

**BEZMİALEM VAKIF ÜNİVERSİTESİ  
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

**İNSAN SOLUNUM SİNSİTYAL VİRÜSÜNÜN PROTEİN DİZİSİ ÇEŞİTLİLİĞİ  
DİNAMİKLERİNİN İNCELENMESİ**

**YÜKSEK LİSANS TEZİ**

**Faruk ÜSTÜNEL**

**Biyoteknoloji Anabilim Dalı**

**Biyoteknoloji Tezli Yüksek Lisans**

**Tez Danışmanı: Prof. Dr. Atilla AKDEMİR  
İkinci Danışman: Doç.Dr. Mohammad Asif KHAN**

**Biyoteknoloji Anabilim Dalı**

**Biyoteknoloji Tezli Yüksek Lisans**

**TEMMUZ 2022**

**BEZMİALEM VAKIF ÜNİVERSİTESİ  
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

**İNSAN SOLUNUM SİNSİTYAL VİRÜSÜNÜN PROTEİN DİZİSİ ÇEŞİTLİLİĞİ  
DİNAMİKLERİNİN İNCELENMESİ**

**YÜKSEK LİSANS TEZİ**

**Faruk ÜSTÜNEL  
(195309005)**

**Biyoteknoloji Anabilim Dalı**

**Biyoteknoloji Tezli Yüksek Lisans**

**Tez Danışmanı: Prof. Dr. Atilla AKDEMİR  
İkinci Danışman: Doç.Dr. Mohammad Asif KHAN**

**Biyoteknoloji Anabilim Dalı**

**Biyoteknoloji Tezli Yüksek Lisans**

**TEMMUZ 2022**

Bezmialem Vakıf Üniversitesi, Sağlık Bilimleri Enstitüsü'nün 195309005 numaralı Yüksek Lisans Öğrencisi Faruk ÜSTÜNEL, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı “İNSAN SOLUNUM SİNSİTYAL VİRÜSÜNÜN PROTEİN DİZİSİ ÇEŞİTLİLİĞİ DİNAMİKLERİNİN İNCELENMESİ” başlıklı tezini aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

**Tez Danışmanı :** **Prof. Dr. Atilla AKDEMİR**  
Bezmialem Vakıf Üniversitesi

**Eş Danışman :** **Doç.Dr. Mohammad Asif KHAN**  
Bezmialem Vakıf Üniversitesi

**Jüri Üyeleri :** **Prof. Dr. Fahri AKBAŞ**  
Bezmialem Vakıf Üniversitesi

**Doç. Dr. Emrah YÜCESAN**  
Bezmialem Vakıf Üniversitesi

**Dr. Öğr. Üyesi Onur SERÇİNOĞLU**  
Gebze Teknik Üniversitesi

## ÖNSÖZ

Yüksek Lisans tez araştırma sürecim boyunca her konuda bana destek olan Sayın Prof. Dr. Binnur TEMEL ve tez danışmanlarım Sayın Prof. Dr. Atilla AKDEMİR ve Sayın Doç.Dr. Mohammad Asif KHAN'a teşekkürlerimi borç bilirim.

Temmuz 2022

Faruk ÜSTÜNEL  
Biyoteknolog

**Teslim Tarihi** : 22.07.2022  
**Savunma Tarihi** : 04.07.2022

## **BEYAN**

Bu tez çalışmasının kendi çalışmam olduğunu, tezin planlanmasından yazımına kadar bütün safhalarda etik dışı davranışımın olmadığını, bu tezdeki bütün bilgileri akademik ve etik kurallar içinde elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara kaynak gösterdiğimi ve bu kaynakları da kaynaklar listesine aldığımı, yine bu tezin çalışılması ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığını beyan ederim.

Faruk ÜSTÜNEL

## İÇİNDEKİLER

### Sayfa

ÖNSÖZ .....	iii
BEYAN .....	iv
İÇİNDEKİLER .....	v
KISALTMALAR .....	vi
SEMBOLLER .....	vii
TABLO LİSTESİ .....	ix
ŞEKİL LİSTESİ .....	x
ÖZET .....	xi
SUMMARY .....	xii
<b>1. GİRİŞ .....</b>	<b>1</b>
1.1 Tezin Kapsamı ve Amacı .....	1
1.2 Hipotez .....	1
<b>2. LİTERATÜR ARAŞTIRMASI .....</b>	<b>2</b>
2.1. Hastalığın Epidemiyolojisi .....	2
2.1.1 Mevsimsellik .....	2
2.2 İnsan Solunum Sinsityal Virüsü .....	2
2.2.1 Tarihçe .....	2
2.2.2 Genom yapısı ve nomenklatur .....	3
2.2.3 hRSV'nin sınıflandırılması .....	5
2.2.4. hRSV patogenezi .....	5
2.2.5 Viral çeşitlilik mekanizması .....	6
2.2.6 hRSV aşısı çalışmalarındaki gelişmeler ve karşılaşılan zorluklar .....	6
2.3 hRSV'nin Antijenik Dizi Çeşitliliği .....	7
2.4 hRSV Dizi Çeşitliliği ve Dinamiklerini Çözmek İçin Kullanılan Biyoinformatik Yaklaşım .....	8
<b>3. ARAŞTIRMA YÖNTEMİ .....</b>	<b>9</b>
3.1 Projeye Genel Bakış .....	9
3.2 Verilerin Toplanması ve Sınıflandırılması .....	12
3.3 Verilerin İşlenmesi .....	12
3.3.1 Veri birleştirilmesi, temizliği ve deduplikasyonu .....	12
3.3.2 Çoklu dizi hizalaması .....	13
3.3.3 Çoklu dizi hizalaması sonuçlarının kontrolü .....	13
3.4 Proteome Dizi Çeşitliliğinin Karakterizasyonu .....	13
3.5 Çeşitlilik Motiflerinin Karakterizasyonu ve Kuantifikasyonu .....	14
3.6 Korunmuş ve Çeşitlilik Gösteren Dizilerin Dağılımı ve Aşısı Hedefi Olabilecek Aday Epitopların Belirlenmesi .....	15

3.7 Epitop Tahmini .....	16
3.8 Moleküler Docking Analizleri.....	16
3.9 Moleküler Dinamik Simülasyonları .....	17
<b>4. BULGULAR .....</b>	<b>18</b>
4.1 Analizde Kullanılan Verilere Ait Bilgiler .....	18
4.2 Proteom Dizi Çeşitliliği.....	22
4.3 Çeşitlilik Motiflerinin Dinamiği.....	27
4.4 Korunmuş ve Değişkenlik Gösteren Dizilerin Dağılımı .....	34
4.5 Çok Yüksek Derecede Korunmuşluk İçeren ve Potansiyel Aşı Hedefi Olabilecek İmmünojenik Diziler .....	35
4.6 Moleküler Docking Sonuçları .....	37
4.7 Moleküler Dinamik Simülasyonlarının Sonuçları.....	39
<b>5. TARTIŞMA VE SONUÇ .....</b>	<b>46</b>
<b>KAYNAKLAR .....</b>	<b>48</b>
<b>EKLER .....</b>	<b>56</b>
<b>ÖZGEÇMİŞ.....</b>	<b>60</b>

## KISALTMALAR

<b>ASN</b>	: Asparajin
<b>ASP</b>	: Aspartik Asit
<b>CTL</b>	: Sitotoksik T hücresi
<b>ÇYDÇS</b>	: Çok Yüksek Derecede Çeşitliliğe Sahip
<b>DiMA</b>	: Çeşitlilik Motif Çözümleyici
<b>GLN</b>	: Glutamin
<b>GLU</b>	: Glutamik Asit
<b>HIV</b>	: İnsan Bağışıklık Yetmezliği Virüsü
<b>HLA</b>	: İnsan Lökosit Antijeni
<b>hRSV</b>	: İnsan solunum sinsityal virüsü
<b>HTL</b>	: Yardımcı T hücresi
<b>IEDB</b>	: İmmun Epitop Veritabanı
<b>KDİ</b>	: Karışık Değişkenleri İçeren
<b>M</b>	: Molarite
<b>MHC I</b>	: Majör Histokompatibilite Kompleksi 1
<b>MHC II</b>	: Majör Histokompatibilite Kompleksi 2
<b>NaCl</b>	: Sodyum Klorür
<b>NCBI</b>	: Ulusal Biyoteknoloji Bilgi Merkezi
<b>OPLS</b>	: Sıvı Simülasyonlar için Optimize edilmiş Potansiyel
<b>RMSD</b>	: Kök-ortalama-kare sapması
<b>RNA</b>	: Ribonükleik Asit
<b>RSV</b>	: Solunum sinsityal virüsü
<b>SER</b>	: Serin
<b>SK</b>	: Süpertip Kısıtlamalı
<b>TCR</b>	: T Hücresi Reseptörü
<b>THR</b>	: Treonin
<b>Tip5P</b>	: 5 üssü için Transfer edilebilen İntermoleküler Potansiyel
<b>TK</b>	: Tamamen Korunmuş
<b>TRP</b>	: Triptofan
<b>TYR</b>	: Tirozin
<b>VIPR</b>	: Virüs Patojen Veritabanı ve Analiz Kaynağı
<b>YDÇS</b>	: Yüksek Derecede Çeşitliliğe Sahip
<b>YDKS</b>	: Yüksek Derecede Korunmuşluğa Sahip

## SEMBOLLER

$\sim$	: Yaklaşık olarak
%	: Yüzde
$\geq$	: Büyük ve eşit
'	: Üs
Å	: Angstrom
c	: Dizi benzerlik eşik değeri
n	: Dizi uzunluğu
H(x)	: Entropi
p(i,x)	: Nonamer i dizisinin x pozisyonunda bulunma olasılığı
n(x)	: Toplam nonamer dizisi
log	: Logaritma
r	: Pearson korelasyon katsayısı

## TABLO LİSTESİ

	<u>Sayfa</u>
<b>Tablo 2.2.2</b> : hRSV proteinlerinin görevleri.....	4
<b>Tablo 4.1.1</b> : Analiz edilen hRSV dizilerine ait bilgiler.....	18
<b>Tablo 4.2.1</b> : Analiz edilen hRSV dizilerine ait entropi ve total varyant istatistikleri.....	22
<b>Tablo 4.3.1</b> : M2-2 proteini için kuantitatif dizi çeşitliliği analizine ait örnek tablo. ....	27
<b>Tablo 4.5.1</b> : IEDB veritabanına göre deneysel olarak ispatlanmış hRSV epitopları. ....	34
<b>Tablo 4.5.2</b> : hRSV MHC-I epitopları ve onlara karşılık gelen süpertip kısıtlamalı... alellere ait örnek tablo.....	35
<b>Tablo 4.5.3</b> : hRSV MHC-II epitopları ve onlara karşılık gelen süpertip kısıtlamalı.. alellere ait örnek tablo.....	36

## ŞEKİL LİSTESİ

### Sayfa

Şekil 2.2.2 : hRSV virion yapısı ve genom kompozisyonu.....	3
Şekil 2.2.4 : hRSV patogenezi etkileyen faktörler.. ..	5
Şekil 3.1 : Araştırma yöntemine genel bakış.....	10
Şekil 4.1.1 : NCBI Virus veritabanında bulunan dizilerin zamana bağlı olarak dağılımı.....	19
Şekil 4.1.2 : NCBI Virus veritabanında bulunan dizilerin coğrafi dağılımı.....	21
Şekil 4.1.3 : VIPR veritabanında bulunan dizilerin zamana bağlı olarak dağılımı.. ..	21
Şekil 4.1.4 : VIPR veritabanında bulunan dizilerin coğrafi dağılımı. ....	22
Şekil 4.2.1 : hRSV'nin her bir proteini için pozisyonlara ait entropi ve total varyant değerlerini içeren stem grafiği.....	25
Şekil 4.2.2 : hRSV proteomunun entropi ve total varyantları arasındaki ilişki.....	26
Şekil 4.3.1 : hRSV dizi çeşitliliği motiflerinin proteom düzeyinde dinamikleri. ....	29
Şekil 4.3.2 : hRSV dizi çeşitliliği motiflerinin protein düzeyinde dinamikleri. ....	31
Şekil 4.3.3 : hRSV dizi çeşitliliği motiflerinin dağılımını inceleyen violon (keman) grafiği. ....	33
Şekil 4.4.1 : Protein ve proteom düzeyinde hRSV nonamerlerinin korunmuşluk derecelerini gösteren grafik.....	34
Şekil 4.6.1 : HLA-A-11-01 aleli ile F-ASISQVNEK nonamerinin moleküler docking pozunu. ....	38
Şekil 4.6.2 : HLA-B-15-03 aleli ile F-VQIVRQQSY nonamerinin moleküler docking pozunu. ....	38
Şekil 4.6.3 : HLA-A-03-01 aleli ile L-KLNEIHLMK nonamerinin moleküler docking pozunu. ....	38
Şekil 4.7.1 : HLA-A-11-01 aleli ile F-ASISQVNEK nonamerinin moleküler dinamik simülasyonu sonuçları. ....	39
Şekil 4.7.2 : F-ASISQVNEK nonamerinin simülasyon boyunca bağ yaptığı amino asitler ve bağlandığı sürenin oranları. ....	41
Şekil 4.7.3 : HLA-B-15-03 aleli ile F-VQIVRQQSY nonamerinin moleküler dinamik simülasyonu sonuçları.....	42
Şekil 4.7.4 : F-VQIVRQQSY nonamerinin simülasyon boyunca bağ yaptığı amino asitler ve bağlandığı sürenin oranları. ....	43
Şekil 4.7.5 : HLA-A-03-01 aleli ile L-KLNEIHLMK nonamerinin moleküler dinamik simülasyonu sonuçları.....	44
Şekil 4.7.6 : L-KLNEIHLMK nonamerinin simülasyon boyunca bağ yaptığı amino asitler ve bağlandığı sürenin oranları. ....	45
Şekil 5.1 : Potansiyel aşı hedefi olan F-ASISQVNEK nonameri.....	47

## İnsan Solunum Sinsityal Virüsünün Protein Dizisi Çeşitliliği Dinamiklerinin İncelenmesi

### ÖZET

İnsan solunum sinsityal virüsü (hRSV), dünya çapında en bulaşıcı virüslerden biridir ve bronşiolit ve pnömoni gibi alt solunum yolu enfeksiyonlarına yol açabilir. Lisanslı profilaktik hRSV aşısı bulunmamaktadır. hRSV'nin büyük ölçekli proteom çapında dizi çeşitliliği analizinde kullanılacak ve epitop bazlı hRSV aşısı tasarımı için potansiyel aşı hedeflerinin tanımlanmasına yol açacak çok sayıda hRSV dizisi halka açık veri tabanlarında mevcuttur.

Bu çalışma, virüsün mevcut tüm protein dizilerine odaklanmıştır. Dizi verileri, halka açık NCBI Virüs ve Virüs Patojen Veritabanı ve Analiz Kaynağı (VIPR) veritabanlarından toplanmıştır. Protein dizisi verileri havuz oluşturularak birleştirilmiş ve benzer diziler CD-HIT kullanılarak çıkarılmıştır. Bireysel protein veri kümeleri oluşturmak için 11 hRSV proteininin UniProt referans kayıtları kullanılarak toplanan diziler üzerinde BLASTp analizi yapılmıştır. Her protein veri seti, MUSCLE programı kullanılarak hizalanmıştır. Dizi çeşitliliği, proteinin uzunluğu boyunca örtüşen her 9-mer (nonamer) (1-9, 2-10, vb.) pozisyonu için Shannon entropisi kullanılarak ölçülmüştür. Bundan sonra, T-hücre epitopları, IEDB veritabanı kullanılarak deneysel olarak ispatlanan epitoplar ile eşleştirilmiş ve IEDB'nin önerdiği yardımcı araçlar kullanılarak epitop tahmini yapılmıştır. Tahmin sonuçlarımızı desteklemek için, seçilen epitopların MHC I ve II'ye karşı moleküler docking analizleri ve moleküler dinamik simülasyonları yapılmıştır.

VIPR ve NCBI Virus veritabanlarından toplanan toplam dizi sayısı 93.382'dir. Benzer dizilerin çıkartılmasından sonra, sayı 12.413'e düşmüştür (~%87'lik bir azalma). hRSV için gözlemlenen maksimum entropi değeri, M2-2 proteininde 43'üncü nonamer pozisyonunda ~4.2 olarak bulunmuştur. Bununla birlikte, proteom çapında ortalama entropi genel olarak düşüktür (~0.8) ve bu nedenle yüksek korunmuşluk göstermektedir ve bu değerler ~0.5 (protein N) ile ~2.3 (protein M2-2) arasındadır. Varyantların çoğu (~%55) 0 ile 1 entropi aralığı arasındadır. Nonamer pozisyonlarının yaklaşık yarısı (~%51) yüksek oranda korunmuş olarak sınıflandırılmıştır (indeks insidansı  $\geq$  %90). Epitop tahmini için 2269 yüksek oranda korunmuş nonamer dizisi seçilmiştir. 2269 nonamer dizisinden, sırasıyla MHC Sınıf I ve II için 997 ve 235 epitop tahmin edilmiştir.

Bu çalışma, hRSV proteomundaki dizi çeşitliliği ve bunların potansiyel aşı hedefleri hakkında bilgi sağlamaktadır. Proteomun ve tahmin edilen epitoplarının yüksek düzeyde korunması, profilaktik hRSV aşısı tasarımı için daha fazla araştırmanın gerekliliğini ortaya koymaktadır.

**Anahtar Kelimeler:** Viral dizi analizi, immünoinformatik, aşı hedef keşfi

# Dissecting the Dynamics of Human respiratory syncytial virus (hRSV) Protein Sequence Diversity

## SUMMARY

Human respiratory syncytial virus (hRSV) is one of the most contagious viruses worldwide, which can lead to lower respiratory tract infections, such as bronchiolitis and pneumonia. There is no licensed prophylactic RSV vaccine. A large number of hRSV sequences are available in public databases, which were used in a large-scale proteome-wide sequence diversity analysis herein of hRSV and that led to the identification of potential vaccine targets for epitope-based hRSV vaccine design.

This study focussed on all available protein sequences of the virus. Sequence data was collected from publicly available NCBI Virus and Virus Pathogen Database and Analysis Resource (VIPR) databases. Protein sequence data were pooled, and duplicates were removed by use of CD-HIT. BLASTp search was performed on the collected sequences using UniProt reference records of the 11 hRSV proteins to generate individual protein datasets. Each protein dataset was multiple sequence aligned using MUSCLE. Sequence diversity was measured by use of Shannon's entropy for each overlapping 9-mer (nonamer) (1–9, 2–10, etc.) position across the length of the protein. From there on, T-cell epitopes were matched to experimentally validated ones in the IEDB database and were also predicted using the tools recommended by the IEDB resource benchmark. To support our prediction results, molecular docking experiments of selected epitopes against human leucocyte antigen (HLA) I and II and molecular dynamic simulations were performed.

The total number of sequences collected from VIPR and NCBI was 93,382. After the removal of duplicates, the number decreased to 12,413 (an ~ 87% reduction). The peak entropy value observed for hRSV was ~4.2 in the M2-2 protein at starting nonamer position 43. However, the proteome-wide average entropy was low (~0.8) and thus, indicating high conservation, with values ranging from ~0.5 (protein N) to ~2.3 (protein M2-2). Most of the variants (~55%) were between the entropy range of 0 to 1. Approximately half (~51%) of the nonamer positions were classified as highly conserved (index incidence  $\geq 90\%$ ). The number of 2269 highly conserved nonamer sequences were selected for epitope prediction. From the 2269 nonamer sequences, 997 and 235 epitopes were predicted to be restricted to HLA Class I and II, respectively.

This study provides information about sequence diversity across the proteome of hRSV and identified potential vaccine targets.. The high conservation of the proteome and their predicted epitopes merits further investigation for prophylactic hRSV vaccine design.

**Keywords:** Viral sequence analysis, immunoinformatics, vaccine target discovery

# 1. GİRİŞ

## 1.1 Tezin Kapsamı ve Amacı

Halka açık veritabanlarında sürekli artmakta olan viral diziler virüslere ait dizi çeşitliliğinin çözülmesi gerekliliğini ortaya koymaktır. hRSV protein ve nükleotit sayılarındaki artışlar virüsün neden olduğu hastalıklara karşı önleyici tedaviler geliştirilmesine yardımcı olmaktadır. Bu tez hRSV genotip A ve B'nin kantitatif analizine odaklanmaktadır. hRSV protein dizilerini karakterize etmek için viral varyant dizilerinin farklı popülasyonlarının oluşumu ve bunların bulunma değerleri bağışıklık sistemine bağlı analizler ile değerlendirilmiştir. Tez kapsamında cevap alınması beklenen araştırma soruları ise şunlardır:

1. hRSV proteom çeşitliliğini oluşturan viral dizi çeşitliliği dinamiklerinin motifleri nelerdir?
2. Bu motifler birbirleri ile nasıl korelasyon göstermektedir?
3. Antijenik dizi çeşitliliğinin hRSV aşı tasarımına etkileri nelerdir?

Ayrıca araştırmanın hedeflediği çıktılar ise sırasıyla hRSV antijenik proteom dizi çeşitliliğinin immun sistem kapsamında karakterize edilmesi, hRSV antijenik dizi çeşitliliği dinamiklerinin kuantifikasyonu ve virüse karşı epitop temelli aday aşı hedeflerinin belirlenmesidir.

## 1.2 Hipotez

Araştırmanın hipotezi, hRSV çeşitlilik dinamiklerinin viral fitness seçiminde görev alan ve proteomun kendi organizasyonu içindeki var olan karakteristik dizi motifleri ile temsil edildiğidir.

## **2. LİTERATÜR ARAŞTIRMASI**

### **2.1. Hastalığın Epidemiyolojisi**

İnsan solunum sinsityal virüsü bronşit ve zatürre gibi ölümcül olabilecek hastalıklara neden olan bir virüstür. Genellikle çocuklarda ve genç yetişkinlerde görülmekle birlikte yaşlı bireylerde de görülme sıklığı gittikçe artmaktadır. Dünya genelinde etki gösteren hRSV, çocuklarda ve genç yetişkinlerde her sene ortalama 30 milyon kişinin akut solunum yolu enfeksiyonu geçirmesine neden olmaktadır. Çocuk yaşta olan hasta bireyler için ise yılda ortalama 60,000'den fazla ölüm gerçekleşmektedir [1]. Enfeksiyon sonucu oluşan morbidite ve mortalite sayıları ülkeler arasında çeşitlilik göstermektedir. Örnek olarak Brezilya'da enfekte olup çocuk servisine yatan hastaların oranı %35 [2], İtalya'da %22, Belçika'da %35 ve İngiltere'de %59 olarak görülmüştür [3].

#### **2.1.1 Mevsimsellik**

İnsan solunum sinsityal virüsünün mevsimselliği çeşitli araştırmalarda ortaya çıkarılmıştır. Bu virüsün neden olduğu enfeksiyonlar yıl boyunca etkinlik göstermekle birlikte en çok kış aylarında dominant özellik göstermektedir [4]. Diğer enfeksiyonlar ise daha çok Kasım ve Aralık aylarında gerçekleşmektedir [5].

## **2.2 İnsan Solunum Sinsityal Virüsü**

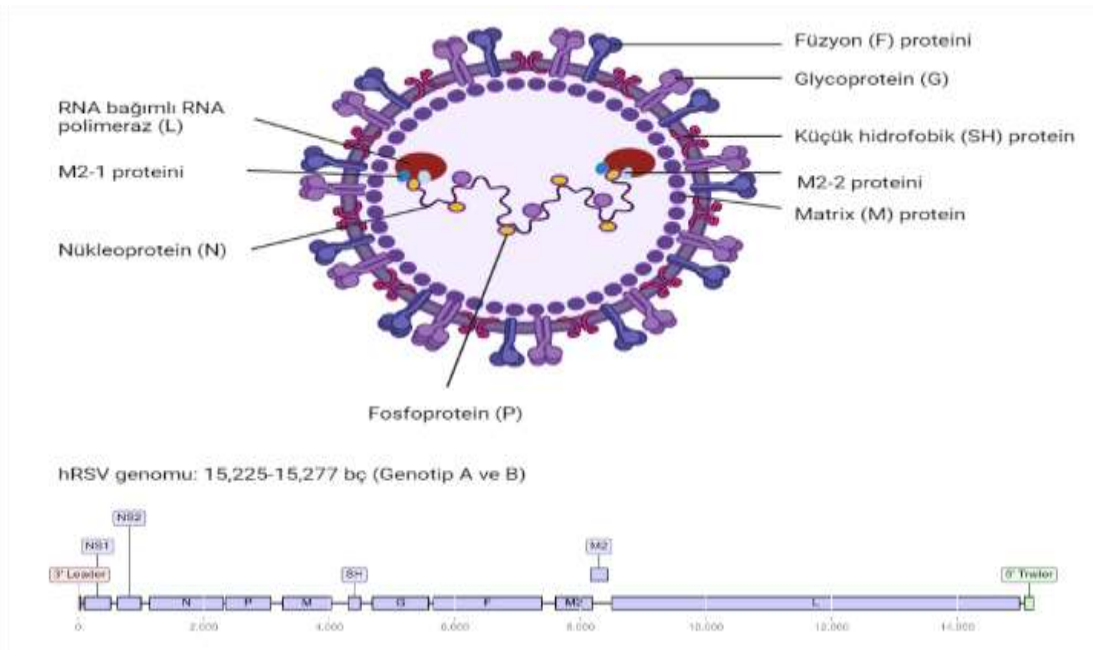
### **2.2.1 Tarihçe**

RSV ilk olarak 1955 yılında Walter Reed Army Araştırma Enstitüsünde (WRAIR, Washington, DC, ABD) 20 şempanzede görülen öksürme, aksırma, mukopürülan akıntı ile karakterize edilmiştir [6]. Viral izolasyon karaciğer hücrelerinde gerçekleştirilerek fare, hamster, tavşan ve şempanze gibi laboratuvar hayvanlarında inoküle edilmiş ve semptom geliştirme durumu kontrol edilmiştir. Daha sonraki çalışmalarda insanda izolasyon gerçekleştirme işlemi başarısız olmasında rağmen

şempanzeden elde edilen ajana karşı kompleman fiksasyonu ile antikor titresinde bir artış tespit edilmiş ve buna bağlı olarak gençlerde ve yetişkinlerde paralel olarak yeni çalışmalar başlatılmıştır. Sonraki yıllarda Chanock ve arkadaşları tarafından gerçekleştirilen çalışmada şempanzede bulunan ajanın bir benzeri zatürre hastası bir çocuktan ve kuş palazı hastalığı olan başka bir çocuktan izole edilmiştir [7,8]. İzole edilen bu ajana ise insan solunum sinsityal virüsü (hRSV) adını vermişlerdir.

## 2.2.2 Genom yapısı ve nomenklatür

İnsan solunum sinsityal virüsü (hRSV), taksanomik sınıflandırmada Negarnaviricota şubesi, Monjiviricetes sınıfı, Mononegavirales takımı, Pneumoviridae ailesi, Orthopneumovirus cinsi içerisinde bulunmaktadır [9,10,11]. hRSV, elektron mikroskopun da hem uzun ipliksi yapıda hemde küre şeklindeki parçacıklar halinde görülebilmektedir [12]. hRSV, non-segmented yapıda, tek zincirli ve negatif polariteli, yaklaşık 15.000 nükleotid içeren bir virüstür. Viral genomu toplamda 10 gen içerir ve bu genler 11 protein kodlamaktadırlar. Bu proteinlerin 9 tanesi yapısal, 2 tanesi (NS1-NS2) ise yapısal olmayan proteinlerdir [13]. 3'-5' yönünde RNA zincirine sahip hRSV genleri genom içersinde NS1-NS2-N-P-M-SH-G-F-M2-L şeklinde sıralanmaktadır [Şekil 2.2.2].



Şekil 2.2.2 : hRSV virion yapısı ve genom kompozisyonu. Referans [14]'ten uyarlanmıştır.

hRSV aşılarının geliştirilmesi için kullanılan stratejilerin arkasında yatan mantığı anlayabilmek için virüsün sahip olduğu proteinlerin görevlerinin çok iyi bilinmesi gerekmektedir. Bu bilgiler ve her bir protein için genetik çeşitlilik oranları Tablo 2.2.2’de özetlenmiştir.

**Tablo 2.2.2 : hRSV proteinlerinin görevleri.**

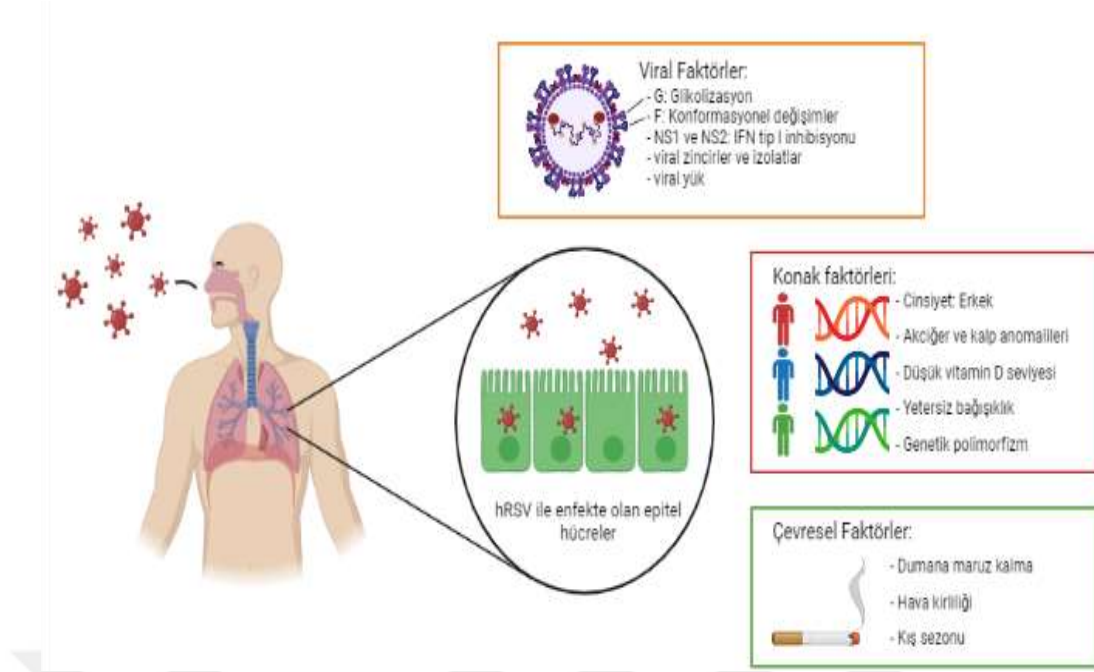
<b>Protein</b>	<b>Görevi</b>
NS1	Tip I interferon üretimini inhibe etmek <sup>[15,16,17]</sup>
NS2	Tip I interferon üretimini ve CD8+ T hücre sitotoksitesini inhibe etmek <sup>[18,19]</sup>
N	Genom kapsidasyonunda ve immünolojik sinaps oluşumuna müdahalede yer alır. <sup>[20,21]</sup>
P	Virial replikaz-transkriptaz kompleksinde L proteini için kofaktör olarak görev yapar. <sup>[22,23]</sup>
M	Virion oluşumunda görev alır ve viral transkriptaz aktivitesini inhibe eder. <sup>[24]</sup>
SH	Kasyon seçici iyon kanalları oluşturarak enfekte olmuş hücrelerin geçirgenliğine müdahale eder. <sup>[25]</sup>
G	Virüsün heparan sülfat gibi glikozaminglikanlara bağlanmasını kolaylaştırır. <sup>[26,27]</sup>
F	Virüs penetrasyonuna ve virus ile sitoplazmik membrane arasında füzyona aracılık eder. <sup>[28]</sup>
M2-1	Translasyonun anti-terminatörü olarak görev yapar. <sup>[29,30]</sup>
M2-2	Virial mRNA sentezini negatif olarak module eder. <sup>[31]</sup>
L	Virial transkripsiyonu ve replikasyonu gerçekleştirir. <sup>[32]</sup>

### **2.2.3 hRSV'nin sınıflandırılması**

Monoklonal antikörlerin kullanılması ile hRSV'nin iki antijenik gruba ayrıldığı gösterilmiştir. Bunlar Genotip A ve B şeklinde adlandırılır. Monoklonal antikörlerin kullanılmasının yanı sıra virüsün G protein dizisine dayanan moleküler düzeydeki diğer çalışmalar da bu sınıflandırmayı ispat etmiştir [33,34]. G proteinin kullanılmasının sebebi ise en fazla dizi çeşitliliğin bu proteine ait olmasıdır [35]. Günümüzdeki çalışmalarda ise bu sınıflandırmanın G geni üzerinden yapıldığı ancak bu genin farklı korunmuş bölgeleri kullanılarak yapıldığı literatürde görülmektedir. Bunlar, G geninin tamamı kullanılarak [36], HVR2 bölgesi kullanılarak [37] ve Ektodomain bölgesi kullanılarak [38] yapılan sınıflandırmalardır.

### **2.2.4. hRSV patogenezi**

hRSV patogenezinin anlaşılmasının güç olmasının nedenlerinden bazıları pulmoner immün hücrelerine dayanan çalışmaların çocuklarda uygulanmasının zor olması [39,40] ve birçok araştırmacının neonatal modeller yerine yetişkinler üzerinde çalışmalar yapmasıdır [41]. hRSV patogenezi etkileyen faktörler ise viral, konak ve çevresel faktörler olarak gösterilebilir [Şekil 2.2.4].



**Şekil 2.2.4 :** hRSV patogenezini etkileyen faktörler. Referans [42]'den uyarlanmıştır.

### 2.2.5 Viral çeşitlilik mekanizması

hRSV'nin non-segmented yapıda olması virüsün influenza virüsleri gibi antijenik şift ve drift gibi reassortment olaylarına uğramamasına yol açmaktadır. Bu yüzden viral çeşitliliğin mekanizmaları segmented virüslerin sahip olduğu mekanizmalardan farklılık göstermektedir. L geninin ürettiği RNA bağımlı RNA polimerazın proofreading ve editing kapasitelerinin az olması ve çevresel faktörlerin oluşturduğu seçim baskısı bu mekanizmalara örnek olarak gösterilebilir [43].

### 2.2.6 hRSV aşı çalışmalarındaki gelişmeler ve karşılaşılan zorluklar

hRSV aşı geliştirme çalışmalarının önem kazanmasının çeşitli nedenleri bulunmaktadır. Bunlara örnek olarak viral proteinlerin yapılarının ve fonksiyonlarının, virion morfolojisi ve yapısının, virüsün neden olduğu hastalıklara karşı oluşan immün tepkinin ve aşı sonucu hastalık derecesinin azalması yerine artması verilebilir. Bu durumlar hRSV için günümüzde etkili ve güvenli bir aşının bulunmasının gerekliliğini ortaya koymaktadır. Ayrıca günümüzde 30'dan fazla aşı adayının klinik ya da klinik öncesi çalışmalarının devam ettiği görülmektedir [44,45].

Parçacık temelli ve subunit temelli aşılarda yaşlılar üzerinde hastalığı ilerletici etkiye sahip olabileceği endişeleri günümüz bilim dünyasında süregelen bir durumdur. Bu

endişelerin yanı sıra, hRSV aşı geliştirme çalışmaları genellikle F ve G proteinleri üzerine yoğunlaşmış bulunmaktadır. İnaktive edilmiş aşılardan ve subunit temelli aşılardan hastalığı iletme ihtimali nedeniyle güvenlik açısından sıkıntı çıkarmaktadır. Canlı aşılardan ise böyle bir sorunu bulunmamaktadır. Son çalışmalara bakılacak olursa canlı aşı adayları olan ve sıcaklık hassasiyetine bağlı olarak mutasyon geliştirme ihtimali stabilize edilmiş bir hRSV aşısı 6 ve 24 aylık bebeklerde test edilmiştir. Çalışmanın sonucunda aşının genetik stabilitesinin artmış olduğu gösterilmiştir [46].

Birçok aşı adayının klinik ve klinik öncesi çalışmalarının devam etmesinin yanı sıra hRSV için efektif, güvenli ve hastalık özetinde bir tedavinin geliştirilememiş olması, ayrıca virüsün yol açtığı hastalığın global ölçekte etkili olması gibi nedenlerden dolayı aşı keşfetme çalışmaları bilim dünyasında büyük bir ilgi uyandırmıştır. Ancak hRSV

aşısı geliştirmenin çeşitli zorlukları bulunmaktadır. Neonatal adaptif ve doğuştan gelen immün tepkisindeki bozukluklar, bebek aşılama sırasında problem çıkarmaktadır. Ayrıca bebeklerde bulunan maternal antikorların varlığı onlara güçlü bir koruma sağlarken kendi antikor tepkilerinin baskılanmasına da yol açabilir [47].

Tüm bu nedenlerden dolayı hRSV için epitop temelli bir aşı geliştirilmesi gelecek açısından ümit verici olarak gözükmektedir.

### **2.3 hRSV'nin Antijenik Dizi Çeşitliliği**

hRSV, A ve B olmak üzere iki temel antijenik gruba ayrılmaktadır. Bu iki grup yaklaşık 350 sene önce G proteininin antijenik ve dizi temelli varyasyonları temel alınarak belirlenmiştir [48-51]. Ayrıca her bir antijenik grup için birçok genotip tanımlanmıştır [52,53]. Yapılan son çalışmalarda her antijenik gruba özel G proteinine ait genetik modifikasyonlar tespit edilmiştir. ON genotipi 72 nükleotitlik duplikasyon içerir ve A grubu ile ilişkilendirilir. BA genotipi ise 60 nükleotitlik duplikasyon içerir ve B grubu ile ilişkilendirilir [54-56]. Bu iki genotip viral dolaşımında global olarak zamanla dominant hale gelmiştir ve in vitro açıdan viral fitness ve bağlanma özelliğinin arttığı gözlemlenmiştir [57]. Ayrıca, hRSV-A ve hRSV-B gruplarındaki genetik varyasyonlar birincil epitopları tanıyan anti-G monoklonal antikorlarındaki

indüksiyon ile korelasyon göstererek hRSV için immun yönelimli evrime kanıt oluşturmaktadır [58].

## **2.4 hRSV Dizi Çeşitliliği ve Dinamiklerini Çözmek İçin Kullanılan Biyoinformatik Yaklaşım**

Dünya genelinde araştırmalar artık hipotez güdümlü olmaktan veri yönelimli olmaya doğru gitmektedir. Bu durum gelişen genomik ve proteomik alanlarından elde edilen veriler ile gittikçe belirgin hale gelmektedir. Bu sayede tüm genler ve proteinlerin çalışılabilmesi mümkün hale gelecektir. Bu durum elde edilen bilimsel veri hacminin artışına yol açmakta olup bu verilerin sistematik olarak analiz edilmesi için biyoinformatik yaklaşımların gerekliliğini ortaya koymaktadır. Ayrıca çeşitli kaynaklardan elde edilen verilerin manipülasyonu ve analizi için de biyoinformatik yaklaşımların gerekliliği elzem bir durumdur [59].

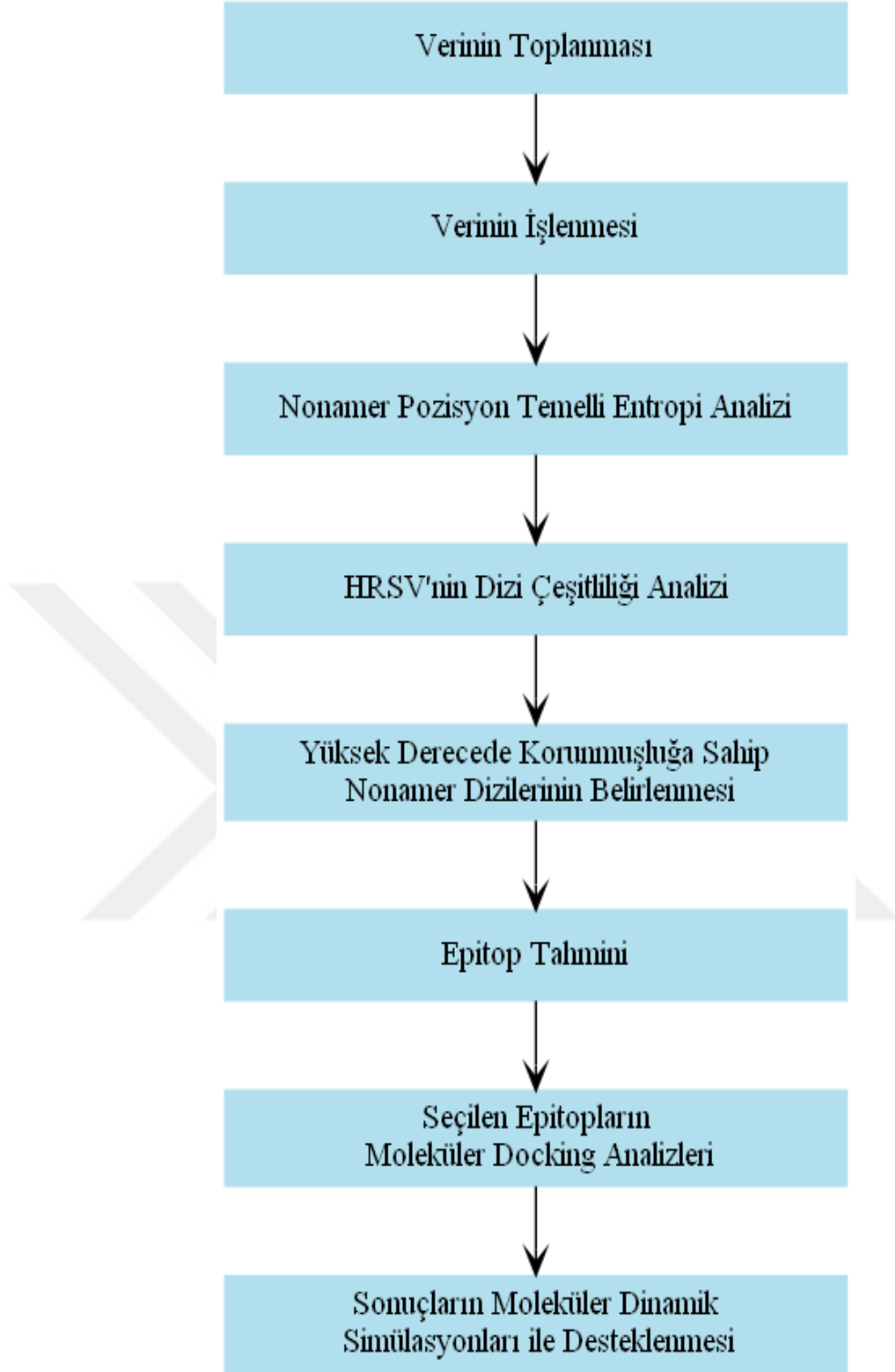
Bu tez çalışması hRSV protein dizi çeşitliliği ve dinamiklerini çözmek için biyoinformatik temelli bir yaklaşıma odaklanmış ve virüsün iki antijenik (A ve B) alt grubuna ait protein dizilerinden oluşturan veriseti kullanılarak gerçekleştirilmiştir. Halka açık veritabanlarından elde edilen bu veriseti (NCBI Virus, VIPR-Virus Pathogen Database and Analysis Resource) analiz edilerek viral proteinlerin korunmuş bölgeleri tespit edilmiş ve bu bölgelerden potansiyel olarak epitope olma ihtimali olan nanomer diziler IEDB – Immune Epitope Database kullanılarak belirlenmiştir. Çalışmadan elde edilen sonuçlar ise moleküler docking ve moleküler simülasyon deneyleri yapılarak desteklenmiş ve epitope temelli aşı çalışmalarına katkıda bulunması amaçlanmıştır.

### **3. ARAŐTIRMA YÖNTEMİ**

#### **3.1 Projeye Genel Bakıő**

Bu projede kullanılan biyoinformatik analizler Őekil 3.1’de gösterilmiőtir. Bu analizler ierisinde, veri toplama ve sınıflandırma ilk aőamada yer almaktadır. Bunu takip eden veri iőleme srecinde ise, verileri analiz iin uygun hale getirme ve deduplikasyon iőleminin gerekleőtirilmesi, verileri birleőtirme, birleőtirilen verilerin dizi hizalamala analizlerinin yapılması ve sonularının kontrol edilmesi, hizalanan dizilerin proteom dzeyinde dizi eőtillilięinin belirlenmesi, dizi eőtillilięi motiflerinin karakterizasyonu ve nicel olarak ölçümü, korunmuőluk ieren ve eőtillilik gösteren dizi blgelerinin daęılımının saptanması, potansiyel aőı adayı epitoplara belirlenmesi, molekler docking analizleri ve sonuların molekler dinamik simlasyonları ile desteklenmesi bulunmaktadır.





**Şekil 3.1** : Araştırma yöntemine genel bakış.

### **3.2 Verilerin Toplanması ve Sınıflandırılması**

İnsan solunum sinsityal virüsü (hRSV) A ve B alt tiplerinin tüm protein dizileri NCBI Virus [60] ve VIPR [61] veritabanlarından 31.03.2021 tarihinde indirilmiştir. İndirilen protein dizileri virüsün 11 proteini (F,G,L,M,M2-1,M2-2,N,NS1,NS2,P,SH) dikkate alınarak sınıflandırılmıştır. Bu araştırmanın virüsün insanda neden olduğu hastalıklara karşı profilaktif aşı çalışmalarına katkı sağlaması hedeflendiği için konak olarak insan seçilmiştir. Her iki veritabanında da protein dizilerinin farklı uzunluklardaki kayıtları bulunmaktadır. Bu durum, dizi hizalamalarında hatalara yol açabileceğinden dolayı her iki alt tipin proteinleri için NCBI veritabanından referans protein dizileri seçilmiştir. Dizi uzunlukları referans değerinde olan ve bu değerden en fazla %10 uzaklıkta olan diziler analiz için seçilmiştir.

### **3.3 Verilerin İşlenmesi**

#### **3.3.1 Veri birleştirilmesi, temizliği ve deduplikasyonu**

NCBI Virus ve VIPR veritabanlarından indirilen veriler tek dosya haline getirilerek birleştirilmiştir. Analizlerde sorun oluşturabilecek protein dizileri kontrol edilmiş ve veri setinden çıkartılmıştır. Her iki veritabanında da bulunan ortak diziler CD-HIT [62] programı kullanılarak tekilleştirilmiştir. Bu işlemde, dizi benzerliği eşik değerini ifade eden -c parametresi 1 olarak seçilmiştir. Benzerlik değerini hesaplamada kullanılacak olan ve dizi uzunluğunu belirleyen -n değeri ise varsayılan değer olan 5 olarak seçilmiştir. Bu değer, aynı zamanda programın 0.7-1 eşik değeri aralığı için önerdiği değerdir. Analiz sonuçlarının güvenilir olması ve verilerin sınıflandırılmasının doğruluğundan emin olmak için hata içeren protein dizilerinin üst verileri, NCBI [63] ve Uniprot [64] veritabanları kullanılarak kontrol edilmiş ve hatalar düzeltilmiştir. Yapılan bu işlemler, veri setinde oluşabilecek yanlışlığın da önüne geçilmesini sağlayarak verileri analize hazır hale getirmiştir.

### 3.3.2 Çoklu dizi hizalaması

Çoklu dizi hizalaması herhangi bir biyoinformatik analiz için hayati öneme sahip bir dizi karşılaştırma analizidir. Hâlihazırda birçok dizi hizalaması analizi yapan programlar bulunmakla birlikte, bu programlar hem açık kaynak kodlu olarak hem de ticari versiyon olarak kullanıcılara sunulmaktadır. Analize hazır protein veri seti, açık kaynak kodlu MUSCLE [65] programı kullanılarak analiz edilmiştir.

### 3.3.3 Çoklu dizi hizalaması sonuçlarının kontrolü

Çoklu dizi hizalaması sonuçları, her bir protein veri seti için AliView [66] programı kullanılarak kontrol edilmiştir. Analiz edilen 11 protein veri seti arasından sadece G proteininde 240'ıncı pozisyondan sonra yer yer boşluklar gözlemlenmiştir. Diğer proteinler için herhangi bir sorunla karşılaşılmamıştır.

### 3.4 Proteome Dizi Çeşitliliğinin Karakterizasyonu

Shannon entropisi [67], İnsan solunum sinsityal virüsü (hRSV) proteomunun dizi çeşitliliğini belirlemek için kullanılmıştır. Sliding k-mer window yaklaşımı ise hücrel bağışıklık tepkisi oluşturabilecek peptidlerin belirlenmesi için çoklu dizi hizalaması sonuçlarına uygulanmıştır. Bu yaklaşımda k değeri olarak 9 seçilmiştir. Çünkü, insan lökosit antijenleri (HLAs) ve T hücresi reseptör (TCR) bağlanma bölgeleri çoğunlukla bu uzunluktaki peptidleri (nonamer) tanımaktadır. Nonamer çeşitliliği, her üst üste gelen pozisyonlar için (1-9,2-10,3-11 vb.) entropi değeri belirlenerek hesaplanmıştır. Entropi değeri olan  $H(x)$ , dizi hizalaması sonuçlarındaki ilgili pozisyonlar için şu formül kullanılarak hesaplanmıştır.

$$H(x) = - \sum_{i=1}^{n(x)} p(i, x) \log_2 p(i, x)$$

Bu formülde  $p(i,x)$ ,  $i$  nonamer dizisinin  $x$  pozisyonunda bulunma olasılığını ifade eder. Entropi değeri  $H(x)$ ,  $n(x)$  değeri olan toplam nonamer dizisi arttıkça artış gösterir. Entropi değeri,  $x$  pozisyonunda bulunan dizi çeşitliliğini belirtir ve aynı zamanda  $x$  pozisyonunda bulunan nonamerlerin birbirleri ile göreceli olasılığına da bağlıdır. İlgili nonamer dizisinin  $x$  pozisyonunda dominant olması o pozisyon için entropi değerini azaltıcı etki gösterir. Bu durum,  $x$  pozisyonunun nonamer dizileri açısından korunmuşluk gösterdiğini kanıtlar. Entropi değerini etkileyen bir başka faktör ise dizi hizalamasında bulunan proteinlerin sayısıdır. Paninski kuralına göre dizi hizalamasındaki proteinlerin sayısı  $N$  olarak ifade edilirse hizalamada oluşabilecek istatistiksel yanlılık  $N$  ile ters orantılı olarak değişim göstermektedir [68]. Her  $x$  pozisyonu için entropi değerinde  $N$  sayısını ele alarak istatistiksel olarak düzenleme yapılması analiz sonuçlarının doğruluğu açısından büyük önem arz etmektedir. Anlatılan tüm bu aşamaların analizi, DiMA (Diversity Motif Analyzer) [69] programı kullanılarak gerçekleştirilmiştir.

### 3.5 Çeşitlilik Motiflerinin Karakterizasyonu ve Kuantifikasyonu

Entropi değeri protein dizi çeşitliliği hakkında genel olarak bilgi sahibi olmamızı sağlar. Bu çeşitliliğin nicel ölçümü ve karakterizasyonu için dizi çeşitliliği motiflerini incelemek gereklidir. Motiflerin karakterizasyonu çoklu dizi hizalamasındaki her bir nonamer pozisyonu için o pozisyonda bulunan farklı nonamer dizilerinin nicel ölçümü yapılarak gerçekleştirilmiştir [70]. En çok tekrar eden nonamer dizisi “index” olarak adlandırılmıştır. Bunun dışında kalan nonamerler ise “varyantlar” olarak isimlendirilmiştir ve index nonamer dizisinden en az 1 amino asiti farklıdır. Varyantların içinde en sık bulunan nonamer dizisi “major”, major motifinden daha az sıklıkta olan ve birden fazla bulunan nonamer dizisi ise “minor” olarak tanımlanmıştır. Tek olarak bulunan çeşitlilik motifine ise “özgün” denilmiştir. Bu motiflere ek olarak “nonatipler” tüm motiflerin (index,major,minor,özgün)

nonamerleri toplamının dizi hizalamasındaki tüm dizilerin toplamına bölünmesi sonucu hesaplanmış ve proteom düzeyinde dizi dinamiklerini açığa çıkarmıştır.

### **3.6 Korunmuş ve Çeşitlilik Gösteren Dizilerin Dağılımı ve Aşı Hedefi Olabilecek Aday Epitopların Belirlenmesi**

Korunmuşluk içeren ve çeşitlilik gösteren peptidleri incelemek için index motifi kendi arasında 5 kategoriye ayrılmıştır:

1. Çok yüksek derecede çeşitliliğe sahip (index değeri %10'dan küçük)
2. Yüksek derecede çeşitliliğe sahip (index değeri %20'den küçük, %10'dan büyük ve eşit)
3. Karışık değişkenleri içeren (index değeri %90'dan küçük, %20'den büyük ve eşit)
4. Yüksek derecede korunmuşluğa sahip (index değeri %100'den küçük, %90'dan büyük ve eşit)
5. Tamamen korunmuş (index değeri %100'e eşit)

Yüksek derecede korunmuşluğa sahip ve tamamen korunmuş olan nonamer dizileri çok düşük entropi değerlerine sahip olmakla birlikte evrimsel açıdan da dirençlilik gösterirler. Bu yüzden aşı adayı olabilecek epitoplar bu kategorilere ait dizilerden seçilmiştir ve amino asit dizileri üst üste gelen nonamerler birbirleri ile birleştirilmiştir. Birleştirilen nonamerlerin immünojenitesi IEDB [71] veritabanı kullanılarak araştırılmıştır. Virüse ait deneysel olarak ispatlanan ve rapor edilen epitoplar da IEDB veritabanı kullanılarak belirlenmiştir.

### 3.7 Epitop Tahmini

CD8+ T hücresi (CTL) epitopları, MHC-I bağlanma kapasitelerini hesaplayan IEDB'nin web tabanlı tahmin programı ile (<http://tools.iedb.org/mhci/>) ortaya çıkarılmıştır. Yüksek derecede korunmuşluğa sahip ve tamamen korunmuş olan nonamer dizileri program için giriş verisi olarak kullanılmış ve tahmin metodu olarak veritabanının önerdiği NetMHCPan 4.1 EL [72] algoritması kullanılmıştır. Epitopların, HLA (İnsan Lökosit Antijeni) bağlanma kapasitelerini ölçmek için 12 süperipten (A1,A2,A3,A24,A26,B7,B8,B27,B39,B44,B58,B62) toplam 95 alel seçilmiştir. Eşik değeri olarak %1'den küçük skora sahip [73,74] (iyi bağlanma kapasitesi gösteren) olan epitoplar ve onlara karşılık gelen HLA alelleri ileri analizler için seçilmiştir. Aynı zamanda, CD4+ T hücresi (HTL) epitopları da, MHC-II bağlanma kapasitelerini hesaplayan IEDB'nin web tabanlı tahmin programı ile (<http://tools.iedb.org/mhcii/>) analiz edilmiştir. MHC-I'den farklı olarak MHC-II giriş verileri için 15 amino asit uzunluğundaki dizilerin kullanılması gerekmektedir. Bu uzunluktaki diziler birleştirilen nonamer dizilerinden üretilmiştir. Tahmin metodu olarak veritabanının önerdiği IEDB recommended seçeneği kullanılmıştır. Bu metodun içeriğinde farklı aleller için olmak üzere Consensus, NN-align, SMM-align, CombLib ve Sturniolo olarak adlandırılan beş farklı algoritma bulunmaktadır. Eğer bağlanma kapasitesi ölçülmek istenen HLA aleli bu algoritmalarından biri ile tahmin edilemiyorsa NetMHCIIpan algoritması ile veriler analiz edilmektedir. Analiz sonuçları algoritmaların ortak skorları kullanılarak üretilmiştir. MHC-II sınıfındaki epitopların tahmini için üç süperipte (DP,DQ,DR) ait toplam 45 alel seçilmiştir. Eşik değeri ise %10 olarak belirlenmiştir [75]. Tahmin edilen epitoplar ve onlara karşılık gelen aleller ileri analizler için seçilmiştir.

### 3.8 Moleküler Docking Analizleri

Her bir proteine ait epitop tahmin sonuçları incelenerek moleküler docking analizleri için hazır hale getirilmiştir. Hesaplama gücü ve analiz süreci gibi etkenler göz önüne alınarak her protein için sonuçların %1'i kadar aday epitoplar hem MHC-I hem de MHC-II için belirlenmiştir. Eğer sonuçların %1'i sıfırın altında ise o proteine ait en iyi tahmin skoruna sahip aday epitop docking analizi için seçilmiştir. Moleküler docking

analizleri için belirlenen epitoplara Schrödinger Maestro [76] arayüzü ile Ligprep [77] Ionizer modülü kullanılarak 3 boyutlu hale getirilmiştir ve OPLS4 [78] kuvvet alanı kullanılarak minimize edilmiştir. Hazırlanmak istenen aday epitoplara için pH değeri 7-7,2 aralığında seçilmiştir. Aday epitoplara karşılık gelen HLA alelleri ise Protein Preparation Wizard [79,80] kullanılarak su molekülleri ve tampon moleküllerinden arındırılmıştır. N ve C terminal uçlarına sırasıyla N-asetil ve N-metil amid eklenmiş ve eksik yan zincirler varsa ekleme yapılarak analizlere hazır hale getirilmiştir. Moleküler docking analizleri ise Glide Ligand Docking [81-84] modülü kullanılarak gerçekleştirilmiştir. Tüm ligandlar hedef proteinin bağlanma bölgesine dock edilmiştir. Protein ve ligand arasındaki maksimum uzaklık 5 Å olarak seçilmiştir. MHC-I için 51, MHC-II için sekiz olmak üzere toplamda 59 docking analizi gerçekleştirilmiştir. Analizler tamamlandıktan sonra pozlar ve skorlar incelenmiştir. Yüksek skora ve hidrojen bağı, elektrostatik etkileşimler ve hidrofobik etkileşimlere sahip, çakışma göstermeyen aday epitoplara belirlenerek moleküler dinamik simülasyonları için hazırlanmıştır.

### **3.9 Moleküler Dinamik Simülasyonları**

Moleküler docking sonuçları incelenerek elde edilen aday epitoplara ve onlara karşılık gelen HLA alelleri birleştirilerek yeni bir kompleks elde edilmiştir. Elde edilen her bir kompleks ortorombik kutu içerisine kompleks ve kutu arasındaki maksimum uzaklık 10 Å olacak biçimde yerleştirilmiştir. Daha sonra her ortorombik kutunun içi Tip5P su molekülleri ve 0,15 M NaCl kullanılarak doldurulmuştur. Na<sup>+</sup> ve Cl<sup>-</sup> iyonlarının miktarı ayarlanarak sistem nötralize hale getirilmiştir. Tüm ağır atomlar analiz için sınırlandırılmış ve 100 pikosaniyede OPLS4 kuvvet alanı kullanılarak sistem minimizasyonu gerçekleştirilmiştir. Oluşturulan sistemlerin her biri 250 nanosaniyede izotermik (Nose-Hoover chain, 1 pikosaniye relaksasyon zamanı) ve izobarik (Martyna-Tobial-Klein, 2 pikosaniye relaksasyon zamanı) ortamda simülasyona uğratılmıştır. Bu analizler için Schrödinger Maestro arayüzü ile açık kaynak kodlu Desmond [85,86] programı kullanılmıştır.

## **4. BULGULAR**

### **4.1 Analizde Kullanılan Verilere Ait Bilgiler**

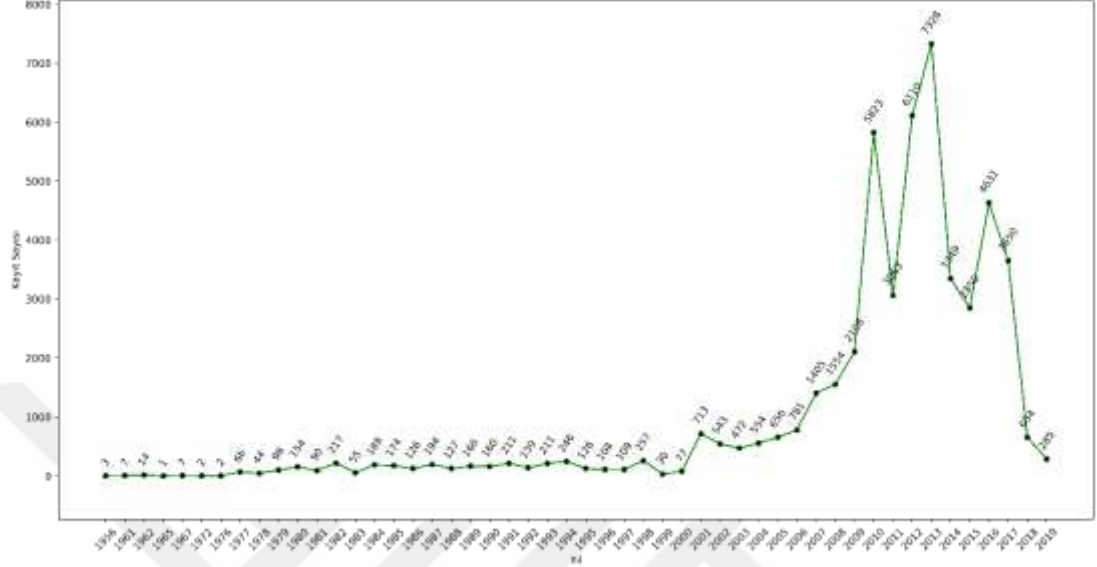
NCBI Virus ve VIPR veritabanlarından indirilen insan solunum sinsityal virüsü protein dizilerine ait birleştirilen ve tekilleştirilen veri seti, virüsün 11 proteini için toplam 13770 kayıttan oluşmaktadır. Bu sayı, veri setinin protein dizilerine göre sınıflandırılması ve tekrar tekilleştirilmesi işleminden sonra 5766'ya düşmüştür. Virüsün her iki alt tipinin de benzer uzunlukluktaki protein dizilerine sahip olduğu görülmektedir (Tablo 4.1.1). Ayrıca, proteinlerin kendi içindeki dizi çeşitliliği de en düşük benzerlik oranına bakılarak incelenebilir. Virüsün en fazla çeşitliliğe sahip proteini en düşük benzerlik oranı %35,39 olan G proteini iken en fazla korunmuş bölge içeren proteini ise %75,54 ile L proteini olarak bulunmuştur.

**Tablo 4.1.1 :** Analiz edilen hRSV dizilerine ait bilgiler.

<b>Protein</b>	<b>Genotip A Uzunluğu</b>	<b>Genotip B Uzunluğu</b>	<b>Minimum Benzerlik (%)*</b>	<b>Toplam İndirilen Dizi</b>	<b>Toplam Özgün Dizi</b>
F	574	574	64.11	2284	1134
G	298	299	35.39	5424	1606
L	2165	2166	75.54	2734	1350
M	256	256	53.91	274	128
M2-1	194	195	58.46	486	237
M2-2	90	90	38.89	513	253
N	391	391	74.68	432	206
NS1	139	139	60.43	334	166
NS2	124	124	68.55	488	241
P	241	241	69.29	510	255
SH	64	65	50.77	291	190

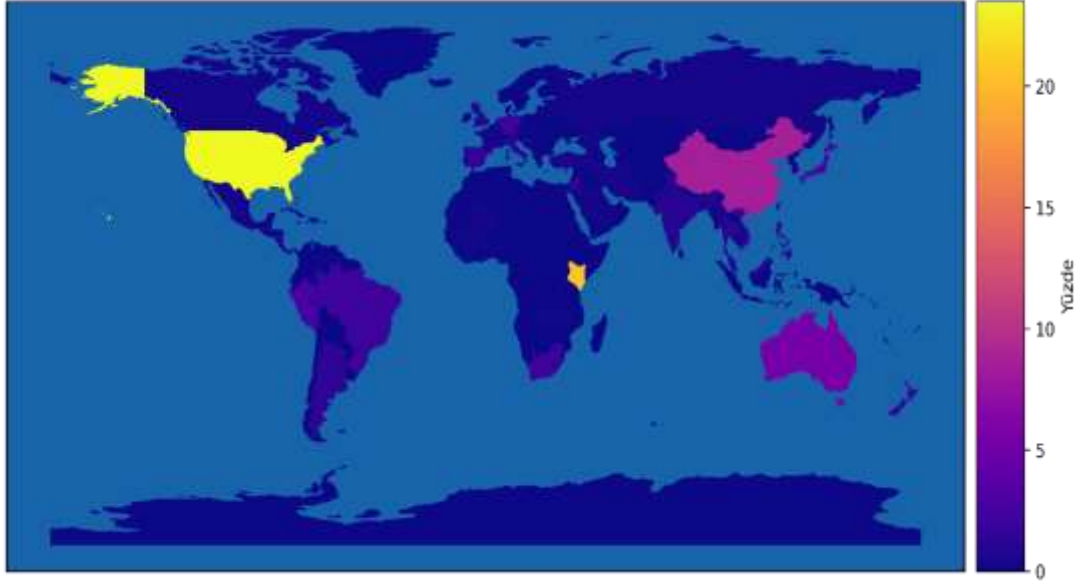
Not. Minimum Dizi Benzerlik Oranı her bir protein için çok dizi hizalaması sonuçları kullanılarak hesaplanmıştır.

NCBI Virus veritabanından indirilen veri seti incelendiğinde protein dizilerinin dağılımının 1956 ve 2019 yılları arasında olduğu görülmektedir. En fazla protein dizisi eklenen yıl ise 2013'tür ve eklenen protein dizisi kaydı 7328'dir (Şekil 4.1.1).



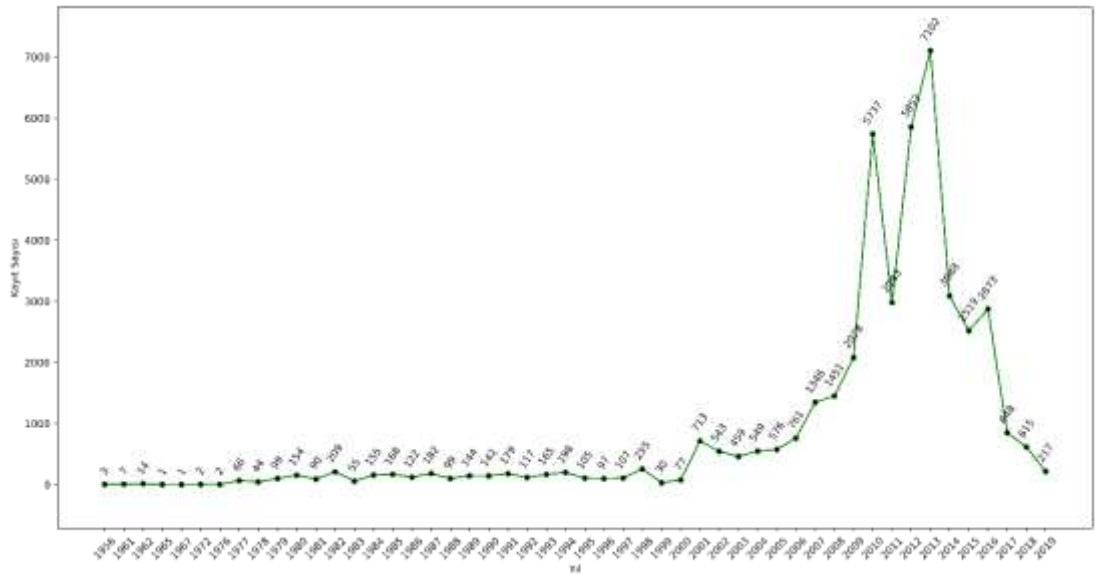
**Şekil 4.1.1 :** NCBI Virus veritabanında bulunan dizilerin zamana bağlı olarak dağılımı.

NCBI Virus veritabanından indirilen veri setindeki protein dizileri Asya, Afrika, Avrupa, Kuzey Amerika, Güney Amerika ve Okyanusya olmak üzere altı farklı kıtada ve 71 farklı ülkede coğrafi dağılım göstermektedir (Şekil 4.1.2). Veri setindeki en büyük yüzdeye sahip olan ülke %23 ile Amerika'dır.



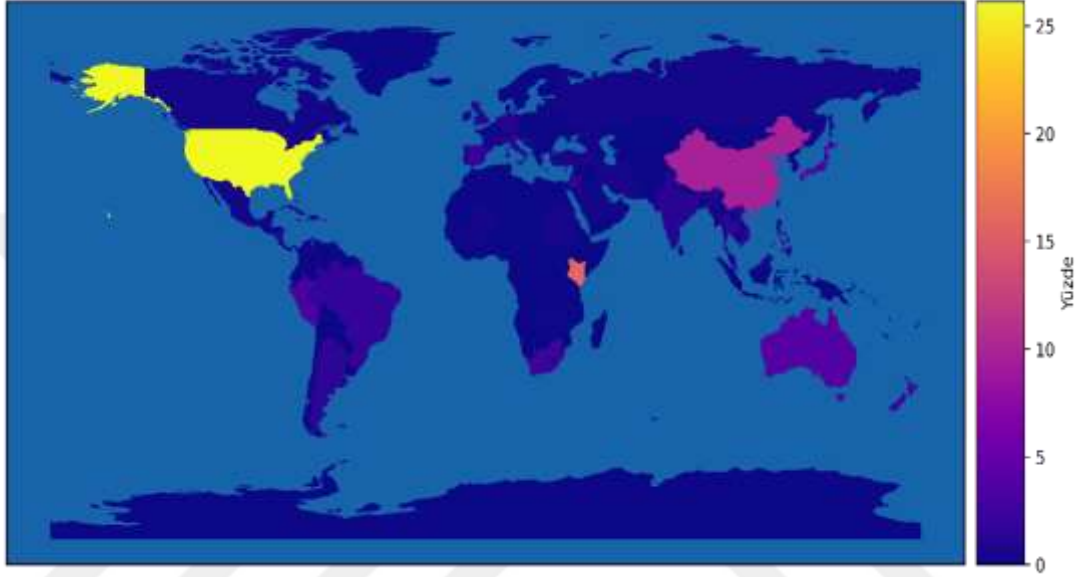
**Şekil 4.1.2 :** NCBI Virus veritabanında bulunan dizilerin coğrafi dağılımı.

VIPR veritabanından indirilen veri seti incelendiğinde protein dizilerinin dağılımının 1956 ve 2019 yılları arasında olduğu görülmektedir. En fazla protein dizisi eklenen yıl ise 2013'tür ve eklenen protein dizisi kaydı 7102'dir (Şekil 4.1.3). Bu sayının, NCBI Virus veritabanından indirilen veri setinde farklı olması veritabanlarında bulunan protein dizileri arasında farklılık olduğuna işaret etmektedir. Çalışmanın geniş kapsamlı olması açısından farklı veritabanlarının kullanılmasının önemini de ortaya koymaktadır.



**Şekil 4.1.3 :** VIPR veritabanında bulunan dizilerin zamana bağlı olarak dağılımı.

VIPR veritabanından indirilen veri setindeki protein dizileri NCBI Virus veritabanı ile aynı olarak Asya, Afrika, Avrupa, Kuzey Amerika, Güney Amerika ve Okyanusya olmak üzere altı farklı kıtada ve 71 farklı ülkede coğrafi dağılım göstermektedir (Şekil 4.1.4). Veri setindeki en büyük yüzdeye sahip olan ülke %26 ile Amerika'dır. Diğer ülkelerde NCBI Virus veritabanındaki protein dizileri ile benzer yüzdelere sahiptir.



Şekil 4.1.4 : VIPR veritabanında bulunan dizilerin coğrafi dağılımı.

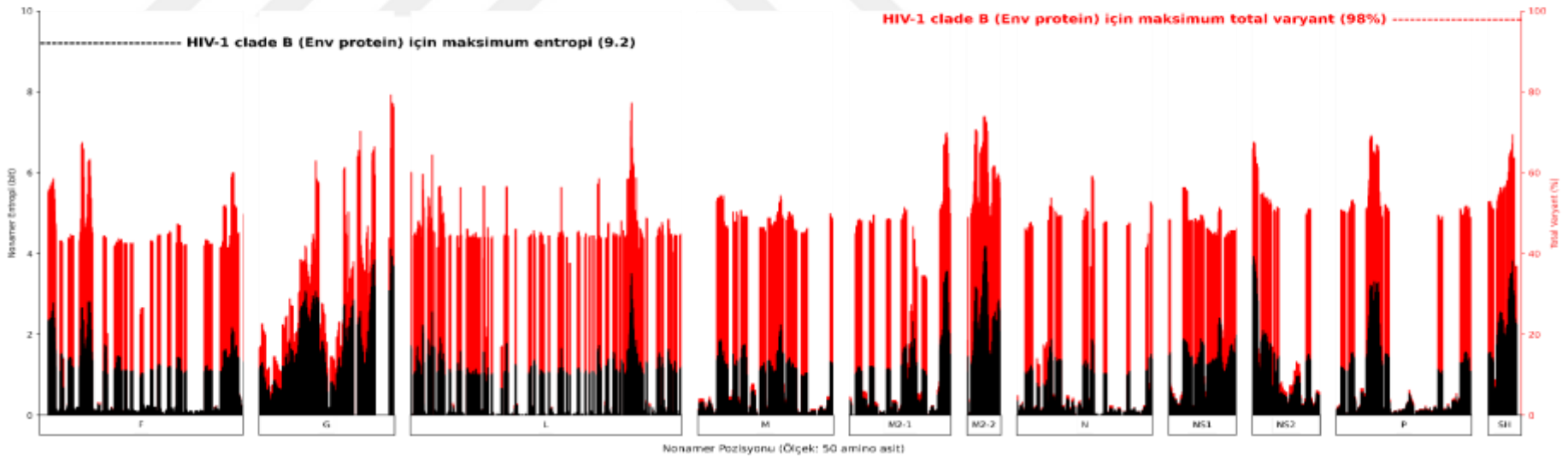
## 4.2 Proteom Dizi Çeşitliliği

hRSV proteome dizi çeşitliliği Shannon entropisi kullanılarak ölçülmüştür. Proteome düzeyinde entropi değeri ortalama 0,79 olarak bulunmuştur ve entropi değeri dağılımı 0 ile 4,19 arasında değişim göstermiştir. Buna bağlı olarak, proteome düzeyinde korunmuşluk yüzdesi yüksek olarak ifade edilebilir. Proteome düzeyindeki total varyantların ortalama değeri ise %22,66'dır ve varyantların değer dağılımı %0 ile %79,33 arasında bulunmuştur. Dört proteinin (L,M,N,P) ortalama entropi değeri olan 0,79'dan düşük değere sahip olduğu gözlemlenmiştir (Tablo 4.2.1).

**Tablo 4.2.1 :** Analiz edilen hRSV dizilerine ait entropi ve total varyant istatistikleri.

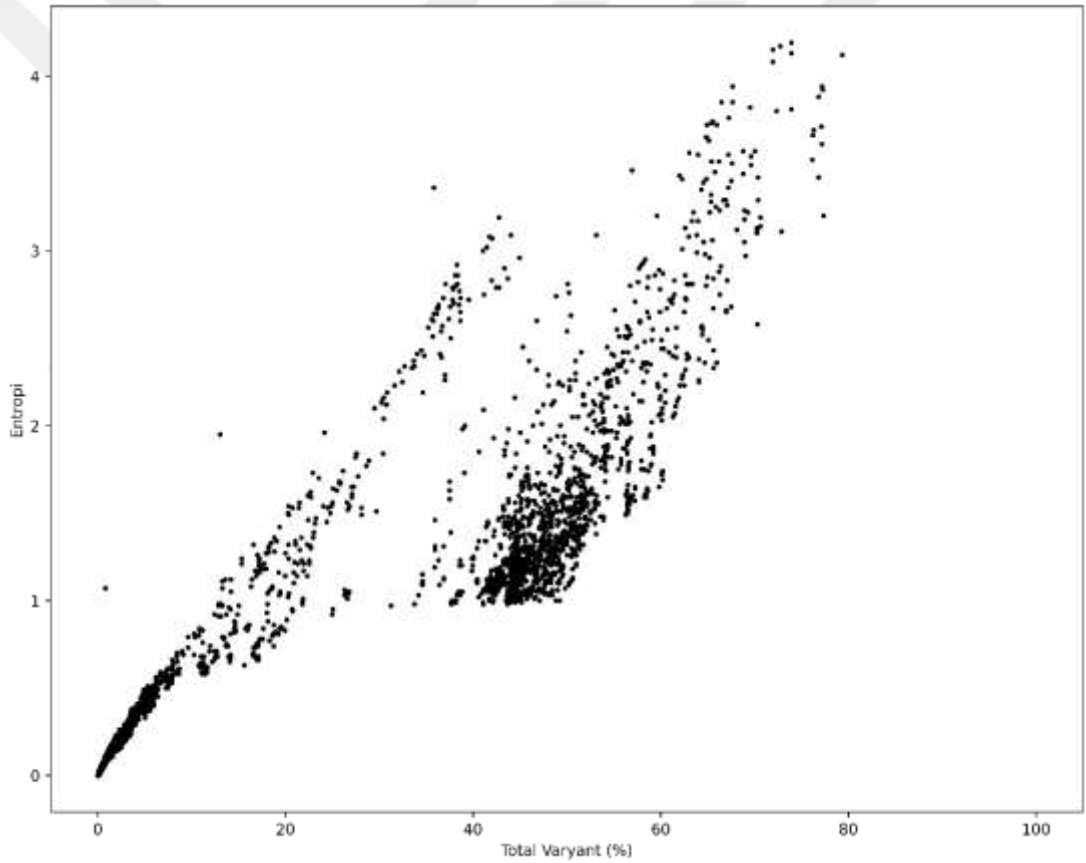
Protein	Analiz Edilen Pozisyon Sayısı	Entropi - H(x)			Total Varyant (%)		
		Minimum	Maksimum	Ortalama	Minimum	Maksimum	Ortalama
Proteom	4449	0,00	4,19	0,79	0,00	79,33	22,66
F	566	0,03	2,81	0,80	0,27	67,52	23,30
G	320	0,11	4,12	1,73	1,00	79,33	31,06
L	2126	0,00	3,52	0,56	0,00	77,34	19,39
M	248	0,00	2,25	0,76	0,00	54,33	24,65
M2-1	187	0,00	3,57	0,87	0,00	70,04	23,53
M2-2	82	0,00	4,19	2,28	0,00	73,91	54,86
N	383	0,00	1,88	0,50	0,00	59,22	15,60
NS1	131	0,15	2,42	1,22	1,81	56,41	35,36
NS2	116	0,15	3,94	1,13	0,83	67,66	25,65
P	233	0,03	3,32	0,77	0,39	69,29	21,75
SH	57	0,47	3,82	2,15	5,43	69,52	48,79

Entropi değeri 1'e eşit ve küçük olan tam 2480 pozisyon hRSV proteomunda bulunmuştur. 1612 pozisyonun entropi değeri 2'ye eşit ve küçük olmakla birlikte 1'den de büyük olarak gözlemlenmiştir. Bunun yanı sıra, entropi değeri 3'e eşit ve küçük olan 2'den de büyük olan 269 pozisyon vardır. 82 pozisyonun entropi değeri 3'den büyük, 4'e eşit ve küçüktür. Sadece altı pozisyonun entropi değeri 4'den büyük, 5'e eşit ve küçük olarak bulunmuştur. Toplamda analiz edilen 4449 pozisyonun %55,7'sinin (entropi değeri 1'e eşit ve küçük olan pozisyonlar) yüksek korunmuşluğa sahip olduğu saptanmıştır. Proteome düzeyinde maksimum entropi değeri olan 4,19 ise M2-2 proteininde bulunmaktadır ve 43-51 pozisyonları arasındaki nonamer dizisine aittir. Total varyantların maksimum değeri olan %79,33 ise G proteininde 312-320 pozisyonları arasındaki nonamerde bulunmuştur. Bu zamana kadar yapılan çalışmalarda en yüksek entropi değerine sahip (en fazla çeşitlilik gösteren) nonamer HIV-1 B kolunda Env proteinine aittir ve değeri 9,2'dir. Aynı şekilde, en yüksek total varyantlar da HIV-1 virüsün Env proteinine ait olup değeri %98'dir. Özellikle entropi değerinde yaklaşık 2 kattan fazla bir boşluk olduğu hRSV ve HIV-1 arasında görülmektedir (Şekil 4.2.1). Bu durum, hRSV proteomunun genel olarak korunmuş bölgelerden oluştuğunu göstermektedir ve uygun aşı adayları epitoplara keşfetmenin hRSV için mümkün olduğunu ortaya koymaktadır.



Şekil 4.2.1 : hRSV'nin her bir proteini için pozisyonlara ait entropi ve total varyant değerlerini içeren stem grafiği.

hRSV proteomunda entropi ve total varyantlar arasında ilişki incelendiğinde iki değer arasında güçlü pozitif korelasyon olduğu görülmektedir (Pearson korelasyon katsayısı  $r = 0,89$ ). Tüm nonamer pozisyonları ele alındığında total varyantlar için entropi değeri 1, toplam nonamerlerin yaklaşık %55'inde görülmüştür. Entropi değeri 2, 3, 4 ve 4,19 (maksimum entropi değeri) sırasıyla tüm nonamerlerin yaklaşık %36, %6, %1.8 ve %1'inde saptanmıştır. Grafik dikkatli incelendiğinde entropi değeri 1-3 aralığı ile total varyantların %20-%40 değer aralığında boşluklar gözlemlenmektedir (Şekil 4.2.2). Bu durum, diğer virüsler ile kıyaslama yapıldığında olağan dışı olarak karşımıza çıkmaktadır ve ilgili bölgenin daha detaylı araştırılmasının gerekliliğini ortaya koymaktadır.



Şekil 4.2.2 : hRSV proteomunun entropi ve total varyantları arasındaki ilişki.

### 4.3 Çeşitlilik Motiflerinin Dinamiği

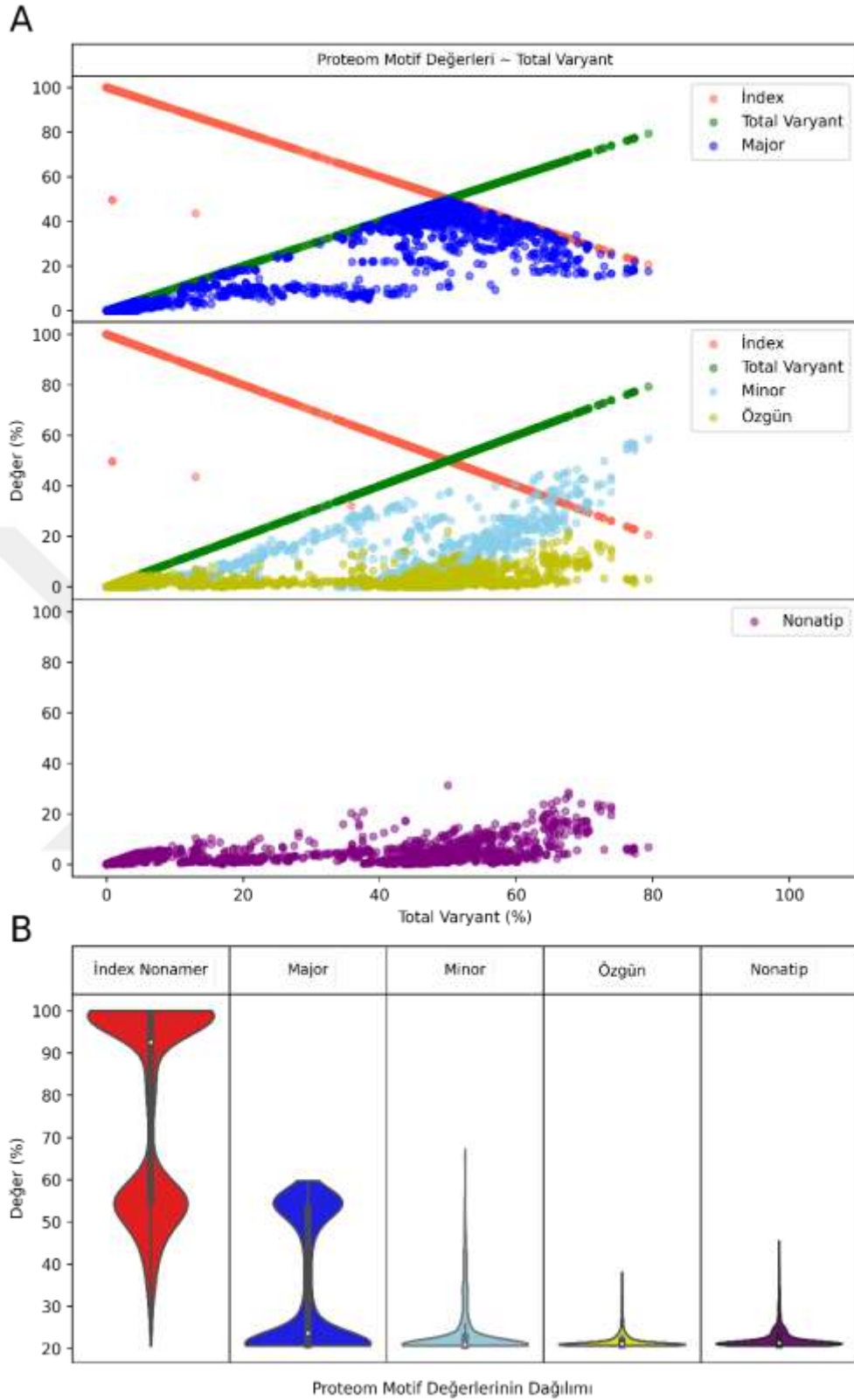
hRSV proteomunun motif çeşitliliği dinamiklerini daha iyi anlamak için 4449 üst üste gelen nonamer dizisinin nicel analizi yapılmıştır. M2-2 proteininin örnek olarak seçilmiş nonamerleri için motif çeşitliliği dinamikleri Tablo 4.3.1'de incelenebilir. Tablo 4.3.1'deki veri setinin M2-2 proteini için başlangıç pozisyonu çoklu dizi hizalaması dikkate alındığında 1-13 arasında olan toplam 13 farklı nonamerden oluştuğu gözlemlenebilir. 13 farklı nonamerin temsil ettiği toplam nonamer sayısı ise 2319'dur. Bu nonamerler arasında korunmuşluk derecesi tamamen korunmuş olan bir tek nonamer bulunmuştur. 7-15 pozisyonları arasında olan nonamerin entropi değeri 0 ve index değeri %100 olarak görülmektedir. Diğer nonamerler ise çok yüksek korunmuşluğa sahip ve karışık değişkenleri içeren korunmuşluk seviyelerine sahiptir. Çok yüksek korunmuşluğa sahip olan nonamerler arasında en yüksek index değeri %99,6'dır ve nonamer pozisyonları 8-16 ve 9-17'dir. Entropi değerleri de sırasıyla 0,04 ve 0,03'tür. Karışık değişkenleri içeren korunmuşluk seviyesine sahip olan nonamerler arasında en yüksek index değeri ise %81,4'tür. Bu nonamerlerin pozisyonları 1-9 ve 2-10'dur. Entropi değerleri sırasıyla 1,08 ve 1,07'dir. Her iki pozisyonun da toplam 105 nonamer dizisi içerdiği tespit edilmiştir. Total varyantların değeri bu pozisyonlar için %18,6, major varyantlar için %10,85, minor varyantlar ve özgün varyantlar için ise %3,88'dir. Nonatip olarak adlandırılan ve ilgili bölgedeki farklı nonamer sayısının o bölgedeki tüm nonamer sayısına bölümü sonucu hesaplanan değer ise %6,2'dir. Tablo 4.3.1 genelinde gösterilen nonamer pozisyonlarının hiçbirinde düşük destek (ilgili pozisyon için nonamer sayısının 30'dan az olması durumu) görülmemiştir.

**Tablo 4.3.1 : M2-2 proteini için kuantitatif dizi çeşitliliği analizine ait örnek tablo.**

Hızalanmış Nonamerler			Entropi H(x)	İndex			Varyant (%)				Nonatip (%) <sup>2</sup>	Düşük Destek <sup>3</sup>
Korunmuşluk Derecesi <sup>1</sup>	Nonamer Pozisyonu	Dizi Sayısı		Dizi	Değer (%)	Sayı	Total	Major	Minor	Özgün		
KDİ	1-9	129	1,08	MTKPKIMIL	81,40	105	18,60	10,85	3,88	3,88	6,20	-
KDİ	2-10	129	1,07	TKPKIMILP	81,40	105	18,60	10,85	3,88	3,88	6,20	-
KDİ	3-11	249	1,46	MPKIMILPD	51,00	127	49,00	42,57	4,42	2,01	4,02	-
YDKS	4-12	250	0,33	PKIMILPDK	96,00	240	4,00	2,00	0,80	1,20	2,00	-
YDKS	5-13	250	0,15	KIMILPDKY	98,40	246	1,60	0,80	0,00	0,80	1,20	-
YDKS	6-14	250	0,07	IMILPDKYP	99,20	248	0,80	0,00	0,00	0,80	0,80	-
TK	7-15	252	0,00	MILPDKYPC	100,00	252	0,00	0,00	0,00	0,00	0,00	-
YDKS	8-16	252	0,04	ILPDKYPCS	99,60	251	0,40	0,00	0,00	0,40	0,40	-
YDKS	9-17	252	0,03	LPDKYPCSI	99,60	251	0,40	0,00	0,00	0,40	0,40	-
KDİ	10-18	252	1,26	PDKYPCSIN	50,00	126	50,00	45,24	4,37	0,40	1,19	-
KDİ	11-19	252	1,33	DKYPCSINS	48,81	123	51,19	45,24	5,56	0,40	1,59	-
KDİ	12-20	252	1,33	KYPCSINSI	48,81	123	51,19	45,24	5,56	0,40	1,59	-
KDİ	13-21	252	1,39	YPCSINSIL	48,41	122	51,59	44,84	5,56	1,19	2,38	-

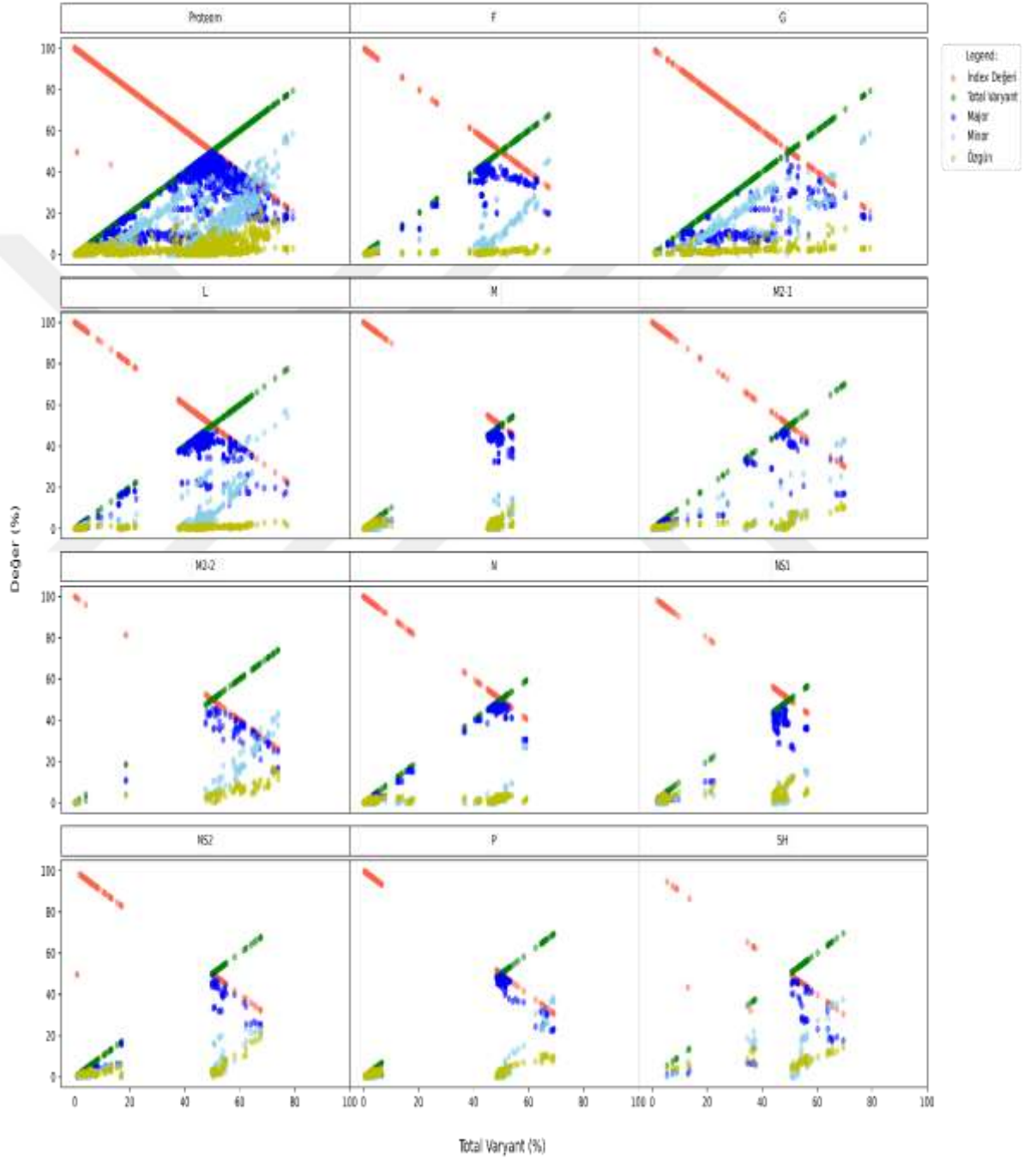
1. TK : Tamamen Korunmuş, YDKS : Yüksek Derecede Korunmuşluğa Sahip, KDİ : Karışık Değişkenleri İçeren
2. Nonatip : Varyantları oluşturan tüm belirgin nonamer dizileri
3. “-” pozisyonun 30 nonamer dizisinden fazla diziyeye sahip olduğunu belirtir. Eğer sayı 30’ dan küçük ise ilgili pozisyon düşük dizi desteğine sahiptir.

Her bir çeşitlilik motifinin total varyantlar ile olan ilişkisi Şekil 4.3.1'de proteom düzeyinde incelenmiştir. Total varyantların artış gösterdiği durumlarda tüm motif çeşitlerinin özgün bir desene sahip olduğu görülmektedir (Şekil 4.3.1 A.). Total varyantların artış gösterdiği durumda major varyantların değerinin piramit deseni oluşturduğu gözlemlenmiştir. Major varyantların değerleri 0 ile %49,21 arasında değişim göstermektedir. Grafik içerisinde görülen ve index değerinin genelinden farklı konumlarda bulunan noktalar ise farklı proteinlerin aynı pozisyonu içinde görülen birden fazla index nonamerini temsil etmektedir. SH proteininde 35-43 ve 50-58 pozisyonları arasındaki nonamerler ile NS2 proteininde 39-47 ve 40-48 pozisyonları arasındaki nonamerler için birden fazla index bulunmuştur. Minor ve özgün varyantlar ise proteom boyunca hemen hemen her pozisyonda görülebilmektedir. Minor varyantların maksimum değeri yaklaşık olarak %58'dir ve bu değere karşılık gelen total varyantların değerinin ise %60'a yakın olduğu görülebilir. Özgün varyantların ise daha dengeli bir dağılım gösterdiği tespit edilmiştir ve maksimum değeri yaklaşık %21 olarak bulunmuştur. Nonatipler de özgün varyantlar gibi proteom düzeyinde dengeli bir dağılım göstermiştir ve görülen maksimum değer yaklaşık %31 olarak tespit edilmiştir. Ayrıca korunmuşluk içeren nonamer dizi değerlerindeki azalmanın %18'e kadar düştüğü ve buna karşılık total varyantların değerinin %80'e kadar çıktığı görülmektedir. Bu iki değer arasında teorik olarak beklenildiği gibi ters orantı gözlemlenmiştir. Tüm çeşitlilik motiflerinin frekans dağılımları keman grafiği kullanılarak analiz edilmiştir (Şekil 4.3.1 B.). İndex nonamerler, major varyantlar, minor varyantlar, özgün varyantlar ve nonatiplerin proteome boyunca görülen ortalama değerleri sırasıyla %77,3, %17,6, %3,7, %1,2 ve %2 olarak ölçülmüştür. İndex nonamerlerin değerleri %50-60 ve %90-%100 arasında yoğunluk göstermektedir. Major varyantların yoğunluğu ise yaklaşık olarak %20-%30 ve %50-%60 değerleri arasında fazlalık göstermektedir. Diğer çeşitlilik motifleri için yoğunluklar yaklaşık olarak %20-%25 arasında değişiklik göstermektedir.



Şekil 4.3.1 : hRSV dizi çeşitliliği motiflerinin proteom düzeyinde dinamikleri.

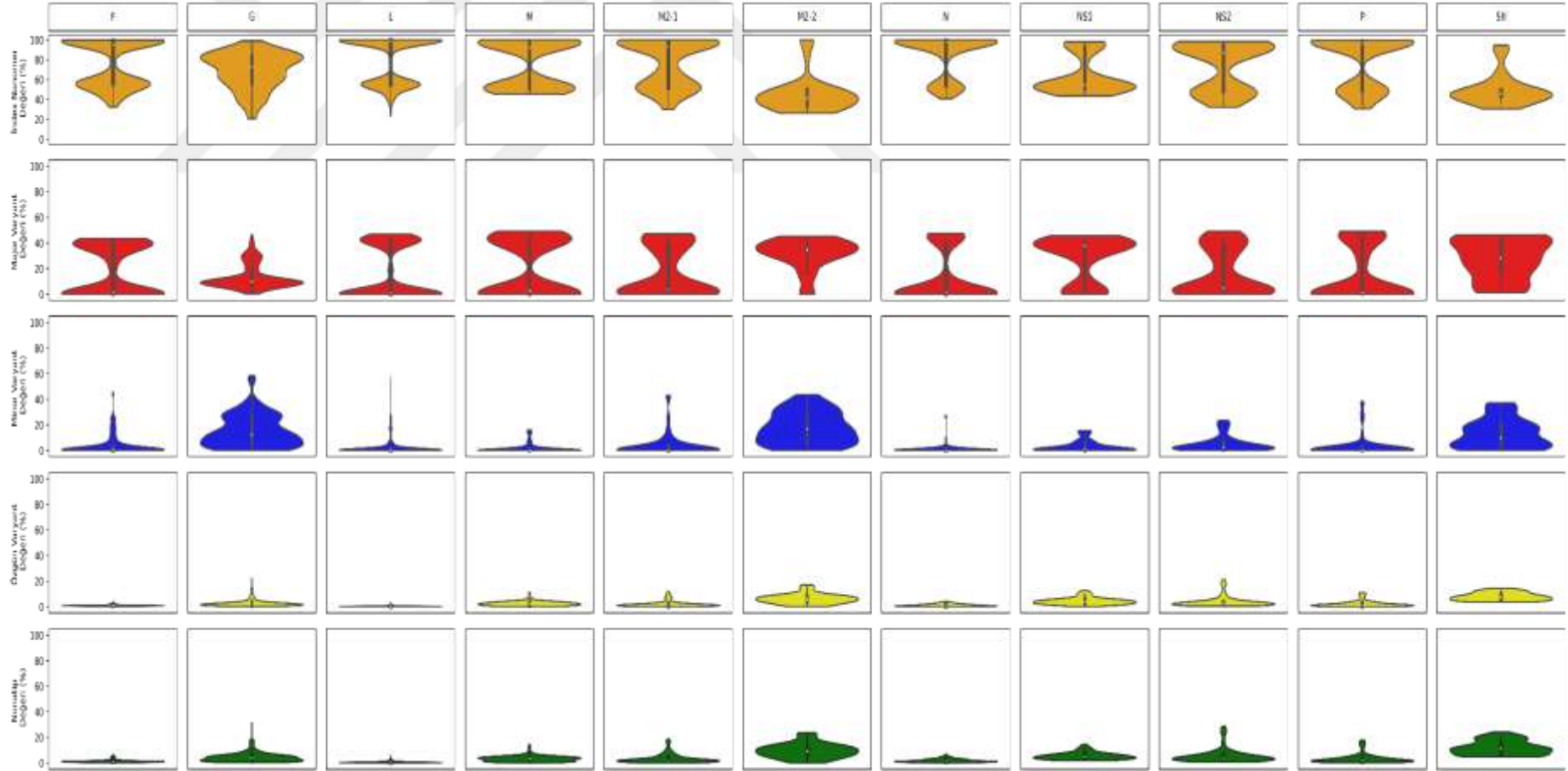
hRSV dizi çeşitliliği motiflerinin protein düzeyindeki dinamikleri, proteom düzeyindeki dinamikleri ile benzerlik göstermektedir (Şekil 4.3.2). G ve M2-1 proteinleri hariç diğer proteinlerin tüm motifleri %20-%40 değer aralığında veri noktalarına sahip değildir. Total varyantlar ise tüm proteinlerde %80 değerinin altında bulunmaktadır.



Şekil 4.3.2 : hRSV dizi çeşitliliği motiflerinin protein düzeyinde dinamikleri.

hRSV dizi çeşitliliği motiflerinin protein düzeyindeki sıklık dağılımları şekil 4.3.3'de keman grafikleri kullanılarak analiz edilmiştir. İndex motifleri incelendiğinde hiçbir protein için %20'inin altında değere rastlanmamıştır. G, M2-2 ve SH proteinleri dışındaki proteinlerin index değerleri %40-%60 ve %80-%100 aralıklarında yoğunluk göstermektedir. Major varyantlar ise tüm proteinlerde %50 değerinin altında veri noktalarına sahiptir. Bu durum minor varyantlar incelendiğinde yaklaşık %59 olarak tespit edilmiştir. Özgün varyantlar ve nonatipler için ise sırasıyla yaklaşık olarak %22 ve %32 değerlerinin üstünde veri noktaları görülmemiştir.

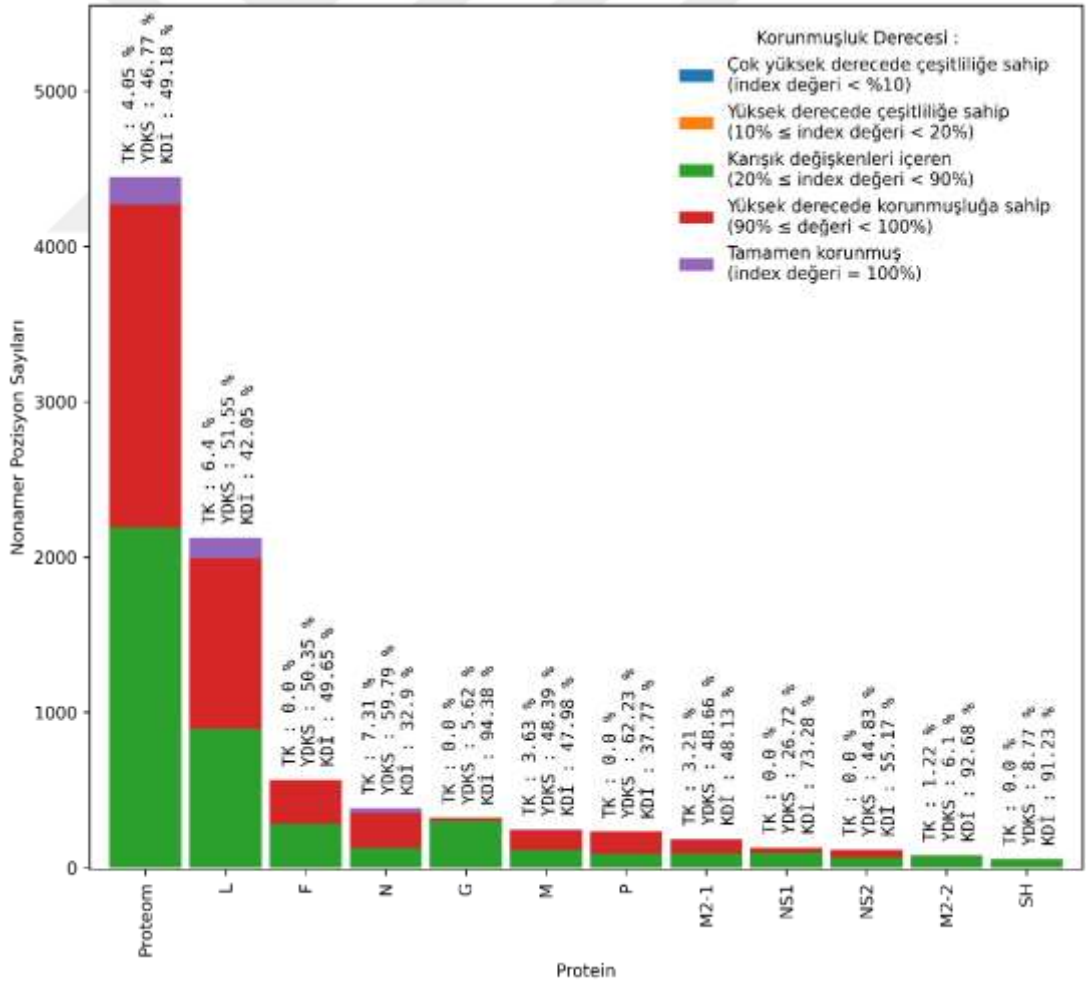




Şekil 4.3.3 : hRSV dizi çeşitliliği motiflerinin dağılımını inceleyen violin (keman) grafiği.

#### 4.4 Korunmuş ve Değişkenlik Gösteren Dizilerin Dağılımı

hRSV proteomunda hemen hemen tüm nonamer dizi pozisyonları karışık değişkenleri içeren ve yüksek derecede korunmuşluğa sahip olarak sınıflandırılmıştır (Şekil 4.4.1). Protome düzeyinde inceleme yapıldığında karışık değişkenleri içeren nonamer dizileri yaklaşık %49 (2180 nonamer dizisi), yüksek derecede korunmuşluk içeren nonamer dizileri yaklaşık %47 (2092 nonamer dizisi), tamamen korunmuş nonamer dizileri ise yaklaşık %4 (177 nonamer dizisi) olarak bulunmuştur. Sadece L, N, M, M2-1 ve M2-2 proteinleri tamamen korunmuş (index değeri %100'e eşit) olan nonamer dizilerini içermektedir. Çok yüksek derecede çeşitliliğe sahip (index değeri %10'dan küçük) ve yüksek derecede çeşitliliğe sahip (index değeri %20'den küçük, %10'dan büyük ve eşit) nonamer dizilerine ise hiçbir proteinde rastlanmamıştır.



Şekil 4.4.1 : Protein ve proteom düzeyinde hRSV nonamerlerinin korunmuşluk derecelerini gösteren grafik.

#### 4.5 Çok Yüksek Derecede Korunmuşluk İçeren ve Potansiyel Aşı Hedefi Olabilecek İmmünojenik Diziler

Tamamen korunmuş ve çok yüksek derecede korunmuşluk içeren MHC-I ve MHC-II'ye ait sırasıyla 997 ve 235 nonamer dizisi belirlenmiştir. MHC-I için belirlenen toplam 997 nonamer dizisinin 572'si L proteinine, 110'u F proteinine, 97'si N proteinine, 62'si M proteinine, 61'i P proteinine, 37'si M2-1 proteinine, 27'si NS2 proteinine, 16'sı NS1 proteinine, dokuzu G proteinine, dördü SH proteinine ve ikisi M2-2 proteinine aittir. MHC-II için belirlenen toplam 235 nonamer dizisinin 86'sı L proteinine, 52'si M proteinine, 48'i M2-1 proteinine, 30'u NS2 proteinine, 13'ü NS1 proteinine, dördü N proteinine ve ikisi G proteinine aittir. Belirlenen bu dizilerin hiçbiri daha önce deneysel olarak T hücrelerine bağlanma afinitesi göstermesi açısından ispatlanmamıştır. hRSV epitopları ile ilgili deneysel veriler çok kısıtlı olup IEDB veritabanına göre deneysel açıdan ispatlanmış sadece F proteinine ait üç epitop bulunmaktadır (Tablo 4.5.1). Ayrıca analizler sonucu elde edilen aday epitoplar ve onlara karşı gelen süpertip kısıtlayıcı alellere ait örnek veriler Tablo 4.5.2 ve Tablo 4.5.3'te incelenebilir.

**Tablo 4.5.1** : IEDB veritabanına göre deneysel olarak ispatlanmış hRSV epitopları.

Protein	Dizi	Epitop ID
F	ISNIETVIEFQQKN	1087566
F	RLLEITREFSVNAGVTTP	153694
F	GKSTTNIMIT	96322

**Tablo 4.5.2 : hRSV MHC-I epitopları ve onlara karşılık gelen süpertip kısıtlanmalı alellere ait örnek tablo.**

Protein	Peptit	Pozisyon	A1	A2	A3	A24	A26	B7	B8	B27	B39	B44	B58	B62
F	SSQNITEEF	24-32	SK	-	-	SK	SK	-	-	-	-	-	SK	SK
G	LAMIISTSL	46-54	-	-	-	-	-	SK	SK	-	-	-	-	-
L	SYLKGVISF	17-25	SK	-	-	SK	-	-	SK	SK	SK	-	-	SK
M	STYTAAVQY	12-20	SK	-	SK	-	SK	-	-	-	-	-	SK	SK
M2-1	KFEIRGHCL	8-16	-	-	-	-	-	-	SK	-	-	-	-	-
M2-2	KIMILPDKY	5-13	SK	-	-	-	-	-	-	-	-	-	SK	SK
N	ITEDANHKF	53-61	SK	-	-	SK	SK	-	-	-	-	-	SK	-
NS1	SLSMIKVRL	5-13	-	SK	-	-	-	-	SK	-	-	-	-	-
NS2	KLDERQATF	51-59	SK	SK	-	SK	-	-	SK	-	-	-	SK	SK
P	KKKDSIISV	35-43	-	-	-	-	-	-	-	SK	SK	-	-	-
SH	WPYFTLIHM	15-23	-	-	-	-	-	SK	-	-	-	-	-	-

Not. SK : Süpertip Kısıtlanmalı

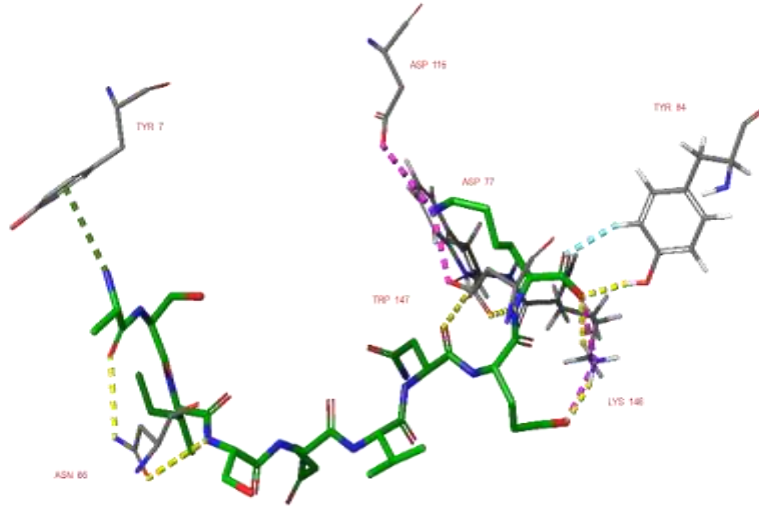
**Tablo 4.5.3 :** hRSV MHC-II epitopları ve onlara karşılık gelen süpertip kısıtlanmalı alellere ait örnek tablo.

Protein	Peptit	Pozisyon	A1	A2	A3
L	DATNNIHWSYLHIKF	1921-1935	SK	-	-
M	LHEGSTYTA AVQYNV	8-22	-	SK	-
M2-1	KQTIHLLKRLPADVL	143-157	SK	-	SK
NS1	NSLSMIKVRLQNLFD	4-18	SK	-	SK
NS2	VRKLDERQATFTFLV	49-63	SK	-	-

Not. SK : Süpertip Kısıtlanmalı

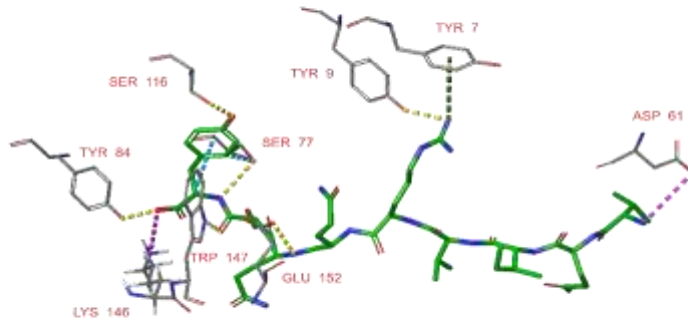
#### 4.6 Moleküler Docking Sonuçları

Epitop olma olasılığı yüksek olan peptitler ve bu olası epitopların bağlanması öngörülen MHC hedef proteinleri (EK B, EK C) arasında bağlanma etkileşimleri docking çalışmaları ile hesaplanmıştır. Yapılan analizler sonucu MHC-I sınıfı için teorik olarak bağlanma affinitesi gösteren 3 aday epitop seçilmiştir. Bunlardan ASISQVNEK ve VQIVRQQSY nonamerleri F proteinine ait olup sırasıyla HLA-A-11-01 ve HLA-B-15-03 alelleri ile bağ yapmışlardır [Şekil 4.6.1, Şekil 4.6.2]. Diğer nonamer olan KLNEIHLMK ise L proteinine ait olup HLA-A-03-01 aleli ile bağ yapmıştır [Şekil 4.6.3]. MHC-II sınıfına ait bağlanma affinitesi gösteren bir nonamer bulunamamıştır.



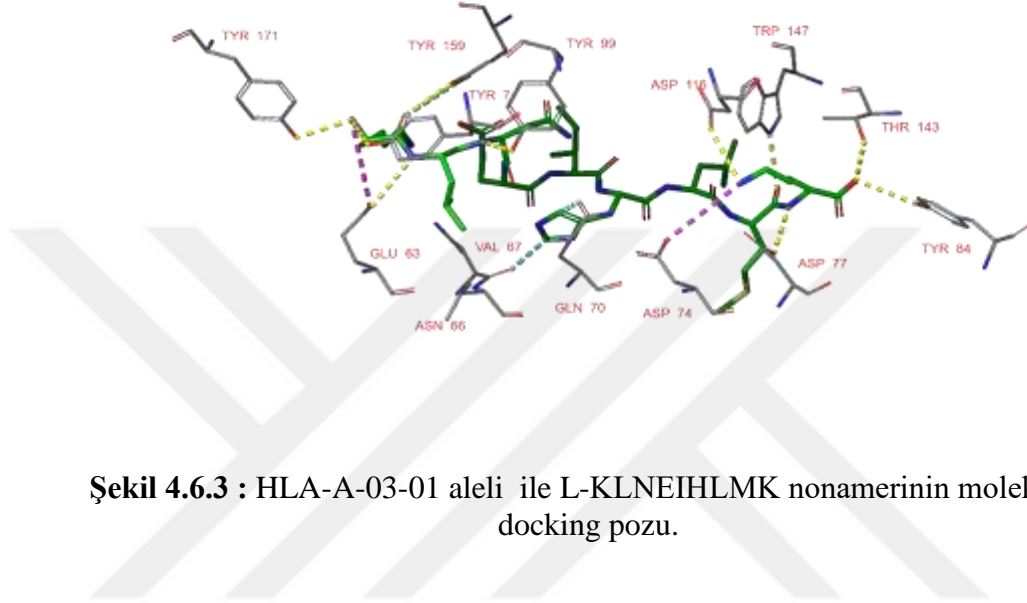
**Şekil 4.6.1** : HLA-A-11-01 aleli ile F-ASISQVNEK nonamerinin moleküler docking pozunu.

F-ASISQVNEK nonameri HLA-A-11-01 alelinin TYR7 ve TYR84 backbone kısımları ile pi-kasyon bağı oluşturduğu, ASN66, TRP147 ve ASP77 amino asitlerinin backbone kısmı ile de hidrojen bağı oluşturduğu görülmektedir.



**Şekil 4.6.2** : HLA-B-15-03 aleli ile F-VQIVRQQSY nonamerinin moleküler docking pozunu.

F-VQIVRQQSY nonameri HLA-B-15-03 alelinin TYR7 backbone kısmı ile pi-kasyon bağı oluşturduğu, GLU152 ve SER116 backbone kısımları ile de hidrojen bağı oluşturduğu görülmektedir.

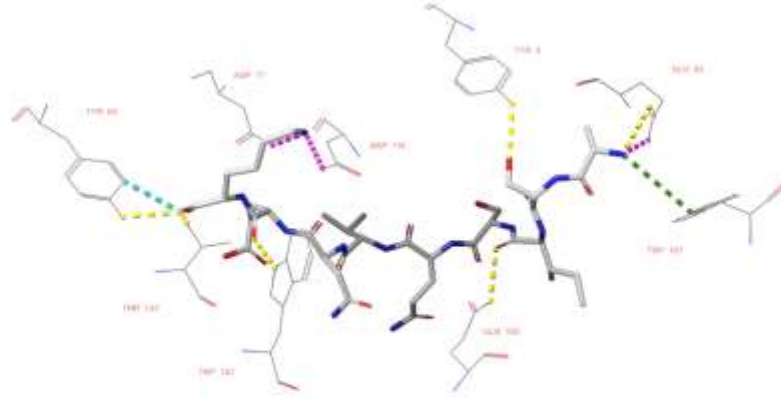
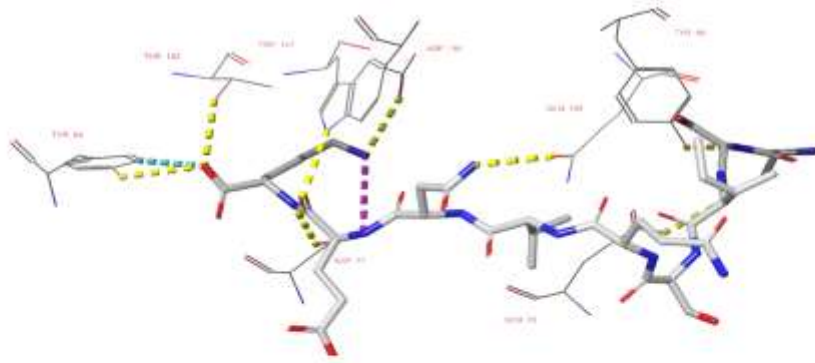
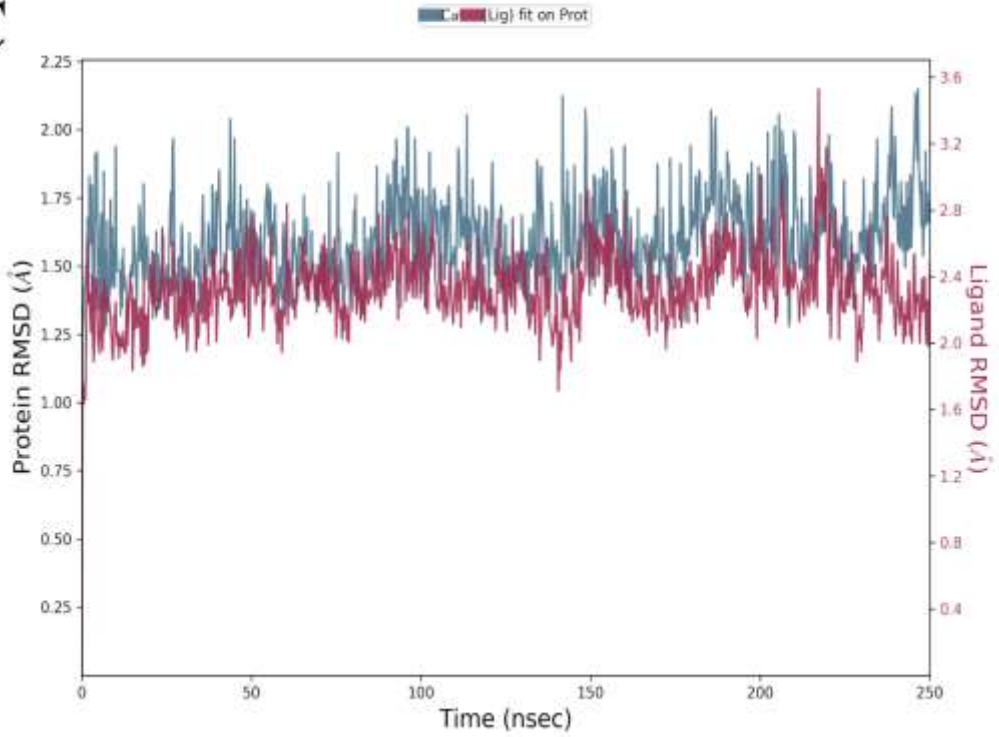


**Şekil 4.6.3 :** HLA-A-03-01 aleli ile L-KLNEIHLMK nonamerinin moleküler docking pozu.

L-KLNEIHLMK nonameri HLA-A-03-01 alelinin TYR171, GLU63, ASP116 ve TYR99 amino asitlerinin backbone kısımları ile hidrojen bağı oluşturdukları görülmektedir.

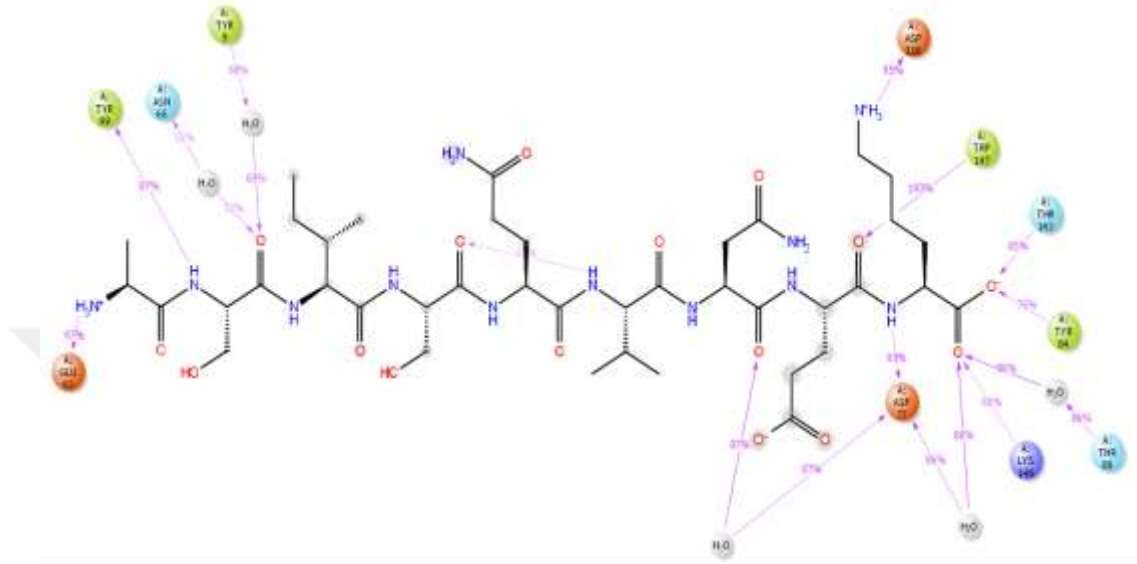
#### 4.7 Moleküler Dinamik Simülasyonlarının Sonuçları

Moleküler dinamik simülasyonlarının sonuçları F-ASISQVNEK nonamerinin F-VQIVRQQSY ve L-KLNEIHLMK nonamerlerinden daha stabil şekilde ilgili proteinine bağlandığını göstermiştir. F-ASISQVNEK nonameri için RMSD değeri 250 nanosaniye boyunca yaklaşık olarak 1 Å ile 3.5 Å arasında değişim göstermiştir. F-ASISQVNEK nonamerinin backbone ve yan zincirleri TYR84, TRP147, TYR9, GLN55, THR143 aminositleri ile simülasyonun başlangıç ve bitiş pozlarında bağlanma göstermiştir [Şekil 4.7.1].

**A****B****C**

**Şekil 4.7.1** : HLA-A-11-01 aleli ile F-ASISQVNEK nonamerinin moleküler dinamik simülasyonu sonuçları. (A) Simülasyon başlangıç pozunu, (B) simülasyon bitiş pozunu, (C) ligand ve proteinin simülasyon boyunca sahip olduğu RMSD değerleri.

F-ASISQVNEK nonamerinin simülasyonunun %97'si boyunca GLU63 ile N-terminal ucundan bağ yaptığı görülmektedir. Ayrıca TYR99 ve ASP77 amino asitleri, nonamerin backbone kısmı ile sırasıyla simülasyonunun %87'si ve %93'ü boyunca bağ yapmıştır [Şekil 4.7.2].

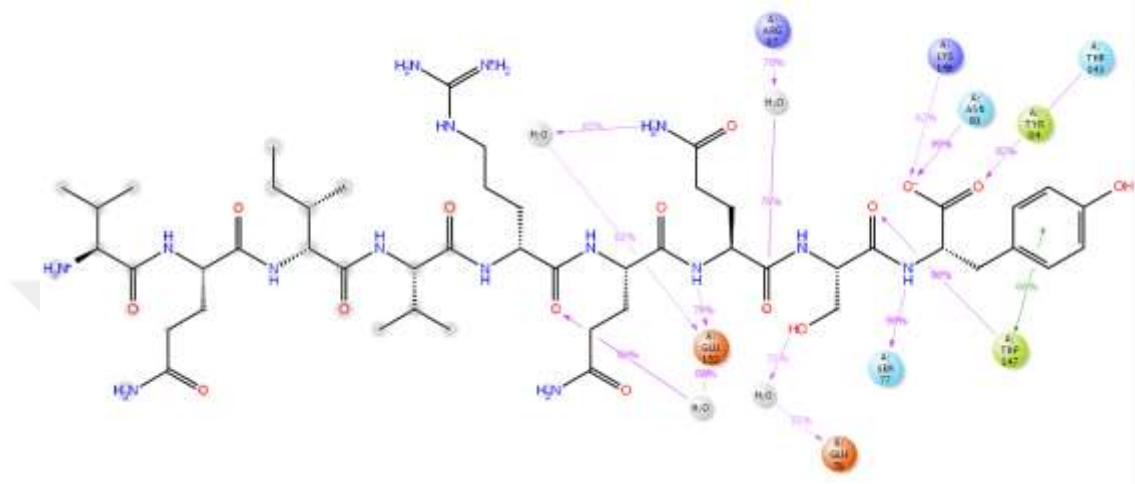


**Şekil 4.7.2 :** F-ASISQVNEK nonamerinin simülasyon boyunca bağ yaptığı amino asitler ve bağlandığı sürenin oranları.

F-VQIVRQQSY nonamerinin simülasyon boyunca gösterdiği RMSD değeri yaklaşık olarak 1 Å ile 13.5 Å arasında değişim göstermektedir ve bundan dolayı stabilitesi F-ASISQVNEK nonamerine göre daha düşüktür. F-VQIVRQQSY nonamerinin backbone ve yan zincirleri SER77, ASN80, TYR84, SER116, TRP147 aminositleri ile simülasyonun başlangıç ve bitiş pozlarında hidrojen bağı ile bağlanma göstermiştir [Şekil 4.7.3].

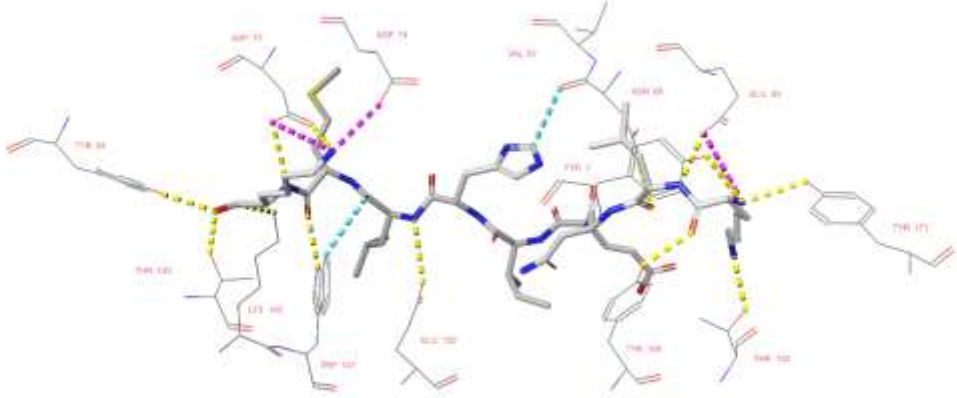
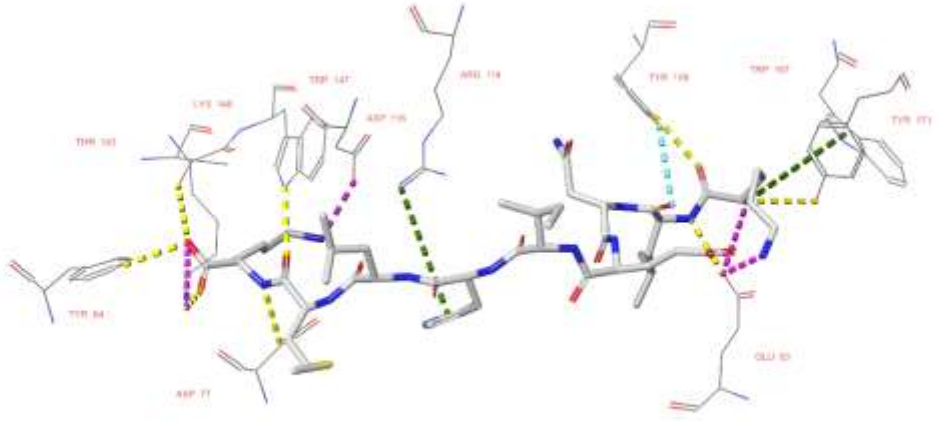
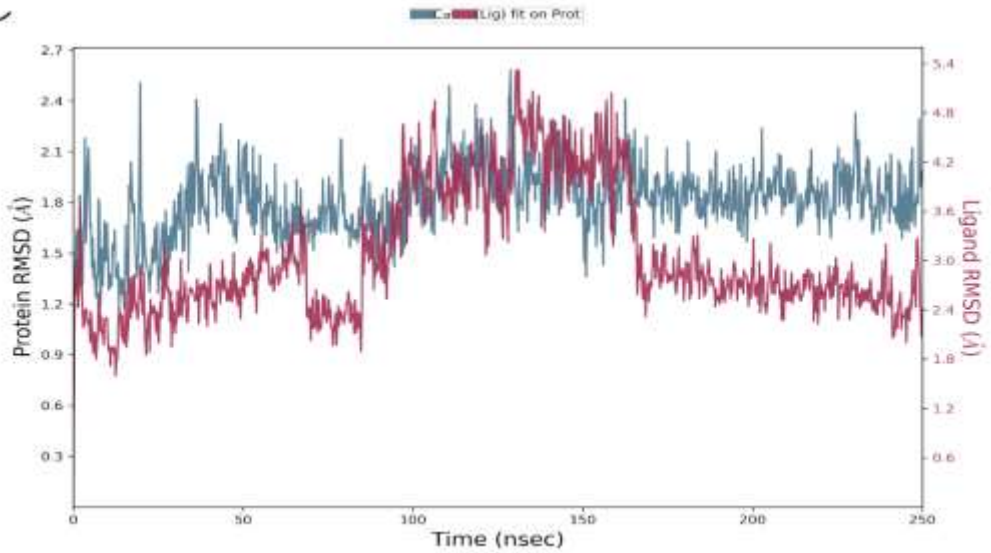


F-VQIVRQQSY nonamerinin simülasyonun %99'u boyunca ASN80 ile yan zincirinden bağ yaptığı görülmektedir. Ayrıca SER77 ve GLU152 amino asitleri, nonamerin backbone kısmı ile sırasıyla simülasyonun %98'i ve %79'u boyunca bağ yapmıştır [Şekil 4.7.4].



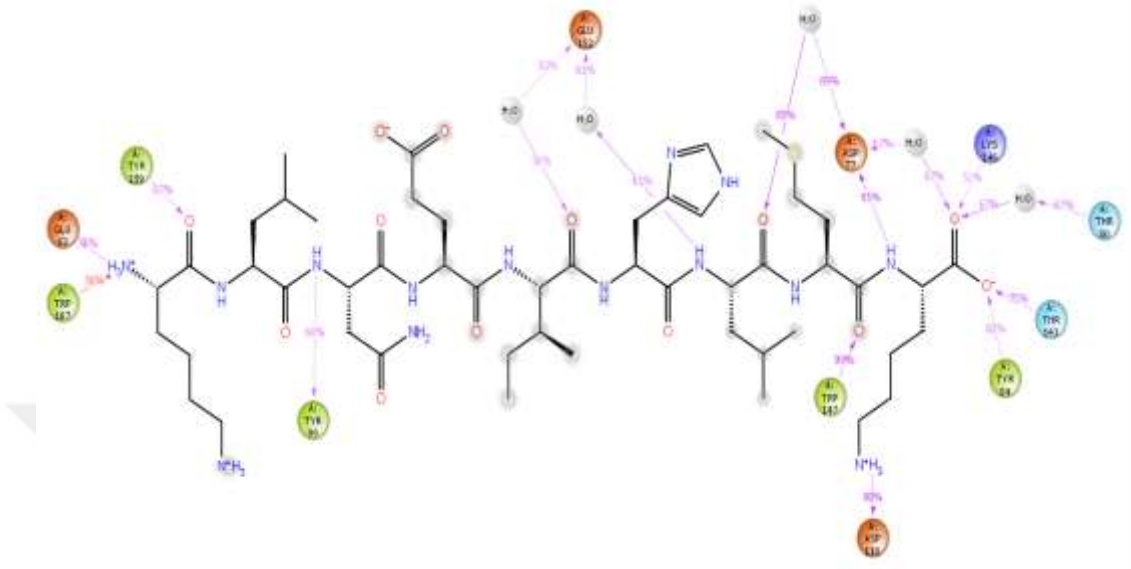
**Şekil 4.7.4 :** F-VQIVRQQSY nonamerinin simülasyon boyunca bağ yaptığı amino asitler ve bağlandığı sürenin oranları.

L-KLNEIHLMK nonamerinin simülasyon boyunca gösterdiği RMSD değeri yaklaşık olarak 0.8 Å ile 5.3 Å arasında değişim göstermektedir ve bundan dolayı stabilitesi F-ASISQVNEK nonamerine göre daha düşük, F-VQIVRQQSY nonamerine göre daha yüksektir. L-KLNEIHLMK nonamerinin backbone ve yan zincirleri TYR84, TYR159, TYR171 aminositleri ile simülasyonun başlangıç ve bitiş pozlarında hidrojen bağı ile bağlanma göstermiştir [Şekil 4.7.5].

**A****B****C**

**Şekil 4.7.5 :** HLA-A-03-01 aleli ile L-KLNEIHLMK nonamerinin moleküler dinamik simülasyonu sonuçları. (A) Simülasyon başlangıç pozu, (B) simülasyon bitiş pozu, (C) ligand ve proteinin simülasyon boyunca sahip olduğu RMSD değerleri.

L-KLNEIHLMK nonamerinin simülasyonun %99'u boyunca TRP147 ve %95'i boyunca ASP77 ile bağ oluşturduğu görülmektedir [Şekil 4.7.6].



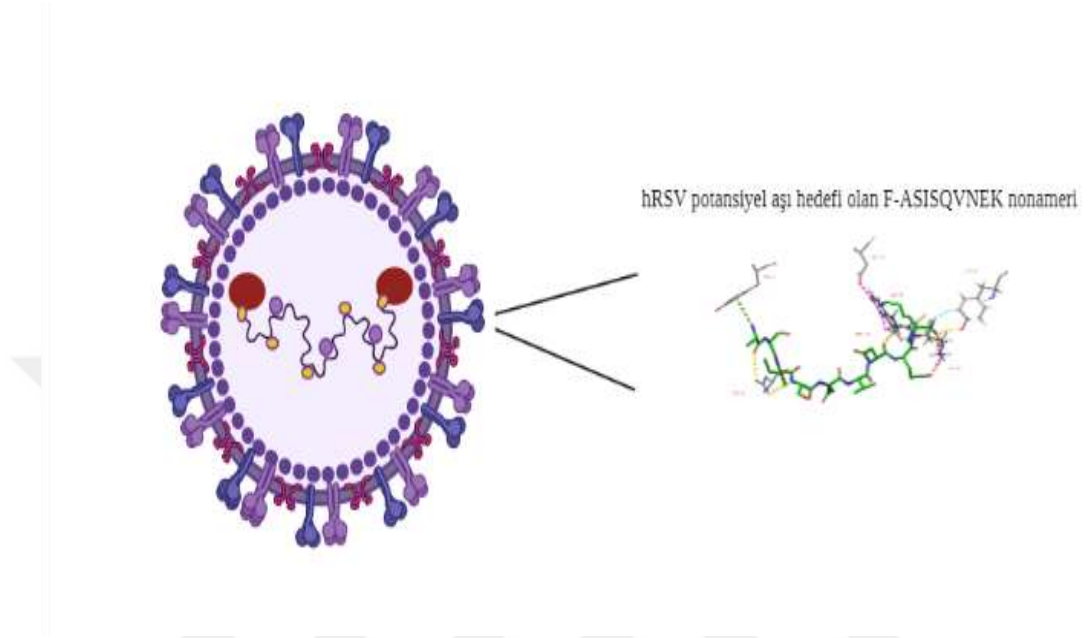
Şekil 4.7.6 : L-KLNEIHLMK nonamerinin simülasyon boyunca bağ yaptığı amino asitler ve bağlandığı sürenin oranları.

## 5. TARTIŞMA VE SONUÇ

hRSV'nin RNA virüsü olması ve yüksek mutasyon hızına sahip olması virüse karşı aşı geliştirme çalışmaları için zorluk çıkarmaktadır. Bu çalışma sonucunda elde edilen tüm korunmuş nonamer dizileri viral popülasyonun dinamiklerini çeşitli şekillerde etkilemekte ve virüse karşı aşı geliştirme çalışmaları için çok geniş bir repertuar oluşturmaktadır. Literatürü destekler bir şekilde F, G, L, N, P ve M proteinlerinin ortalama entropileri 1'in altında olup aşı geliştirme ve aday aşı hedefleri belirleme açısından üstüne çalışılması gereken proteinlerdir.

HIV clade B envelope proteinin literatürde bulunan en yüksek entropi değerine 9.2 ile sahip olması ve hRSV'nin bu değere olan uzaklığının fazla olması virüse karşı aşı geliştirme çalışmalarında olumlu sonuç elde edilmesinin mümkün olduğunu göstermektedir. Yapılan çalışmada proteom düzeyindeki tüm nonamerler incelendiğinde tamamen korunmuş ve yüksek derecede korunmuşluk içeren dizilerin oranın yaklaşık olarak %51 olması da bu durumu destekler niteliktedir. Ayrıca elde edilen nonamerlerin potansiyel epitop olma durumları da MHC-I ve MHC-II'ye karşı affinitelerini ölçen ve skorlayan tahmin programları ile belirlenmiştir. Elde edilen potansiyel epitoplara ait deneysel bir araştırma yapıp yapılmadığı IEDB veritabanı kullanılarak incelenmiştir. Sonuçların desteklenmesi açısından elde edilen tüm potansiyel epitoplar arasından en yüksek skora sahip ve her bir proteine ait olacak şekilde %1 oranında nonamer seçilmiş ve moleküler docking analizi yapılmıştır. F proteinine ait iki nonamer ve L proteinine ait bir nonamerin proteinlerine karşılık gelen MHC alelleri ile bağlanma gösterdiği tespit edilmiştir. Buna ek olarak 3 nonamer dizisi için moleküler dinamik simülasyonları 250 nanosaniyede yapılmış ve F proteinine ait ASISQVNEK nonamerinin simülasyon boyunca stabilite gösterdiği tespit edilmiştir. Moleküler docking ve moleküler dinamik simülasyonu analizleri yüksek hesaplama gücü ve fazla zaman gerektirdiğinden en yüksek skorlu nonamerler her bir protein için seçilerek analizler gerçekleştirilmiştir. Bu yüzden, sonuç olarak gösterilen F proteinine ait ASISQVNEK nonameri dışında incelenmeyi bekleyen toplamda 1231 potansiyel epitop bulunmaktadır.

Sonuç olarak bu tez çalışması F proteinine ait olan ASISQVNEK nonamerinin potansiyel aşı hedefi olduğunu göstermektedir (Şekil 5.1) ve deneysel olarak ispat edilmesi açısından ümit vadetmektedir.



Şekil 5.1 : Potansiyel aşı hedefi olan F-ASISQVNEK nonameri.

## KAYNAKLAR

- [1] **Nair H, Nokes DJ, Gessner BD, Dherani M, Madhi S a, Singleton RJ, et al.** (2010) Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. *The Lancet*. [Online] Elsevier Ltd; 375(9725): 1545–1555. Available from: 10.1016/S0140-6736(10)60206-1
- [2] **Oliveira DBL, Durigon EL, Carvalho ACL, Leal AL, Souza TS, Thomazelli LM, Moraes CTP, Vieira SE, Al G, Stewien KE** (2009) Epidemiology and genetic variability of human metapneumovirus during a 4-year-long study in Southeastern Brazil. *J Med Virol* 81:915–921
- [3] **Durigon, E.L., V.F. Botosso, and D.B.L. de Oliveira** (2017) Human Respiratory Syncytial Virus: Biology, Epidemiology, and Control, in *Human Virology in Latin America*. p. 235-254.
- [4] **Robinson JL, Lee BE, Bastien N, Li Y** (2005) Seasonality and clinical features of human metapneumovirus infection in children in northern Alberta. *J Med Virol* 76:98–105
- [5] **van den Hoogen BG, van Doornum GJ, Fockens JC, Cornelissen JJ, Beyer WE, de Groot R, Osterhaus AD, Fouchier RA** (2003) Prevalence e clinical symptoms of human metapneumovirus infection in hospitalized patients. *J Infect Dis* 188:1571–1577
- [6] **Morris JA Jr, Blount RE, Savage RE** (1956) Recovery of cytopathogenic agent from chimpanzees with coryza. *Proc Soc Exp Med* 92:544–550
- [7] **Chanock RM, Roizman B, Myers R** (1957) Recovery from infants with respiratory illness of virus related to chimpanzee coryza agent. I. isolation, properties and characterization. *Am J Hyg* 66:281–290
- [8] **Chanock RM, Finberg L** (1957) Recovery from infants with respiratory illness of a virus related to chimpanzee coryza agent (CCA). II. Epidemiological aspects of infection in infants and young children. *Am J Hyg* 66:291–300
- [9] **Collins PL, Dickens LE, Buckler-White A, et al.** (1986) Nucleotide sequences for the gene junctions of human respiratory syncytial virus reveal distinctive features of intergenic structure and gene order. *Proceedings of the National Academy of Sciences of the United States of America* 83(13):4594-4598.

- [10] **Collins PL, Wertz GW.** (1983) cDNA cloning and transcriptional mapping of nine polyadenylated RNAs encoded by the genome of human respiratory syncytial virus. *Proceedings of the National Academy of Sciences of the United States of America* 80(11):3208-3212.
- [11] **Collins PL, Graham BS.** (2008) Viral and host factors in human respiratory syncytial virus pathogenesis. *J Virol.* 82(5):2040-2055.
- [12] **Collins PL, Kerron RA.** (2013) Respiratory Syncytial Virus and Metapneumovirus. In: Knipe DM, Howley PM, eds. *Fields Virology*. 6th ed. Philadelphia: Lippincott, Williams and Wilkins; 1087-1123.
- [13] **Graham BS, Modjarrad K, McLellan JS** (2015) Novel antigens for RSV vaccines. *Curr Opin Immunol* 35:30–38
- [14] **Pablo F. Céspedes, Claudia A. Rivera, Rodrigo A. Díaz, And Alexis M. Kalergis,** (2018) *New Insights for the Rational Design Of Human Respiratory Syncytial Virus Vaccines: From Molecular Biology to Clinical Trials.*
- [15] **Atreya PL, Peeples ME, Collins PL.** (1998) The NS1 protein of human respiratory syncytial virus is a potent inhibitor of minigenome transcription and RNA replication. *Journal of Virology* 72(2):1452-1461.
- [16] **Spann KM, Tran KC, Chi B, Rabin RL, Collins PL.** (2004) Suppression of the induction of alpha, beta, and lambda interferons by the NS1 and NS2 proteins of human respiratory syncytial virus in human epithelial cells and macrophages [corrected]. *Journal of Virology* 78(8):4363-4369.
- [17] **Teng MN, Whitehead SS, Bermingham A, et al.** (2000) Recombinant respiratory syncytial virus that does not express the NS1 or M2-2 protein is highly attenuated and immunogenic in chimpanzees. *Journal of Virology* 74(19):9317-9321.
- [18] **Kotelkin A, Belyakov IM, Yang L, Berzofsky JA, Collins PL, Bukreyev A.** (2006) The NS2 protein of human respiratory syncytial virus suppresses the cytotoxic T-cell response as a consequence of suppressing the type I interferon response. *Journal of Virology* 80(12):5958-5967.
- [19] **Wright PF, Karron RA, Madhi SA, et al.** (2006) The interferon antagonist NS2 protein of respiratory syncytial virus is an important virulence determinant for humans. *The Journal of Infectious Diseases* 193(4):573-581.
- [20] **Céspedes PF, Bueno SM, Ramirez BA, et al.** (2014) Surface expression of the hRSV nucleoprotein impairs immunological synapse formation with T cells. *Proceedings of the National Academy of Sciences of the United States of America* 111(31):E3214-3223.

- [21] **Fearns R, Peeples ME, Collins PL.** (1997) Increased expression of the N protein of respiratory syncytial virus stimulates minigenome replication but does not alter the balance between the synthesis of mRNA and antigenome. *Virology* 236(1):188-201.
- [22] **Meng J, Lee S, Hotard AL, Moore ML.** (2014) Refining the balance of attenuation and immunogenicity of respiratory syncytial virus by targeted codon deoptimization of virulence genes. *mBio* 5(5):e01704-01714.
- [23] **Jin H, Cheng X, Traina-Dorge VL, et al.** (2003) Evaluation of recombinant respiratory syncytial virus gene deletion mutants in African green monkeys for their potential as live attenuated vaccine candidates. *Vaccine* 21(25-26):3647-3652.
- [24] **Ghildyal R, Mills J, Murray M, Vardaxis N, Meanger J.** (2002) Respiratory syncytial virus matrix protein associates with nucleocapsids in infected cells. *The Journal of General Virology* 83(Pt 4):753-757.
- [25] **Gan SW, Tan E, Lin X, et al.** (2012) The small hydrophobic protein of the human respiratory syncytial virus forms pentameric ion channels. *The Journal of Biological Chemistry* 287(29):24671-24689.
- [26] **Batonick M, Wertz GW.** (2011) Requirements for Human Respiratory Syncytial Virus Glycoproteins in Assembly and Egress from Infected Cells. *Adv Virol* 2011;2011.
- [27] **Feldman SA, Hendry RM, Beeler JA.** (1999) Identification of a linear heparin binding domain for human respiratory syncytial virus attachment glycoprotein G. *Journal of Virology* 73(8):6610-6617.
- [28] **Feldman SA, Audet S, Beeler JA.** (2000) The fusion glycoprotein of human respiratory syncytial virus facilitates virus attachment and infectivity via an interaction with cellular heparan sulfate. *Journal of Virology* 74(14):6442-6447.
- [29] **Asenjo A, Calvo E, Villanueva N.** (2006) Phosphorylation of human respiratory syncytial virus P protein at threonine 108 controls its interaction with the M2-1 protein in the viral RNA polymerase complex. *The Journal of General Virology* 87(Pt 12):3637-3642.
- [30] **Asenjo A, Gonzalez-Armas JC, Villanueva N.** (2008) Phosphorylation of human respiratory syncytial virus P protein at serine 54 regulates viral uncoating. *Virology* 380(1):26-33.
- [31] **Collins PL, Hill MG, Cristina J, Grosfeld H.** (1996) Transcription elongation factor of respiratory syncytial virus, a nonsegmented negative-strand RNA virus. *Proceedings of the National Academy of Sciences of the United States of America* 93(1):81-85

- [32] **Grosfeld H, Hill MG, Collins PL.** (1995) RNA Replication by respiratory syncytial virus (RSV) is directed by the N, P, and L proteins; transcription also occurs under these conditions but requires RSV superinfection for efficient synthesis of full-length mRNA. *J Virol.* 69(9):5677-5686.
- [33] **Mufson MA, Orvell C, Rafnar B, Norrby E.** (1985) Two distinct subtypes of human respiratory syncytial virus. *J Gen Virol.* 66(10): 2111-2124.
- [34] **Peret TC, Hall CB, Schnabel KC, Golub JA, Anderson LJ.** (1998) Circulation patterns of genetically distinct group A and B strains of human respiratory syncytial virus in a community. *J Gen Virol.* 79(9): 2221-2229.
- [35] **Melero JA, Garcia-Barreno B, Martinez I, Pringle CR, Cane PA.** (1997) Antigenic structure, evolution and immunobiology of human respiratory syncytial virus attachment (G) protein. *J Gen Virol.* 78(10):2411-2418.
- [36] **Thomson E. et al.** (2016) ‘Comparison of Next-Generation Sequencing Technologies for Comprehensive Assessment of Full-Length Hepatitis C Viral Genomes’, *Journal of Clinical Microbiology*, 54: 2470–84.
- [37] **Gimenez H. B. et al.** (1986) ‘Antigenic Variation between Human Respiratory Syncytial Virus Isolates’, *Journal of General Virology* , 67: 863–70.
- [38] **Agoti C. N. et al.** (2015) ‘Successive Respiratory Syncytial Virus Epidemics in Local Populations Arise from Multiple Variant Introductions, Providing Insights into Virus Persistence’, *Journal of Virology*, 89: 11630–42.
- [39] **Guzman E, Taylor G** (2015) Immunology of bovine respiratory syncytial virus in calves. *Mol Immunol* 66(1):48–56
- [40] **Openshaw PJ** (2013) The mouse model of respiratory syncytial virus disease. *Curr Top Microbiol Immunol* 372:359–369
- [41] **Ruckwardt TJ, Morabito KM, Graham BS** (2016) Determinants of early life immune responses to RSV infection. *Curr Opin Virol* 16: 151–157
- [42] **Sara A. Taleb, A.A.A.T., Khalid Al Ansari, Hadi M. Yassine,** (2018) Human respiratory syncytial virus: pathogenesis, immune responses, and current vaccine approaches. *European Journal of Clinical Microbiology & Infectious Diseases.*
- [43] **Domingo E, Holland J J.** (1994) Mutation rates and rapid evolution of RNA viruses. In: Morse S S, editor. *The evolutionary biology of viruses.* New York, N.Y: Raven Press; pp. 161–184.
- [44] **Mazur, N. I. et al.** (2018) The respiratory syncytial virus vaccine landscape: lessons from the graveyard and promising candidates. *Lancet Infect. Dis.* 18, e295–e311.

- [45] **Graham, B. S.** (2017) Vaccine development for respiratory syncytial virus. *Curr. Opin. Virol.* 23, 107–112.
- [46] **Buchholz UJ et al** (2018) Live respiratory syncytial virus (RSV) vaccine candidate containing stabilized temperature-sensitivity mutations is highly attenuated in RSV-seronegative infants and children. *J Infect Dis* 217(9):1338–1346
- [47] **Jorquera PA, Anderson L, Tripp RA.** (2016) Understanding respiratory syncytial virus (RSV) vaccine development and aspects of disease pathogenesis. *Expert review of vaccines.* [Online] 15(2): 173–187.
- [48] **Mufson, M.A.; Orvell, C.; Rafnar, B.; Norrby, E.** (1985) Two Distinct Subtypes of Human Respiratory Syncytial Virus. *J. Virol.* 66, 2111–2124.
- [49] **Zlateva, K.T.; Lemey, P.; Vandamme, A.-M.; Van Ranst, M.** (2004) Molecular Evolution and Circulation Patterns of Human Respiratory Syncytial Virus Subgroup A: Positively Selected Sites in the Attachment G Glycoprotein. *J. Virol.* 78, 4675–4683.
- [50] **Zlateva, K.T.; Lemey, P.; Moës, E.; Vandamme, A.-M.; Van Ranst, M.** (2005) Genetic Variability and Molecular Evolution of the Human Respiratory Syncytial Virus Subgroup B Attachment G Protein. *J. Virol.* 79, 9157–9167.
- [51] **Sande, C.J.; Mutunga, M.N.; Medley, G.F.; Cane, P.A.; Nokes, D.J.** (2013) Group- and genotype-specific neutralizing antibody responses against respiratory syncytial virus in infants and young children with severe pneumonia. *J. Infect. Dis.* 207, 489–492.
- [52] **Trento, A.; Abrego, L.; Rodríguez-Fernández, R.; González-Sánchez, M.I.; González-Martínez, F.; Delfraro, A.; Pascale, J.M.; Arbiza, J.; Melero, J.A.** (2015) Conservation of G-Protein Epitopes in Respiratory Syncytial Virus (Group A) Despite Broad Genetic Diversity: Is Antibody Selection Involved in Virus Evolution? *J. Virol.* 89, 7776–7785.
- [53] **Peret, T.C.; Golub, J.A.; Anderson, L.J.; Hall, C.B.; Schnabel, K.C.** (1998) Circulation patterns of genetically distinct group A and B strains of human respiratory syncytial virus in a community. *J. Virol.* 79, 2221–2229.
- [54] **Agoti, C.N.; Otieno, J.R.; Gitahi, C.W.; Cane, P.A.; Nokes, D.J.** (2014) Rapid Spread and Diversification of Respiratory Syncytial Virus Genotype ON1, Kenya. *Emerg. Infect. Dis.* 20, 950–959.
- [55] **Eshaghi, A.; Duvvuri, V.R.; Lai, R.; Nadarajah, J.T.; Li, A.; Patel, S.N.; Low, D.E.; Gubbay, J.B.** (2012) Genetic Variability of Human Respiratory Syncytial Virus A Strains Circulating in Ontario: A Novel Genotype with a 72 Nucleotide G Gene Duplication. *PLoS ONE* 7, e32807.

- [56] **Trento, A.; Galiano, M.; Videla, C.; Carballal, G.; García-Barreno, B.; Melero, J.A.; Palomo, C.** (2003) Major changes in the G protein of human respiratory syncytial virus isolates introduced by a duplication of 60 nucleotides. *J. Virol.* 84, 3115–3120.
- [57] **Hotard, A.L.; Laikhter, E.; Brooks, K.; Hartert, T.V.; Moore, M.L.; Lopez-Ona, A.** (2015) Functional Analysis of the 60-Nucleotide Duplication in the Respiratory Syncytial Virus Buenos Aires Strain Attachment Glycoprotein. *J. Virol.* 89, 8258–8266.
- [58] **Botosso, V.F.; Zanotto, P.M.D.A.; Ueda, M.; Arruda, E.; Gilio, A.E.; Vieira, S.E.; Stewien, K.E.; Peret, T.C.T.; Jamal, L.F.; Pardini, M.I.D.M.C.; et al.** (2009) Positive Selection Results in Frequent Reversible Amino Acid Replacements in the G Protein Gene of Human Respiratory Syncytial Virus. *PLOS Pathog.* 5, e1000254.
- [59] **Raman, H., Tan, S., Sjaugi, M., Mohamed, N. N., & Khan, A.** (2014, September). Bioinformatics for vaccine target discovery. *Asia Pacific Biotech News – Bioinformatics*, 18(9), 43–49.
- [60] **Hatcher EL, Zhdanov SA, Bao Y, Blinkova O, Nawrocki EP, Ostapchuck Y, Schäffer AA, Brister JR.** (2017) Virus Variation Resource - improved response to emergent viral outbreaks. *Nucleic Acids Res.* Jan 4;45(D1):D482-D490. doi: 10.1093/nar/gkw1065. Epub 2016 Nov 28. PMID: 27899678; PMCID: PMC5210549.
- [61] **Pickett, B. E., Sadat, E. L., Zhang, Y., Noronha, J. M., Squires, R. B., Hunt, V., Liu, M., Kumar, S., Zaremba, S., Gu, Z., Zhou, L., Larson, C. N., Dietrich, J., Klem, E. B., & Scheuermann, R. H.** (2012). ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic acids research*, 40, D593–D598. <https://doi.org/10.1093/nar/gkr859>
- [62] **Limin Fu, Beifang Niu, Zhengwei Zhu, Sitao Wu and Weizhong Li,** (2012) CD-HIT: accelerated for clustering the next generation sequencing data. *Bioinformatics*, 28 (23): 3150-3152. doi: 10.1093/bioinformatics/bts565
- [63] **National Center for Biotechnology Information (NCBI)** (1988). Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information
- [64] **The UniProt Consortium, UniProt:** the universal protein knowledgebase in 2021, *Nucleic Acids Research*, Volume 49, Issue D1, 8 January 2021, Pages D480–D489, <https://doi.org/10.1093/nar/gkaa1100>
- [65] **Robert C. Edgar,** (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Research*, Volume 32, Issue 5, Pages 1792–1797, <https://doi.org/10.1093/nar/gkh340>
- [66] **Larsson, A.** (2014). AliView: a fast and lightweight alignment viewer and editor for large data sets. *Bioinformatics*30(22): 3276-3278.

- [67] **C. E. Shannon** (1948) "A mathematical theory of communication," in *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379-423, doi: 10.1002/j.1538-7305.1948.tb01338.x.
- [68] **Khan, A. M., Olivo, M., EJ, N., KN, S., AT, H., GL, Z., ... JT, A.** (2008). Conservation and Variability of Dengue Virus Proteins: Implications for Vaccine Design. *PLoS Negl Trop Dis*, 2(8), e272. <http://doi.org/10.1371/journal.pntd.0000272>
- [69] **Tharanga, Shan., Sjaugi, Muhammad Farhan.** (2022) DiMA - Diversity Motif Analyser, <https://github.com/pu-sds/DiMA>.
- [70] **Hu, Y., Tan, P. T., Tan, T. W., August, J. T., & Khan, A. M.** (2013). Dissecting the Dynamics of HIV-1 Protein Sequence Diversity. *PLoS ONE*, 8(4), e59994. <http://doi.org/10.1371/journal.pone.0059994>
- [71] **Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B.** (2018) The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* Oct 24. doi: 10.1093/nar/gky1006.
- [72] **Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M.** (2020) NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* Jul 2;48(W1):W449-W454. doi: 10.1093/nar/gkaa379. PMID: 32406916; PMCID: PMC7319546.
- [73] **Moutaftsi M, Peters B, Pasquetto V, Tschärke DC, Sidney J, Bui HH, Grey H, Sette A.** (2006) A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus. *Nat Biotechnol.* Jul;24(7):817-9. doi: 10.1038/nbt1215. Epub 2006 Jun 11. PMID: 16767078.
- [74] **Kotturi MF, Peters B, Buendia-Laysa F Jr, Sidney J, Oseroff C, Botten J, Grey H, Buchmeier MJ, Sette A.** (2007) The CD8+ T-cell response to lymphocytic choriomeningitis virus involves the L antigen: uncovering new tricks for an old virus. *J Virol.* May;81(10):4928-40. doi: 10.1128/JVI.02632-06. Epub 2007 Feb 28. PMID: 17329346; PMCID: PMC1900207.
- [75] **Southwood S, Sidney J, Kondo A, del Guercio MF, Appella E, Hoffman S, Kubo RT, Chesnut RW, Grey HM, Sette A.** (1998) Several common HLA-DR types share largely overlapping peptide binding repertoires. *J Immunol.* Apr 1;160(7):3363-73. PMID: 9531296.
- [76] **Schrödinger Release 2021-4:** (2021) Maestro, Schrödinger, LLC, New York, NY.
- [77] **Schrödinger Release 2021-4:** (2021) LigPrep, Schrödinger, LLC, New York, NY.

- [78] **Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W.,** (2010) "Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field," *J. Chem. Theory Comput.*, 6, 1509–1519
- [79] **Sastry, G.M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W.** (2013) "Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments," *J. Comput. Aid. Mol. Des.*, , 27(3), 221-234
- [80] **Schrödinger Release 2021-4:** (2021) Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2021; Impact, Schrödinger, LLC, New York, NY; Prime, Schrödinger, LLC, New York, NY.
- [81] **Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T.** (2006) "Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes," *J. Med. Chem.*, 49, 6177–6196
- [82] **Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L.** (2004) "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening," *J. Med. Chem.*, 47, 1750–1759
- [83] **Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shaw, D. E.; Shelley, M.; Perry, J. K.; Francis, P.; Shenkin, P. S.** (2004) "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy," *J. Med. Chem.*, 47, 1739–1749
- [84] **Schrödinger Release 2021-4:** (2021) Glide, Schrödinger, LLC, New York, NY.
- [85] **Kevin J. Bowers, Edmond Chow, Huafeng Xu, Ron O. Dror, Michael P. Eastwood, Brent A. Gregersen, John L. Klepeis, Istvan Kolossvary, Mark A. Moraes, Federico D. Sacerdoti, John K. Salmon, Yibing Shan, and David E. Shaw,** (2006) "Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters," *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)*, Tampa, Florida, November 11-17
- [86] **Schrödinger Release 2021-4:** (2021) Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY.



## **EKLER**

**EK A :** hRSV index dizileri ve motiflerinin sayısal bilgileri

**EK B :** MHC-I epitop tahmini sonuçları

**EK C :** MHC-II epitope tahmini sonuçları

**EK A : hRSV index dizileri ve motiflerinin sayısal bilgileri**

hRSV proteinlerinin tamamı için index dizilerini, entropi deęerlerini ve motiflerinin sayısal bilgilerine ařaęıdaki linkten ulařılabilir.

<https://raw.githubusercontent.com/farukustunel/proteinsequencediversityofhrsv/main/hrsvmotifsresults.csv>

## **EK B : MHC-I epitop tahmini sonuçları**

Bu çalışma kapsamında elde edilen MHC-I epitop tahmini sonuçlarının tamamına aşağıda verilen linkten ulaşılabilir.

[https://raw.githubusercontent.com/farukustunel/proteinsequencediversityofhrsv/main/mhc1\\_goodbinders.csv](https://raw.githubusercontent.com/farukustunel/proteinsequencediversityofhrsv/main/mhc1_goodbinders.csv)

### **EK C : MHC-II epitop tahmini sonuçları**

Bu çalışma kapsamında elde edilen MHC-II epitop tahmini sonuçlarının tamamına aşağıda verilen linkten ulaşılabilir.

[https://raw.githubusercontent.com/farukustunel/proteinsequencediversityofhrsv/main/mhc2\\_goodbinders.csv](https://raw.githubusercontent.com/farukustunel/proteinsequencediversityofhrsv/main/mhc2_goodbinders.csv)

## ÖZGEÇMİŞ

**Ad-Soyad** : Faruk Üstünel

**Doğum Tarihi ve Yeri** :

**E-posta** :

### ÖĞRENİM DURUMU:

- **Lisans** : 2019, İstanbul Üniversitesi, Fen Fakültesi, Moleküler Biyoloji ve Genetik Bölümü
- **Yüksek Lisans** : 2022, Bezmialem Vakıf Üniversitesi, Biyoteknoloji Anabilim Dalı, Biyoteknoloji Tezli Yüksek Lisans Programı

### MESLEKİ DENEYİM VE ÖDÜLLER:

- 2021-Günümüz, Araştırma Görevlisi, Bezmialem Vakıf Üniversitesi, Yaşam Bilimleri ve Biyoteknoloji Enstitüsü, Biyoteknoloji Anabilim Dalı

### YÜKSEK LİSANS TEZİNDEN TÜRETİLEN YAYINLAR, SUNUMLAR VE PATENTLER:

- **Üstünel F., Akdemir A., Asif M.K** (2022). Dissecting the Dynamics of Human Respiratory Syncytial Virus Protein Sequence Diversity and Identifying Their Potential Vaccine Targets. *10th International Drug Chemistry Conference*, March 10-13, Antalya, Turkey.

