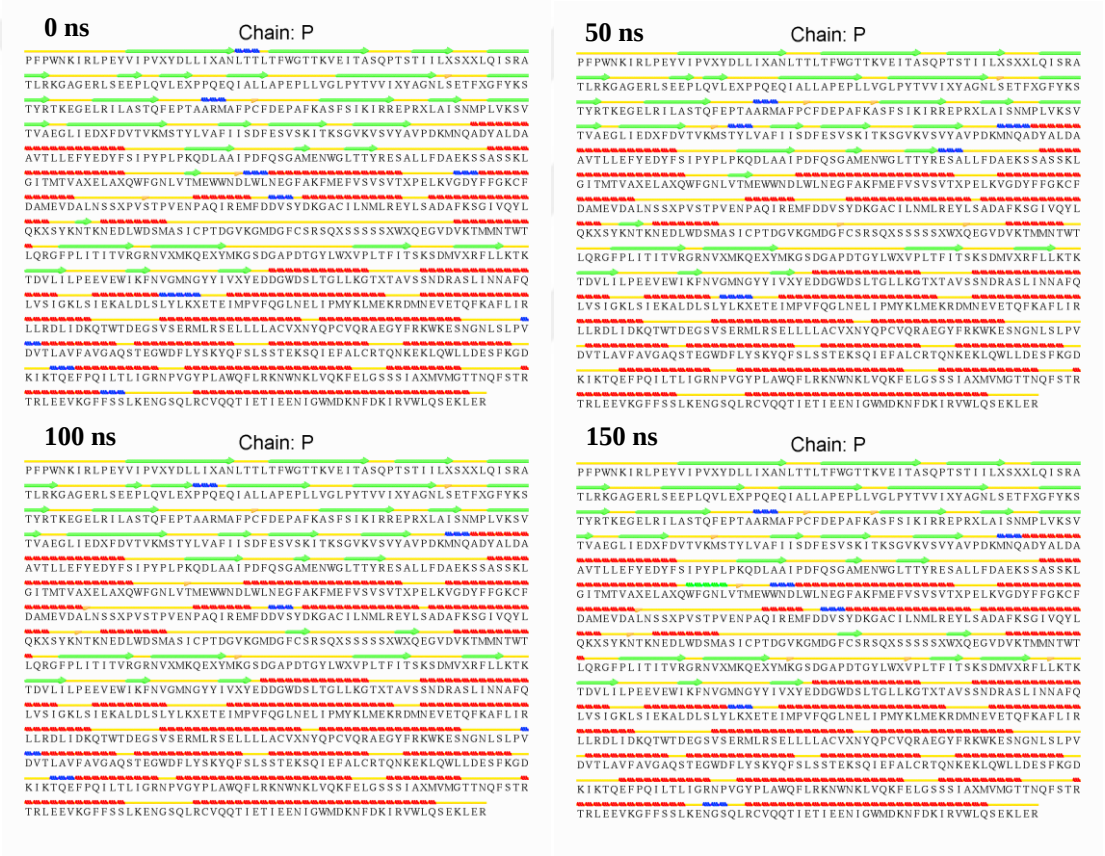



## Appendix 7b. STRIDE results of Hap3 monomer (first run).

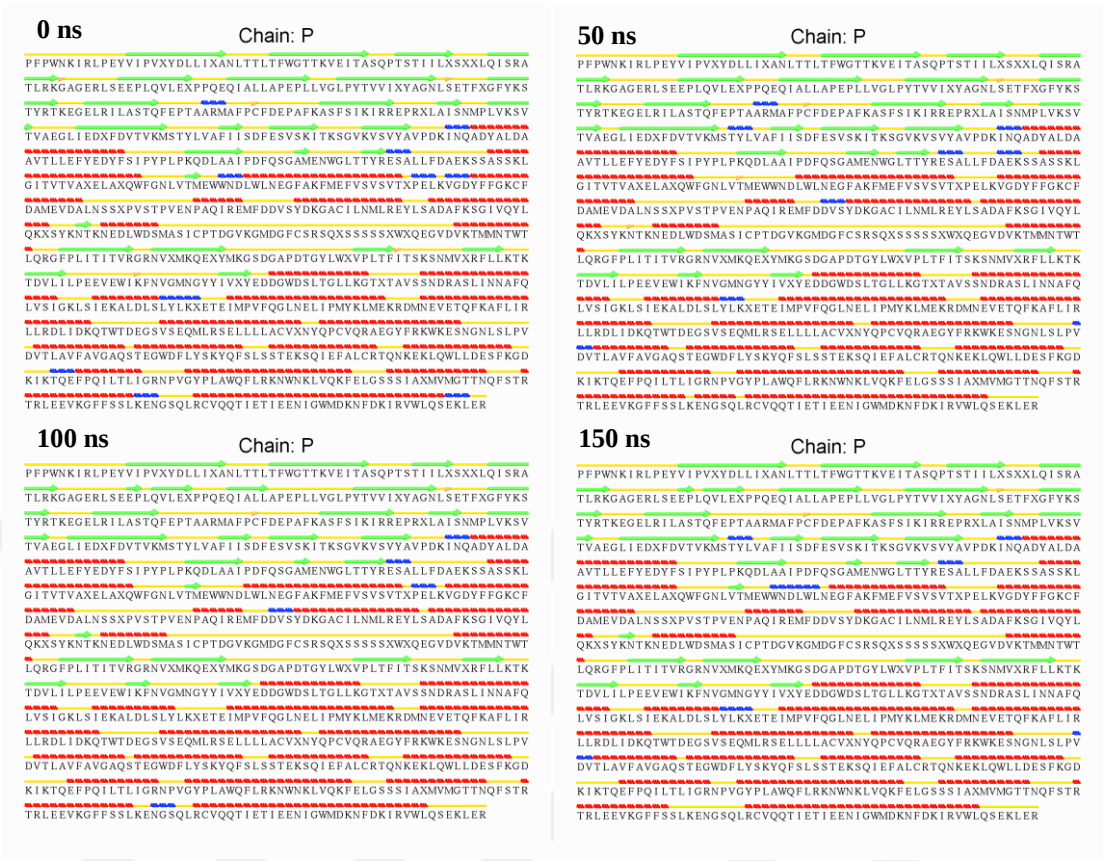




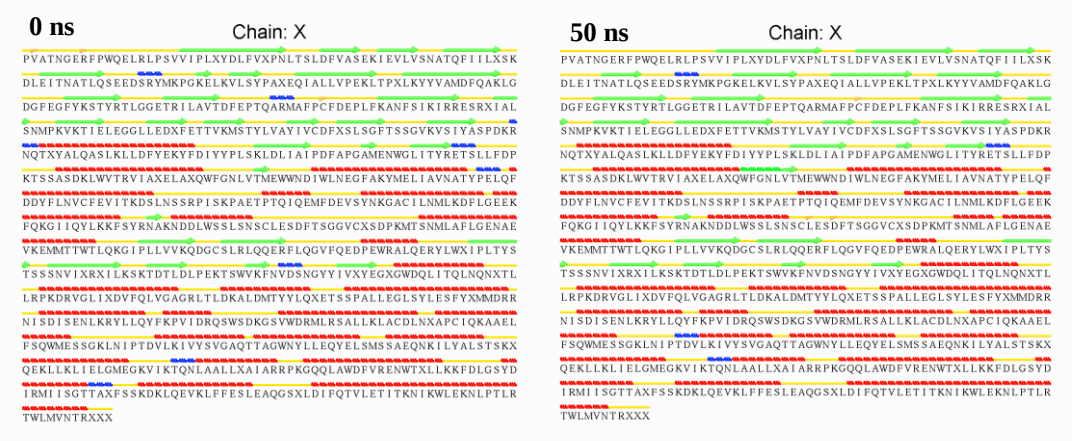

## Appendix 7c. STRIDE results of Hap8 monomer (first run).

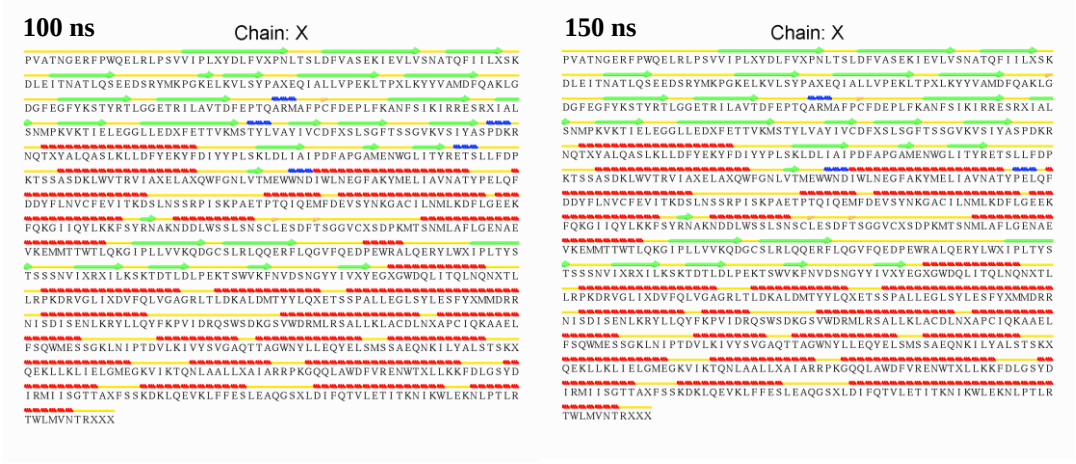


## Appendix 7d. STRIDE results of Hap10 monomer (first run).

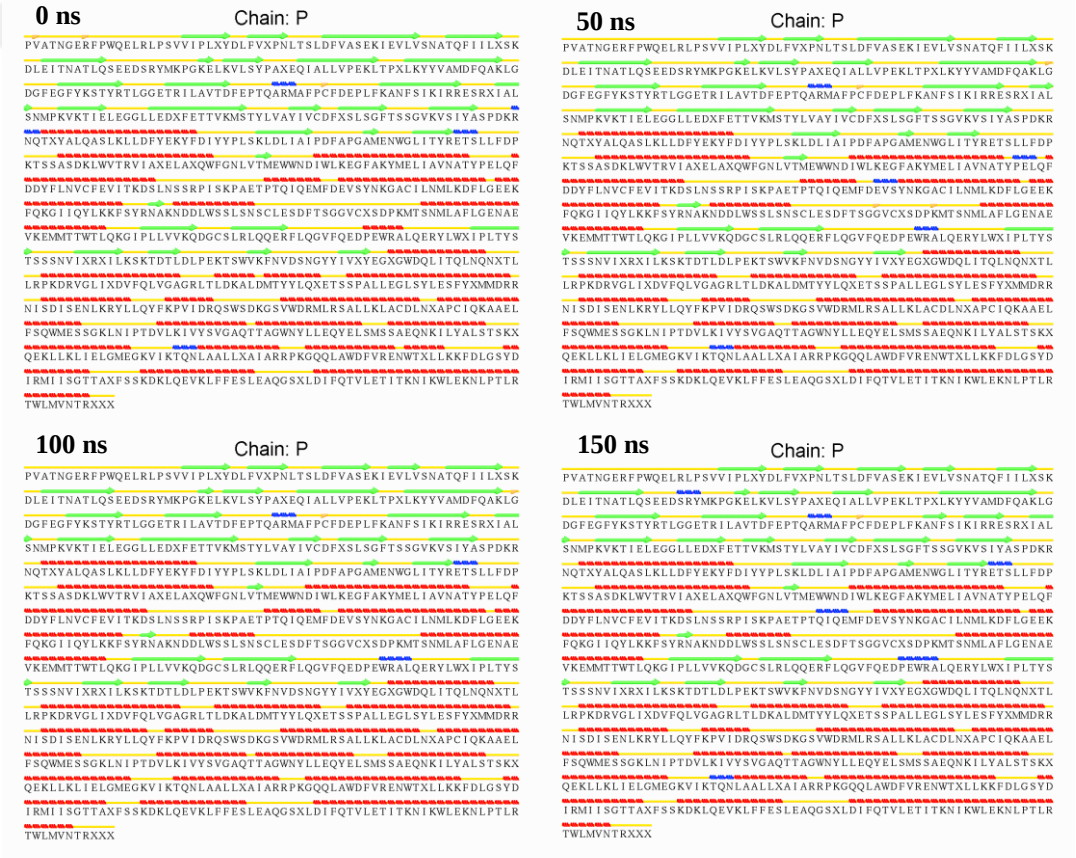



## Appendix 7e. STRIDE results of N392 monomer (first run).

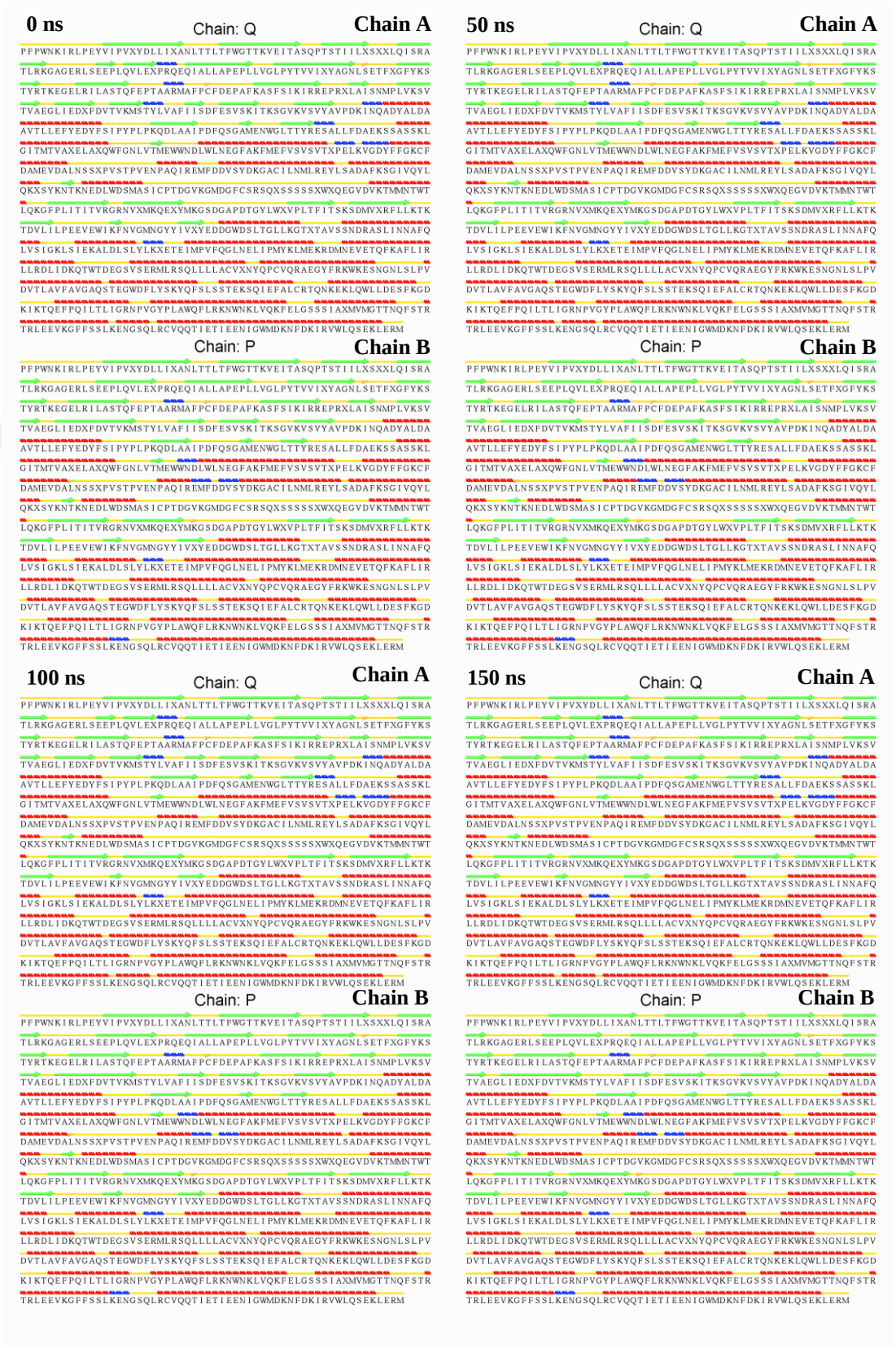




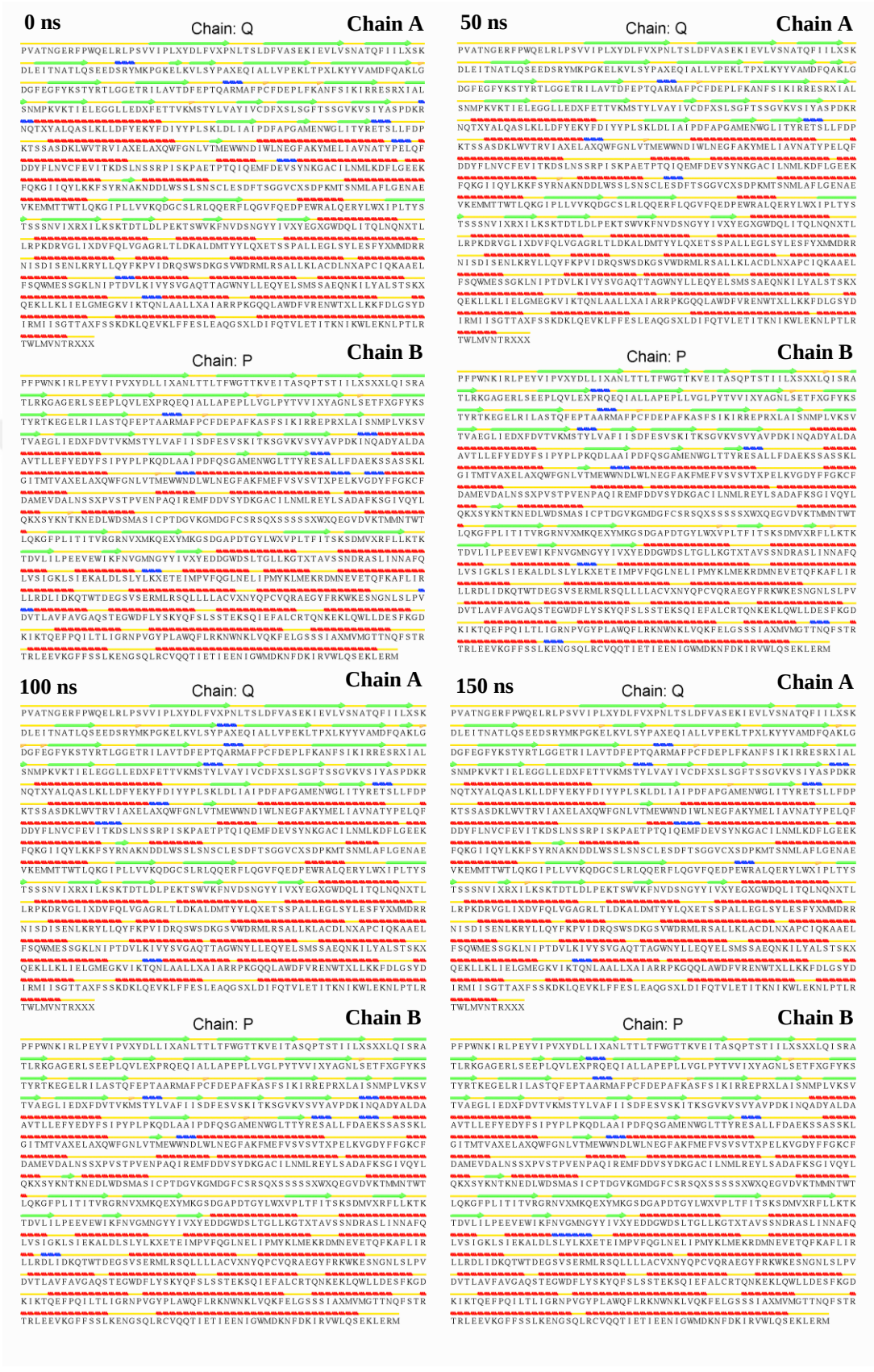

## Appendix 7f. STRIDE results of K392 monomer (first run).



## Appendix 7g. STRIDE results of Hap2–Hap2 homodimer (first run).



## Appendix 7h. STRIDE results of Hap2–N392 heterodimer (first run).



# 9 CURRICULUM VITAE