



**DELAUNAY TESELASYONU SONUCUNDA ELDE
EDİLEN TETRAHEDRONLARIN KÜRESELLİK
DEĞERLERİ İLE PROTEİN YÜZEYLERİNİN VE
PROTEİN KOMPLEKSLERİNİN ARAYÜZLERİNİN
KARAKTERİZASYONU**

Mehmet GEZGİNCİ

Doktora Tezi

Biyoloji Anabilim Dalı

**Danışman: Doç. Dr. Elife Zerrin BAĞCI
2021**

T.C.
TEKİRDAĞ NAMIK KEMAL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DOKTORA TEZİ

**DELAUNAY TESELASYONU SONUCUNDA ELDE EDİLEN
TETRAHEDRONLARIN KÜRESELLİK DEĞERLERİ İLE PROTEİN
YÜZEYLERİNİN VE PROTEİN KOMPLEKSLERİNİN
ARAYÜZLERİNİN KARAKTERİZASYONU**

Mehmet GEZGİNCİ

BİYOLOJİ ANABİLİM DALI

DANIŞMAN: Doç. Dr. Elife Zerrin BAĞCI

TEKİRDAĞ-2021

Her hakkı saklıdır.



Bu tez (TÜBİTAK / SANTEZ / NKÜBAP vb.) tarafından
..... numaralı proje ile desteklenmiştir.

ÖZET

Doktora Tezi

DELAUNAY TESELYONU SONUCUNDA ELDE EDİLEN TETRAHEDRONLARIN KÜRESELLİK DEĞERLERİ İLE PROTEİN YÜZEYLERİNİN VE PROTEİN KOMPLEKSLERİNİN ARAYÜZLERİNİN KARAKTERİZASYONU

Mehmet GEZGİNCİ

Tekirdağ Namık Kemal Üniversitesi

Fen Bilimleri Enstitüsü

Biyoloji Anabilim Dalı

Danışman: Doç. Dr. Elife Zerrin BAĞCI

Tüm yaşam süreçlerinde kritik rol oynayan proteinlerin nasıl kompleks yapı oluşturdukları henüz tam olarak açıklanamamıştır. Bu çalışmada tekil proteinler ve protein komplekslerinin yapılarından elde edilmiş olan tetrahedronlar ve bu tetrahedronlara ait hacim değerleri ve küresellik değerleri incelenerek protein komplekslerinin yüzey ve arayüzlerinin karakterize edilmesi amaçlanmıştır. Bu çalışmada kullanılan proteinler ComSin veri tabanından seçilmiş ve her bir proteinin Protein Data Bankasındaki (PDB) elde edilmiş atom koordinatlarından yararlanılarak Delaunay teselasyonu ile tetrahedronlar belirlenmiştir. Bu tetrahedronlar kullanılarak tetrahedronların hacim ve küresellik değerleri hesaplanmıştır. Çalışma sonucunda protein yüzeyini oluşturan tetrahedronların, toplam tetrahedronların ~%10'unu oluşturduğu görülmüştür. Yüzey tetrahedronlarının ~%95'nin oluşturduğu homo-dimer protein arayüzlerine ait tetrahedronlarının hacmi, yüzeydekilerin hacimlerine göre büyük bulunmuşken küresellik değerleri küçük bulunmuştur. Yüzey tetrahedronları belirlenirken az sayıda tetrahedronun protein yüzeyinde iki üçgene sahip oldukları fark edilmiş ve bunlara “keskin tetrahedronlar” denilmiştir. Keskin tetrahedronların protein arayüzüne geçme oranlarının daha yüksek olduğu gözlemlenmiştir. Retroviral proteaz ailesine ait proteinlerde tespit edilen keskin tetrahedronlara ait olan rezidülerin arasında deneysel olarak önemli oldukları belirlenen rezidülerin de olduğu gösterilmiştir. Bu sonuç keskin tetrahedronların proteinlerin kompleks yapılarında önemli bölgeleri temsil ettiklerine işaret etmektedir. Protein komplekslerindeki protein yüzeyinin ve arayüzünün belirlenmesi, kanser ve Alzheimer hastalığı gibi yaşamı tehdit eden durumlar için yeni tedavilerin ve yeni ilaçların geliştirilmesine olanak sağlayacaktır.

Anahtar kelimeler: Protein yüzeyi, Protein arayüzü, Protein etkileşimleri, Delaunay teselasyonu, Tetrahedron, Küresellik

2021, 107 sayfa

ABSTRACT

PhD Thesis

CHARACTERIZATION OF THE SURFACES OF PROTEINS AND THE INTERFACES OF PROTEIN COMPLEXES WITH THE SPHERICITY VALUES OF TETRAHEDRA OBTAINED BY DELAUNAY TESSELLATION

Mehmet GEZGİNCİ

Tekirdağ Namık Kemal University

Graduate School of Natural and Applied Sciences

Department of Biology

Supervisor: Associate Professor Elife Zerrin BAĞCI, PhD

It has not been fully explained how proteins, which play a critical role in all life processes, form complex structures. In this study, it is aimed to characterize the surfaces and interfaces of protein complexes by examining the tetrahedra obtained from the structures of individual proteins and protein complexes, and the volume and sphericity values of these tetrahedra. The proteins used in this study were selected from the ComSin database and tetrahedra were determined by Delaunay tessellation of the atomic coordinates of each protein obtained from the Protein Data Bank (PDB). Using these tetrahedra, the volume and sphericity values of the tetrahedra were calculated. As a result of the study, it was found that the tetrahedra forming the protein surface constitute ~10% of the total tetrahedra. While the volume of the tetrahedra belonging to the homo-dimer protein interfaces formed by ~95% of the surface tetrahedra was found to be large compared to the volumes of the surface tetrahedra, the sphericity values were found to be small. When the surface tetrahedra was examined, it was noticed that a few tetrahedra had two triangles on the protein surface and these were called "sharp tetrahedra" in this study. It was observed that sharp tetrahedra had a higher rate of transition to the protein interface. It has been shown that residues belonging to sharp tetrahedra detected in proteins belonging to the Retroviral Protease family consisted the residues studied experimentally and determined to be important. This result indicates that sharp tetrahedra represent important regions in the complex structures of proteins. Determining the protein surface and interface in protein complexes will enable the development of new treatments and new drugs for life-threatening conditions such as cancer and Alzheimer's disease.

Keywords: Protein surface, Protein interface, Protein interactions, Delaunay tessellation, Tetrahedron, Sphericity

2021, 107 pages

İÇİNDEKİLER

ÖZET.....	i
ABSTRACT	ii
İÇİNDEKİLER.....	iii
ÇİZELGE DİZİNİ.....	vi
ŞEKİL DİZİNİ.....	vii
SİMGELER ve KISALTMALAR.....	ix
TEŞEKKÜR.....	x
1. GİRİŞ.....	1
1.1. Çalışmanın Amacı	1
1.2. Tezin Organizasyonu	2
2. KAYNAK ÖZETLERİ.....	Error! Bookmark not defined.
3. PROTEİNLER VE BİRBİRLERİYLE OLAN ETKİLEŞİMLERİ.....	5
3.1. Protein Yapısı	5
3.1.1. Amino Asitler ve Birincil Yapı.....	6
3.1.2. İkincil Yapı	9
3.1.3. Üçüncül Yapı	9
3.1.4. Dördüncül Yapı	10
3.2. <i>In Vitro</i> Yöntemler	11
3.2.1. X-ray Kristallografisi.....	11
3.2.2. SAXS (Small-Angle X-ray Scatering).....	12
3.2.3. NMR (Nükleer Manyetik Rezonans).....	12
3.2.4. Elektron Mikroskobu (EM)	12
3.3. Protein-Protein Etkileşim Çalışmalarında Kullanılacak Yapıların Belirlenmesi	13
3.3.1. Protein Homolojisi.....	13
3.3.2. Paralog ve Ortolog Genler	14
3.3.3. Sekans Dizi Aynılığı.....	15
3.4. İkili Protein Komplekslerinin Sınıflandırılması	18
3.5. Domeyn-Domeyn Arayüzlerinin Yapısı	18
3.6. Arayüzlerde Bulunan Rezidüleri.....	19
3.6.1. Rezidü Sayımı.....	19
3.6.2. Rezidü Çiftleri	20
3.6.3. Hot Spot'lar	20

4. PROTEİN ETKİLEŞİMLERİ İÇİN BİYİNFORMATİK YAKLAŞIMLAR	22
4.1. Protein Veri Bankası	22
4.1.1. PDB Formatı	23
4.2. ComSin Veri Tabanı	25
4.2.1. ComSin'in Oluşturulmasında Kullanılan Veri Tabanları	25
4.2.2. ComSin Veri İçeriği	27
4.3. Protein Arayüz Tahmininde Hesaplamalı Yaklaşımlar	29
4.3.1. Yapı-içi Temelli Yaklaşım	30
4.3.2. Kalıp Temelli Yaklaşım	31
4.4. Voronoi/Delaunay Teselasyonu	32
4.4.1. Voronoi Teselasyonu	32
4.4.2. Delaunay Teselasyonu	33
4.5. Tetrahedronların Geometrik Karakterizasyonlarına Örnekler	34
4.5.1. Tetrahedronluk Özelliği ("Tetrahedrality")	35
4.5.2. Küresellik ("Sphericity")	35
5. MATERYAL VE METOT	37
5.1. Materyal	37
5.1.1. Proteinlerin Belirlenmesi	37
5.1.2. Protein Veri Setinde Yapılan Bazı Manipülasyonlar	47
5.2. Metot	50
5.2.1. Tetrahedronların Hacimlerinin Hesaplanması	50
5.2.2. Tetrahedronların Küresellik Değerlerinin Hesaplanması	54
5.2.3. Protein Yüzeyinin Belirlenmesi	55
5.2.4. Protein Arayüzünün Belirlenmesi	59
5.2.5. Protein Yüzeyinde ve Arayüzünde Bulunan Keskin Tetrahedronların Belirlenmesi.	61
6. ARAŞTIRMA BULGULARI	Error! Bookmark not defined.
6.1. Tekil ve Kompleks Yapılarda Tetrahedron Sayıları, Ortalama Hacim ve Ortalama Küresellik Değerlerindeki Değişim	62
6.2. Tekil Proteine Ait Tetrahedron Sayılarının Protein Yüzeyindeki ve Arayüzündeki Değişimleri	65
6.3. Tekil Proteine Ait Tetrahedron Hacim Ortalamalarının Protein Yüzeyindeki ve Arayüzündeki Değişimleri	70
6.4. Tekil Proteine Ait Tetrahedron Ortalama Küresellik Değerlerinin Protein Yüzeyindeki ve Arayüzündeki Değişimleri	74

6.5. Rezidü Sayıları.....	78
6.6. Arayüzdeki “Kilit Noktaların” Rezidü Kompozisyonu	82
7. COMSIN VERİ TABANINDA BULUNAN RETROVİRAL PROTEAZ PROTEİN AİLESİNE AİT PROTEİN KOMPLEKSLERİNİN ANALİZİ	87
8. TARTIŞMA VE SONUÇ	92
KAYNAKLAR.....	95
ÖZGEÇMİŞ	106



ÇİZELGE DİZİNİ

Çizelge 3.1. Amino asitler	6
Çizelge 3.2. Tersinir hetero-dimerik komplekslerinde en çok bulunan 10 rezidü çifti	21
Çizelge 4.1. ComSin veri tabanında bulunan proteinlere ait bilgiler	28
Çizelge 5.1. Çalışmada kullanılacak proteinlere ait bilgiler	39
Çizelge 5.2. İlgili PDB dosyalarında değiştirilen HETATM'lar ve sayıları	48
Çizelge 5.3. Tetrahedronu oluşturan C α (alfa karbon) atomlarına ait x, y ve z koordinatları..	51
Çizelge 5.4. 1ejc PDB kodlu proteine ait başlangıç tetrahedronları.....	56
Çizelge 6.1. Homo-dimer kompleksleri oluşturan zincirler ve arayüzlere ait bilgiler	62
Çizelge 6.2. Tekil, yüzey ve arayüzdeki tetrahedron değerleri	66
Çizelge 6.3. Anlamlı bir biçimde önemli olduğu düşünülen protein çiftlerinde kilit noktaların bulunması.....	69
Çizelge 6.4. Tekil, yüzey ve arayüzdeki tetrahedronlara ait ortalama hacim değerleri.....	71
Çizelge 6.5. Tekil, yüzey ve arayüzdeki tetrahedronlarına ait sayı ortalama küresellik değerleri	75
Çizelge 6.6. Tekil protein ve komplekslerinin arayüz bölgelerindeki rezidü sayıları	78
Çizelge 6.7. Arayüzdeki “kilit noktaların” rezidü içeriği	82
Çizelge 7.1. Retroviral Proteaz protein ailesine ait bazı protein komplekslerinin tetrahedron, rezidü sayısı, hacim ve küresellik bilgileri	89
Çizelge 7.2. Tekil protein ve arayüzdeki rezidülerin sayısı ve isimleri	90
Çizelge 7.3. 2o4p protein kompleksindeki arayüzündeki rezidüler	91

ŞEKİL DİZİNİ

Şekil 3.1. Amino asitlerin yapısı (Carneiro 2019)'dan modifiye edilmiştir	7
Şekil 3.2. En basit amino asit: Glisin (Carneiro 2019)'dan modifiye edilmiştir	8
Şekil 3.3. Amino asitlerin sınıflandırılması	8
Şekil 3.4. İkincil yapıları renklendirilmiş 1zp8 PDB kodlu kompleks (Rose ve ark. 2018)	9
Şekil 3.5. 2o4p PDB kodlu homomonomerin yapısı (Rose ve ark. 2018)	10
Şekil 3.6. Protein yapısını anlamadaki çeşitli deneysel metotların güçlü yanlarının ve limitleri (Srivastava ve ark. 2018)	11
Şekil 3.7. Homolog Genler	14
Şekil 3.8. Rubredoxin Proteini	15
Şekil 3.9. Hipotetik Protein	16
Şekil 3.10. Hipotetik Proteini	17
Şekil 3.11. Standart Yama ve Arayüz Büyüklüğü (Meyerkord ve Fu 2015)	19
Şekil 3.12. Arayüzde rezidü kullanımı (Meyerkord ve Fu 2015)	20
Şekil 4.1. PDB'de yıllık olarak depolanan toplam protein yapıları (mavi) ve yıl başına yeni protein yapıları (lacivert)	23
Şekil 4.2. 1trm PDB'sine ait kompleks reseptör molekülünden PDB dosyasından bir kesit. Amino asitlere (yeşil renk) ait atomların (mor renk) x,y ve z koordinatlarını (kırmızı renk) gösterilmektedir.	23
Şekil 4.3. 1hp2 PDB dosyasındaki eksik rezidüleri (solda) ve eksik atomlar (sağda)	24
Şekil 4.4. 1trm PDB kodlu proteine ait "Occupancy" değerleri	24
Şekil 4.5. Pirüvat kinaz; C-Terminal (mavi), N-Terminal (kırmızı) ve PK domeyn (yeşil) rezidülerinden oluşmuş üç domeyn içermektedir (Bahar ve ark. 2017)	26
Şekil 4.6. Tekil (A ve B; sırasıyla PDB kodları 1RRF ve 1R8M) ve kompleks (C; PDB kodu 1R8Q) yapıdaki ADP-ribozilasyon faktör (yeşil ve mavi zincirler) ve değişim faktörü ARNO (mavi ve sarı zincirler). Kesik çizgiler özünde düzensiz rezidüleri ifade etmektedir.	29
Şekil 4.7. Voronoi teselasyonunun ve Voronoi'den elde edilen yapılar. a) Bir noktanın Voronoi hücresi b) Voronoi hücrelerinin tekrarlanması sonucu elde edilen Voronoi teselasyonu. c) Ortak bir Voronoi yüzü paylaşan noktalar arasındaki bütün segmentlerin çizimiyle elde edilen Delaunay teselasyonu (Poupon 2004).	33
Şekil 4.8. Bazı yaygın "kötü" şekilli tetrahedronlar (Parthasarathy 1994)	34
Şekil 5.1. Protein veri setinin oluşturulması	38
Şekil 5.2. 1aiq PDB kodlu proteine ait "CXM" HETATM'ları	49
Şekil 5.3. Bir proteinden tetrahedronların elde edilmesi (Aydinkal ve Bağcı 2018)'den modifiye edilmiştir	50

Şekil 5.4. Bir paralel yüzünün içerisine yerleşen 6 adet tetrahedronun şematik gösterimi (Dune-Project.org)	53
Şekil 5.5. Paralel yüzünün hacminin hesaplanması.....	53
Şekil 5.6. 2B’da çok sayıda birbirine komşu olan üçgenler kümesinden oluşmuş bir yapıda yüzey üçgenlerinin belirlenmesi	55
Şekil 6.1. 1a4u ve 1a2w pdb kodlu protein komplekslerine ait tetrahedronların hacimlerinin ortalamalarına ait histogram grafikleri. Mavi renkle gösterilen histogram kompleksin birinci zincirine, sarı renkle gösterilen ikinci zincirine, yeşil renkle gösterilen ise arayüzüne aittir. Üsteki protein kompleksi 1a4u pdb kodlu protein kompleksine aitken, alttakiler 1a2w pdb kodlu protein kompleksini göstermektedir	65
Şekil 6.2. Tetrahedron sayıları.....	70
Şekil 6.3. Tetrahedron hacim ortalamaları	74
Şekil 6.4. En düşük ortalama küresellik değerine sahip (0,41) 3ezm PDB kodlu protein (solda), en büyük ortalama küresellik değerine sahip (0,50) 1vd5 PDB kodlu protein (sağda) (Sehnal ve ark. 2021).....	75
Şekil 6.5. Protein kompleksinde bulunan sıcak noktaların rezidülerinin dağılımı	86
Şekil 7.1. 2o4k PDB kodlu HIV proteaz proteininde 4 iplikli beta-tabakalar (Sehnal ve ark. 2021).....	87
Şekil 7.2. 2avs PDB kodlu HIV Proteaz ve ligandları.....	88
Şekil 7.3. HIV proteazın arayüzündeki 4 iplikli beta-tabaka dimerleşmesi (Bowman ve Chmielewski 2002).....	88

SİMGELER VE KISALTMALAR

ARF	: ADP-Ribosylation Factor-1
ARNO	: The Arf Exchange Factor ARNO
Å	: Angström
CATH	: Class, Architecture, Topology, Homologous superfamily
CBM	: Conserved Binding Modes
CDD	: Conserved Domain Database
COG	: Cluster of Ortholog Groups
ComSin	: Database of protein structures in bound (Complex) and unbound (Single) states in relation to their intrinsic disorder: COMplexSINgle
C α	: Alfa Karbon
ÇSD	: Çoklu Sekans Dizilemeleri
EM	: Elektron Mikroskobu
kDa	: Kilo Dalton
NMR	: Nükleer Manyetik Rezonans
Pfam	: Protein Families Database
PDB	: Protein Data Bank
PRISM	: Protein Interactions by Structural Matching
pml	: PyMol
RCSB	: Research Collabatory of Structural Biology
SAXS	: Small-Angle X-ray Scatering
SCOP	: Structural Classification of Proteins
SMART	: Simple Modular Architecture Research Tool
UniProt	: Universal Protein Resource
2B	: 2 Boyutlu
3B	: 3 Boyutlu

TEŞEKKÜR

Tez çalışmam ve lisans üstü öğrenimimde bilgilerini ve deneyimlerini benimle paylaşan ve desteğini benden esirgemeyen, bana yol gösteren değerli danışman hocam Doç. Dr. Elife Zerrin BAĞCI'ya,

Tez İzleme Komitemde yer alan ve her dönem değerleri fikirlerini ve katkılarını sunan saygı değer hocalarım Doç. Dr. Duygu YAŞAR ŞİRİN' e (Tekirdağ Namık Kemal Üniversitesi Fen-Edebiyat Fakültesi) ve Doç. Dr. Gökhan KAÇAR' a (Edirne Trakya Üniversitesi Mühendislik Fakültesi),

Tez sürecinde manevi desteklerini hiç esirgemeyen Dr. Medine Münevver UMA KOÇUM, Filiz ATALAY ve Mehmet Ali ŞAHİN'e,

Son olarak çalışmam süresince bana güç kaynağı olan aileme sonsuz teşekkürlerimi sunarım.

Temmuz, 2021

Mehmet GEZGİNCİ
Biyoloji Öğretmeni

1. GİRİŞ

Proteinler ve proteinlerin diğer moleküllerle etkileşimleri canlılığın temelini oluşturmaktadır. Neredeyse hücre içindeki bütün biyolojik süreçler proteinler ya da protein kompleksleri tarafından sürdürülmektedir. Dünyadaki yaşamın tüm yönlerinin ayrılmaz bir parçası olan protein-protein etkileşimlerinde bilinmeyen örnekler sürekli olarak bulunmaktadır. Proteinlerin komplekslerinin yapılarına ilişkin bilgi, proteinlerin işlevlerinin anlaşılması için gereklidir. Ancak deneysel protein yapı belirlemedeki güncel yöntemler bu gereksinime yeterince cevap verememektedir.

Protein yapılarının belirlenmesinde kullanılan *in vitro* ve *in vivo* yöntemler pahalı, uzun zaman alan ve çok vakit gerektiren metotlardır. Protein Data Bank (PDB)'taki yapıların ~88'i *in vitro* yöntemlerden birisi olan X-ray kristalografisi tarafından belirlenmiştir. X-ray kristalografisi için yüksek kalitede kristal oluşumu gereklidir. Proteinlerin boyutları ve kompleksliği yüzünden bu işlem uzun bir süreç gerektirmektedir ve oldukça zordur. Ayrıca X-ray kristalografisinde proteinlerin anlık (İng. snapshot) etkileşim profilleri elde edilmektedir.

Protein komplekslerinin yapılarını belirlemedeki bu zorluk, tek tek proteinlerin yapıları ile protein komplekslerinin yapıları arasında bir bilgi boşluğuna yol açar. Bu boşluğu kapatmak ve deneysel süreçlerin sınırlamalarının üstesinden gelmek için araştırmacılar bu proteinler hakkında var olan bilgilerden yararlanarak hesaplamalı yöntemlere yönelmişlerdir.

Protein komplekslerinin yapılarını tahmin etmeye çalışan hesaplama yöntemleri arasında geometrik temelli yaklaşımlar da bulunmaktadır. Delaunay/Voronoi teselasyonu ile yapılan çalışmalarda atomların 3 boyutlu (3B) koordinatlarından protein yüzeyleri ve protein-protein arayüzleri belirlenir (Nadassy ve ark. 1999; Nadassy ve ark. 2001; Singh ve ark. 1996).

Protein yapı çalışmaları proteinlerin katlanmış yapılarını, nasıl etkileştiklerini ve nasıl işlev gördüklerini anlamak için elzemdir. Protein yapılarından elde edilen bilgiler hastalıkların anlaşılmasında ve tedavisinde, ilaçların ve insan hayatının iyileştirilmesinde kullanılacak diğer ürünlerin geliştirilmesinde öncü olabilir.

1.1. Çalışmanın Amacı

Bu tez çalışmasının amacı geometrik temelli hesaplamalı yaklaşım kullanılarak, proteinlerin yüzeylerindeki ve protein komplekslerinin arayüzlerindeki tetrahedronlar,

tetrahedronların hacmi ve bunlara ait küresellik deęerleri incelemek ve proteinlerin ve protein komplekslerinin yapılarından elde edilecek olan küresellik deęerlerinin incelenmesi ile protein komplekslerinin arayüzlerini tahmin etmek için yeni bir yaklaşım ortaya koymaktır.

1.2. Tezin Organizasyonu

Okuyucuların bu tezin genel akışına aşına olabilmesi için bu bölümde her bir bölüme ait bazı temel bilgilendirilmeler yapılacaktır. 2. bölümde; bu tez konusuna benzer çalışmalar hakkında genel literatür araştırması özetlenecek, 3. bölümde; proteinleri oluşturan amino asitler ve amino asitlerin isimlendirilmeleri, protein yapısı ve protein homolojisi genel bilgiler verilecek, 4. bölümde; bu tezde kullanılan veri tabanları, teselasyon yöntemleri, küresellik, tetrahedronluk hakkında bilgi verilecek, 5. bölümde; bu tezde kullanılacak proteinler, proteinler üzerinde yapılan manipülasyonlardan bahsedilip, çalışmada kullanılan yöntemler açıklanacak; 6. bölümde; elde edilen bulgular ortaya konulacak 7. bölümde; çalışmada elde edilen bulguların güvenilirliği için HIV 1 Proteaz protein ailesine ait bazı protein komplekslerle kıyaslamalar yapıldıktan sonra son bölüm olan 8. bölümde çalışmanın sonuçları tartışmalarıyla birlikte verilecektir.

2. KAYNAK ÖZETLERİ

Protein-protein arayüzlerini iki tek zincirli protein arasındaki bağlanma yüzeyi (Ing., architectural binding surface) elemanlarını geniş ölçekte karakterize etmek amacıyla, 2007 yılındaki yapılan çalışmada, bağlanma enerjisinin büyük çoğunluğuna katkı sağlayan amino asitler olan, arayüz sıcak noktalarının (“hot spots”) tahminine ve protein bağlanma kısımlarının üzerlerindeki etkileşim halindeki örüntüleri ifade eden arayüzdeki uzaysal desenlerinin keşfedilmesine yönelik yeni hesaplamalı teknikler sunulmuştur (Güney, 2007).

2007 yılında yapılan başka bir çalışmada, arayüz yapılarının sayısının protein yapılarının sayısı ile karşılaştırıldığında çok daha hızlı bir şekilde arttığını ve henüz bu artışın durmadığını göstermiştir. Arayüzlerin yapısal çeşitliliğine rağmen, bazı arayüzlerin tek zincirli proteinler tarafından daha çok kullanıldığı görülmüştür. Bir başka önemli sonuç da bazı türlerin, özellikle hücre kompleksliğinin bir adaptasyonu sonucu ökaryotların zincir içi etkileşimi, daha az gelişmiş organizmaların ise zincirler arası etkileşimi tercih ettiği görülmüştür (Tunçbağ 2007).

2009 yapılan çalışmada, proteinlerin Gaussian Ağ Modellenmesi ile elde edilen amino asitlerin yay hareketi yaptığı varsayımına dayanan metodun güvenilirliği iki veri tabanı üzerinde test edilmiştir (Soner, 2009). PPI-Pred veri tabanındaki 111 proteinden zorunlu proteinlerde başarılı tahmin oranı yüzde 82, zorunlu olmayan proteinlerde başarılı tahmin oranı yüzde 76,5 iken CAPRI yarışmasında ise, 1600 model arasından seçilen on modelden altısı başarılı olmuştur (Ozbek ve ark. 2013).

Biyoinformatik yaklaşımlardan birisini oluşturan çalışmalardan, PRISM (“Protein Interactions by Structural Matching”) adlı yazılım; protein bileşiklerinden elde edilmiş arayüzlerin sorgulanması, incelenmesi ve görselleştirilmesi için geliştirilmiştir. PRISM; protein-protein arayüz tahmini için şablon (“template”) ve hedef (“target”) olmak üzere iki tip veri-seti kullanır. Şablon veri seti; yapısal olarak mevcut olan protein-protein etkileşimlerinin tamamını tanımlamak ve potansiyel olarak etkileşen protein çiftlerini tahmin etmek için kullanılırken, hedef veri seti; proteinler arasındaki potansiyel etkileşimlerin tahmin edilebileceği proteinleri içermektedir. PRISM yazılımıyla yapılmış çalışmalardan ilki, ikili benzerlik verilerinden yararlanarak etkileşim haritası çıkartılması, bunların görselleştirilmesi ve grafiksel ortamda araştırılmasına yöneliktir (Ogmen 2006). İkincisi ise iki hedef proteinin yüzeylerindeki yamalar (“patches”), şablonun arayüzündeki iki tamamlayıcı

(“complementary”) yüze benziyor ise, bunların kompleks oluşturmak için etkileşiyor olabileceğini varsayar (Şekil 3.11) (Baspınar ve ark. 2014).

2015 yılındaki çalışma temel olarak, arayüz yapıları açısından protein-protein etkileşimlerinin kümelenmesine odaklanmıştır. PDB'de 22604 adet eşsiz arayüz yapısı bulunmuş ve bu eşsiz yapılar, birbirleriyle etkileşmek için çoklu bağlanma bölgelerine sahip olan protein çiftlerini bulmak için kullanılmıştır. PFAM domeynleri sınıflandırması temel alınarak hem aynı hem de tamamlayıcı monomerlerdeki arayüzlerdeki domeyn kombinasyonları ilk kez analiz edilmiştir. Bu domeyn kombinasyonları bir ağ oluşturularak sunulmuştur. PFAM domeyn analizine ek olarak, SCOP (“Structural Classification of Proteins”)’un katlanma (“fold”), aile ve süper aile sınıflandırılması, her bir sınıflandırma seviyesindeki SCOP domeynlerini bulmak amacıyla protein arayüz yapılarının karakterize edilmesi için haritalandırılmıştır. Protein arayüz yapıları ve arayüz domeyn kombinasyonları üzerine yapılmış bu geniş kapsamlı çalışmanın benzeri çalışmalar için kaynak olması düşünülmüştür (Çukuroğlu 2015).

Budowski ve ark. (2018) yaptıkları çalışmada “kelime torbası” olarak adlandırılan bir metodu kullanarak protein arayüzünü tahmin etmeye çalışmışlardır. Bu çalışma için proteinin yüzeyinde bulunan rezidülerin alfa karbonlarının ($C\alpha$) 3-boyutlu koordinasyonları kullanılmıştır. Merkezi rezidünün $C\alpha$ 'una yakın olan diğer rezidülerin $C\alpha$ 'larıyla 4-5-6 rezidülü bir yama oluşturulmuştur. Bu yamalar vektörlerle temsil edilmiştir. Böylece uzayda farklı yönlene vektör kümeleri elde edilmiştir. Protein arayüzündeki bu farklı vektör kümelerinin histogramlarına bakılarak, proteinler hakkında ekstra bir bilgiye sahip olmadan, arayüzleri tahmin edilebilmiştir (Budowski ve ark. 2018).

$C\alpha$ koordinat noktalarının Delaunay teselasyonu sonucu elde edilmiş tetrahedronların küreselliğinin kullanılması sonucu α , β , α/β , $\alpha+\beta$ yapısal sınıflarına ait olan globüler proteinlerin küresellik olası dağılımları gözlemlenmiştir (Aydinkal ve Bağci 2018) ve daha sonra kullanılan yöntem geliştirilerek α ve β yapısal sınıflar başarılı bir şekilde birbirlerinden ayrıştırılabildiği (Bağci ve ark. 2021). Ayrıca 3B'lu hemoglobinlerin R ve T konformasyonları analiz edilmiş ve bu proteinlerin farklılıkları bu yöntemle ortaya konulmuştur (Bağci ve ark. 2021).

3. PROTEİNLER VE BİRBİRLERİYLE OLAN ETKİLEŞİMLERİ

İkili protein-protein etkileşimleri interaktom olarak adlandırılır ve insan için tahmin edilen etkileşim sayısı 600.000 civarındır. Bu etkileşimlerin küçük bir oranı deneysel olarak belirlenmiştir. Hücre içinde oluşan kompleksleri yani üretken protein-protein etkileşimlerini anlamak için başlangıç noktası proteinlerin binlerce kristal yapılarını içeren ve haftalık olarak güncellenen Protein Data Bank (PDB)'dir. Protein komplekslerinin arayüzey yapılarının sayısı ile PDB'daki protein yapılarının sayısı karşılaştırıldığında, arayüzey yapılarının sayısının çok daha hızlı bir şekilde arttığı ve henüz bu artışın durmadığı görülmüştür (Tunçbağ ve ark. 2007). STRING veri tabanına göre; 2007 yılında yaklaşık 36 milyon kadar protein etkileşimi bulunmuşken bu sayı 2013 yılında yaklaşık 225 milyona, 2021 yılında yaklaşık 3,5 milyara ulaşmıştır (STRING 2021). Örneğin, CATH veri tabanındaki homolog protein süper ailelerinin sayısındaki büyüme stabil hale gelmiştir (Kundrotas ve ark. 2013) ve daha da artması beklenmemektedir. Bu da protein "katlanma-uzayı"nın sınırlı ve çalışılabilir büyüklükte olduğunu göstermektedir. Ayrıca "etkileşim-uzayı"nın da sınırlı olduğu düşünülmektedir. Ancak her tür için proteomda beklenen etkileşim sayısı değişir. Örneğin bu sayı *S. cerevisiae*'de 25000 civarı iken insanda 600.000 civarındır (Stumpf ve ark. 2008). İnsan proteomunun bir taslak haritasının elde edildiği çalışmada insanda farklı dokulardan elde edilen peptitlerin sayısının 290.000'den fazla olduğu sonucuna ulaşılmıştır (Kim ve ark. 2014). Dolayısıyla on binlerce proteinin ikili kombinasyonlarının sayısının 600.000'den ne kadar büyük olduğu düşünülürse, protein-protein komplekslerinin tahmin edilmesinin ne kadar zorlu olduğu ortaya çıkar. Bu nedenle, hesaplamalı çalışmalar önem taşımaktadır. Bu bölümde proteinlerin ve protein-protein komplekslerinin yapısal özellikleri hakkında bilgiler sunulmaktadır.

3.1. Protein Yapısı

Proteinler canlıların temel bileşenlerinden biridir. Vücudumuzdaki her hücre birçok biyolojik fonksiyonu yerine getirmek için protein sentezlerler. Bu moleküler araçlar bütün yaşamsal süreçlerde kritik rollerden sorumludurlar. Bu sorumluluklara örnek olarak: hücrelerin kendi şekillerini sürdürmelerine izin veren yapısal desteği sağlarlar; biyokimyasal reaksiyonları kolaylaştırmada kataliz olarak rol oynarlar; protein transkripsiyonuna ve hormonların düzenlenmesine yardımcı olurlar; amino asit ve metal iyonların biyolojik rezervleri olarak hizmet görürler ve organizma içerisinde iyonları ve diğer molekülleri taşırlar (Branden ve Tooze 2012; Burkowski ve ark. 2008; Lesk 2001).

Yapıları aktivitelerine bağlıdır ve işlevleri kadar çeşitlidir. Protein yapı çalışmaları proteinlerin nasıl oluştuğunu, nasıl etkileştiklerini ve nasıl işlev kazandıklarını anlamak için oldukça elzemdir. Protein yapıların elde edilen bilgiler hastalıkların anlaşılmasında ve korunmasında ya da yeni ve geliştirilmiş tedavi yöntemlerinin, ilaçların ve insan hayatının iyileştirilmesinde kullanılacak diğer ürünlerin geliştirilmesinde kıymetli olabilir.

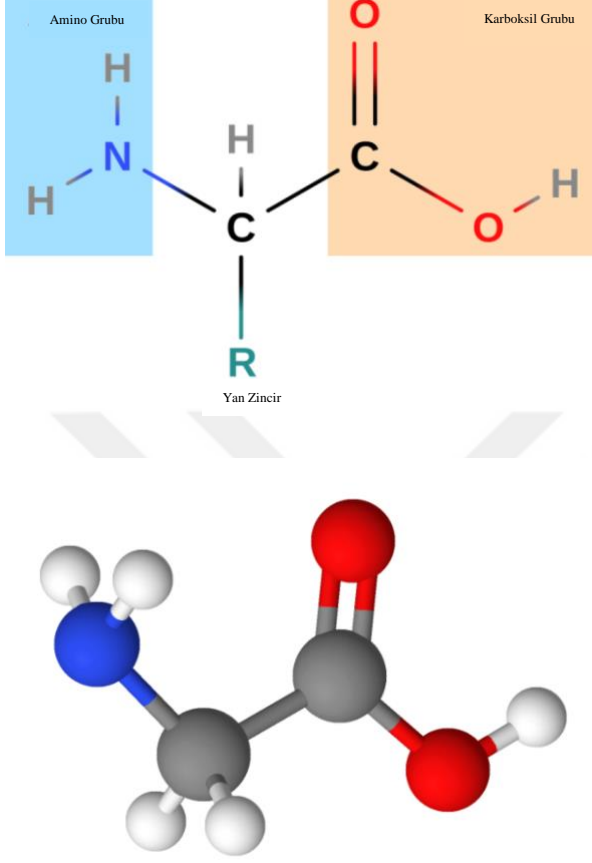
3.1.1. Amino Asitler ve Birincil Yapı

Amino asitler proteinlerin yapı taşlarıdır. 20 çeşit amino asit bulunmaktadır (Çizelge 3.1).

Çizelge 3.1. Amino asitler

Amino asit adı	Kısaltma		Amino asit adı	Kısaltma	
Glisin	GLY	G	Treonin	THR	T
Alanin	ALA	A	Metiyonin	MET	M
Valin	VAL	V	Sistein	CYS	C
Lösin	LEU	L	Asparajin	ASN	N
İzolösin	ILE	I	Glutamin	GLN	Q
Prolin	PRO	P	Aspartat	ASP	D
Fenilalanin	PHE	F	Glutamat	GLU	E
Tirozin	TYR	Y	Lizin	LYS	K
Triptofan	TRP	W	Arjinin	ARG	R
Serin	SER	S	Histidin	HIS	H

Her bir amino asitte ortak olarak merkezi karbon ($C\alpha$), buna bağı amino grubu (NH_2), karboksil grubu ($COOH$) ve her amino asitte farklılık gösteren yan zincir (R) grupları bulunur (Şekil 3.1).



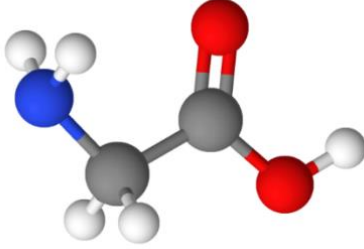
Şekil 3.1. Amino asitlerin yapısı (Carneiro 2019)'dan modifiye edilmiştir

Yan zincirler her bir amino aside has kimyasal ve sterik karakter vererek amino asitlerin birbirlerinden farklılaşmasını sağlarlar. Bu özellikler, farklı amino asit dizilimleri ile birlikte, proteinlerin yapılarını, fonksiyonlarını ve etkileşimlerini etkiler ve tanımlar.

Amino asitler en basit ve yaygın olarak; yan zincirlerinin özelliklerine göre ve polaritelerine göre sınıflandırılırlar. (Şekil 3.2).

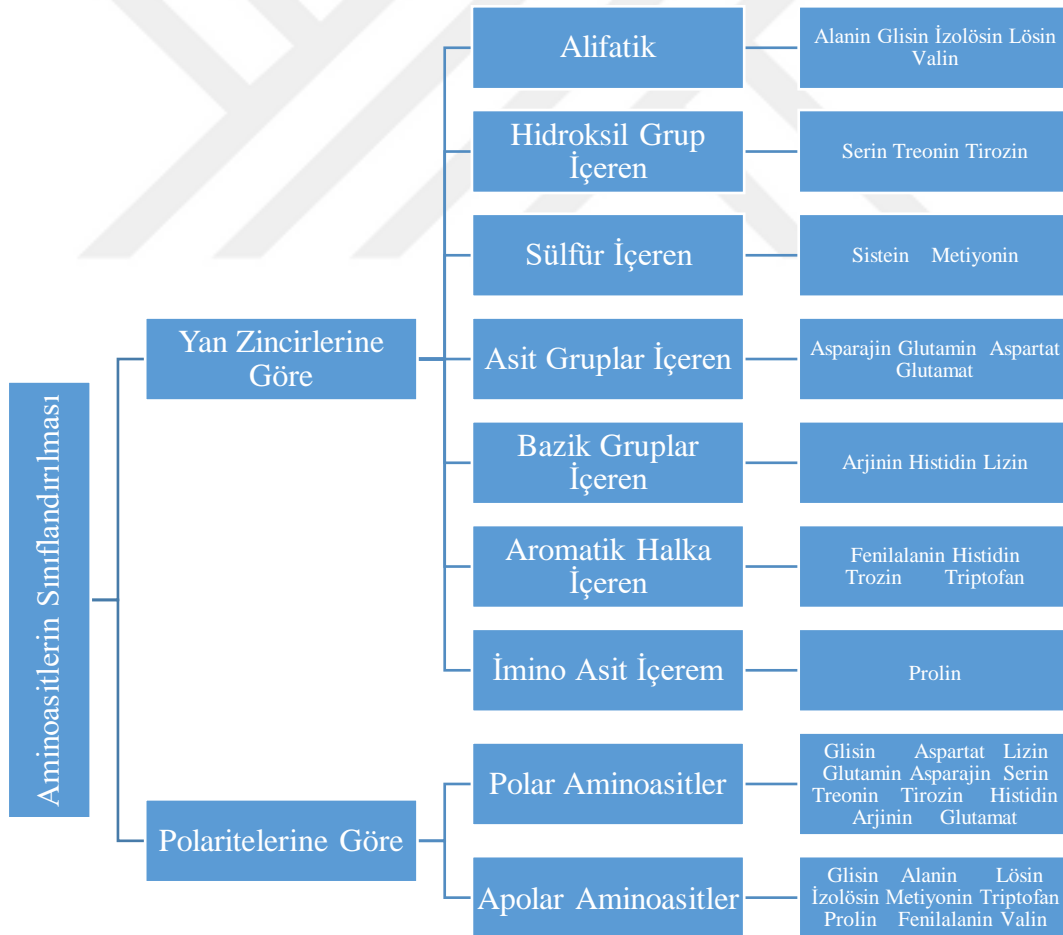
Protein sentezi sürecinde amino asitler bir amino asidin amino grubuyla diğer amino asidin karboksil grubuyla peptit bağı oluşturarak dehidrasyonla polipeptit zincirlerini oluştururlar. Bu zincirler omurga olarak bilinen Azot-Karbon-Karbon atomlarının tekrarlayan dizilerini sergilerler.

Bazı amino asitler farklı özelliklere sahiptirler. Örneğin; Prolin omurganın esnekliğini limitleyen polipeptit zincirlerinde geri dönüşler oluştururlar, Glisin (Şekil 3.3) boyutundan dolayı diğer amino asitlerden daha fazla esnekliğe izin vermektedir (Burkowski ve ark. 2008).



Şekil 3.2. En basit amino asit: Glisin (Carneiro 2019)'dan modifiye edilmiştir

Proteinin omurga zincirinde sırayla görülen amino asitlerin doğrusal sekansının tanımı proteinin birincil yapısı olarak bilinir.

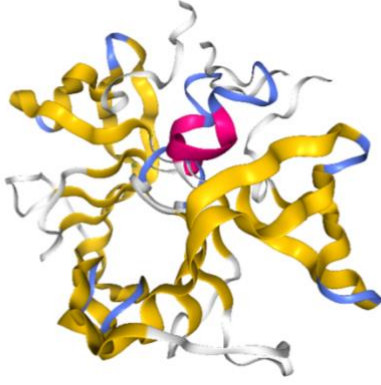


Şekil 3.3. Amino asitlerin sınıflandırılması

3.1.2. İkincil Yapı

Amino asitler enerjik olarak uyumlu ve makul derecede düzgün yapı konfigürasyonları oluşturarak, kendi üzerlerinde katlanarak polipeptit zincirleri oluştururlar. Bu yapılar omurga zincirindeki atomlar arasındaki hidrojen bağlarıyla tutulur ve oluşur.

Bu yapıların yaygın sınıflandırılması bu yapıları α -heliksler, β -tabakalılar ve döngü (Ing. loop) olarak üç kategoriye ayırmaktadır. Bu üç ikincil yapının da görülebileceği protein örneği Şekil 3.4'de gösterilmiştir. Bu üç yapı proteinlerin büyük çoğunluğunda mevcuttur. Çoğunlukla protein yapılar α -heliksler ve β -tabakalılar döngülerle bağlantılı olacak şekilde bir hidrofobik merkez oluştururlar (Burkowski ve ark. 2008). Bu düzenli yapıların karakterizasyonu proteinlerin ikincil yapıları olarak bilinir.



Şekil 3.4. İkincil yapıları renklendirilmiş 1zp8 PDB kodlu kompleks (Rose ve ark. 2018)

3.1.3. Üçüncül Yapı

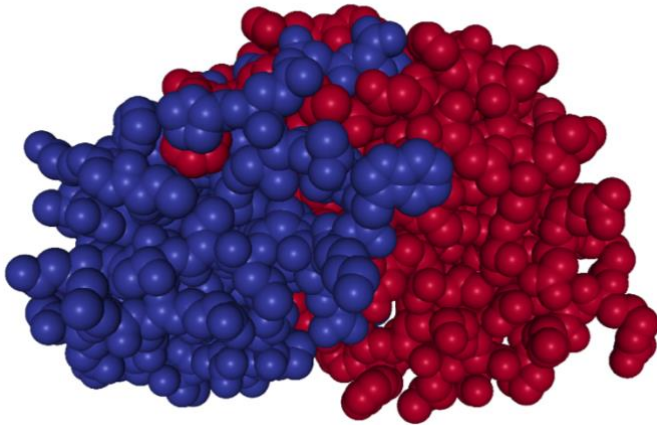
Proteinler sentezlendiği gibi kimyasal ve sterik tercihlerine göre çözücüyle ve birbirleriyle etkileşim kuran bütün farklı rezidüel katlanma sürecine girerler. Bu etkileşimler atomların çok sayıda geçici uzaysal düzenlenmeleriyle sonuçlanır. Sonuçta bazı biyolojik fonksiyonları sürdürülme yeteneğine sahip yarı-stabil üç boyutlu konfigürasyonların oluşumuna öncülük eden diğer katlanmalardan enerjik olarak daha kararlı olan özel bir katlanma modeli olan katlanmış yapı (Ing. folded structure) oluşur. Bütün atomların düzenlenmesi ve pozisyonlarının tanımı üçüncül yapı olarak bilinir.

3.1.4. Dördüncül Yapı

Dördüncül yapı protein yapısının zorunlu olmayan tek seviyesidir. Birçok polipeptit zinciri birleştiği zamanlarda bir protein dördüncül yapıya sahip olur (Şekil 3.5). Bu yapılar kompleksler olarak bilinirler.

Proteinlerin büyük çoğunluğu bu kategoride de bulunur çünkü fonksiyonlarını sergileyebilmeleri için diğer proteinlerle etkileşime girerler. Bu etkileşimler “arayüz” olarak bilinen protein yüzey yamalarıyla (“patch”) oluşur. Proteinler arasındaki etkileşimler zorunlu ve geçici olmak üzere iki gruba ayrılırlar. Zorunlu etkileşimler proteinlerin fonksiyonlarını ve yapılarını sürdürdürebilmeleri için kompleks durumunda olmaları gerektiği anlamına gelmektedir. Diğer yandan geçici etkileşimlerde birçok farklı moleküller etkileşim kurabilirler (Ansari ve Helms 2005; De ve ark. 2005).

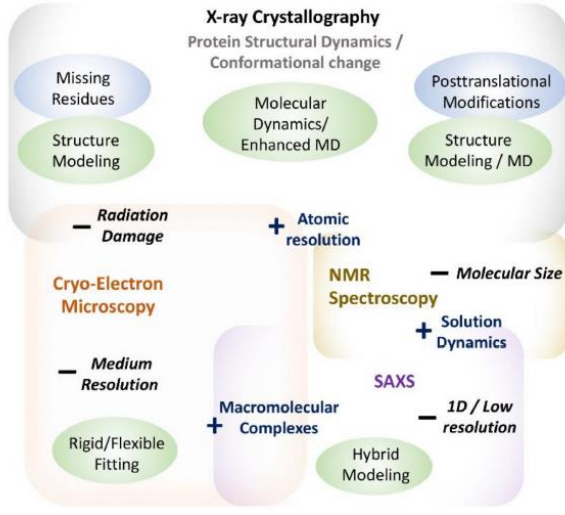
Bu iki tip etkileşimin ayrımı konuyla alakalıdır çünkü her iki tip arayüz alan, yapı ve rezidü kompozisyonu gibi birçok farklılık sergilemektedirler. Örneğin polar ve yüklü rezidü dağılımı daha fazla olan (Ansari ve Helms 2005) geçici etkileşimlere karşılık, zorunlu etkileşimler daha büyük ve daha hidrofobik olarak karakterize edilirler (De ve ark. 2005) Hidrofobiklik protein yapılarının stabilitesinde en temel faktördür. Bu rezidü kompozisyonundaki farklılıkları açıklamaya yardım edebilir çünkü geçici etkileşimler hızlı bir şekilde birleşmek ve ayrılmak zorundadırlar (Branden ve Tooze 2012).



Şekil 3.5. 2o4p PDB kodlu homomonomerin yapısı (Rose ve ark. 2018)

Arayüz yamalarındaki rezidüler spesifik eşleşme tercihlerine sahiptir. Hidrofobik rezidüler hidrofobik rezidülerle eşleşme eğilimindeyken yüklü rezidüler tamamlayıcı yüklü rezidüleri tercih ederler (Burkowski ve ark. 2008; Glaser 2001; Madaoui ve Guerois 2008).

Protein arayüz tahmini için kullanılan yöntemler *in vitro* (laboratuvar ortamında), *in vivo* (canlı üzerinde) ve *in silico* (bilgisayar ortamında)'dur. Her yöntemin maliyet, doğrulama ve zaman gibi kendine has avantajları ve dezavantajları bulunmaktadır (Şekil 3.6).



Şekil 3.6. Protein yapısını anlamadaki çeşitli deneysel metotların güçlü yanlarının ve limitleri (Srivastava ve ark. 2018)

3.2. *In Vitro* Yöntemler

In vitro çalışmaları proteinlerin normalde bulunduğu canlı organizmadan izole edilerek laboratuvar ortamında biyolojik yapılar içinde gerçekleştirilir. Bu izolasyon sayesinde daha detaylı analizler yapılabilir. *In vitro* çalışmalar organizmanın ideal hücresel ortamını yaratamadığından elde edilen sonuçlar doğru bir şekilde ortaya konulmamış olabilir. Bu çalışmaların doğruluğunun saptanabilmesi için *in vivo* çalışmaların yapılması gerekmektedir. X-ray kristallografi, SAXS, NMR ve Elektron mikroskopu *in vitro* yöntemleridir. Bu yöntemlerle proteinlerin etkileştiği arayüzler bulunur. Fakat bu metotlar pahalı ve çok meşakkatlidir (Atanasov ve ark. 2015; Lipinski ve Hopkins 2004; Smith 2001; Relman 1998).

3.2.1. X-ray Kristallografisi

X-ray kristallografisi interaktomları belirlemek için kullanılan temel araçtır. Bu araçla protein komplekslerinin atomik çözünürlükte görselleştirilmesini yapar. Böylece protein fonksiyonunun anlaşılmasını sağlar. Enzimlerin geçirdikleri değişimler, proteinlerin diğer moleküllerle etkileşimlerinin izlenebilmesine olanak sağlar. Protein-protein etkileşimlerinin bir kristal yapıdan ayırt edilmesi kolay olmadığından bir dizi hesaplama yapılması gerekmektedir.

Bu da hata oranını yükseltmektedir (Kobe ve ark. 2001; Kobe ve ark. 2008; Urakubo ve ark. 2008).

3.2.2. SAXS (Small-Angle X-ray Scatering)

Küçük-Açı X-Işını Saçılma (SAXS), yapı analizinde kullanılan X-ışını yöntemlerinden birisidir. Bu yöntemde incelenen örnekten küçük açıyla saçılan X-ışınının şiddetleri, saçılma vektörünün büyüklüğüne göre elde edilir ve bu verilerden yapı ile ilgili bilgiye ulaşılır. Saçılma yapıdaki elektronlar tarafından gerçekleştirildiğinden, yöntem elektron yoğunluğu farkının algılanması temeline dayanmaktadır. SAXS, polimer oluşumların içyapısını, şeklini, büyüklüklerini ve dağılımlarını belirlemek için kolloid biliminde güçlü bir yöntem olduğundan, gelişmiş güzel yönelimlere sahip, çözelti içindeki lateksler, bir araya toplanan miseller, proteinler ve sentetik polimerlerin de bu yöntem ile kolayca incelenebilmesine olanak sağlamıştır. (Altundag 2012; Rambo ve Tainer 2013). Hesaplamalı yöntemler yardımıyla, SAXS ile elde edilen protein yapılarının yüksek çözünürlüklü modelleri oluşturulabilmiştir (Jiménez-García ve ark. 2015; Schindler ve ark. 2016)

3.2.3. NMR (Nükleer Manyetik Rezonans)

NMR spektroskopisinin temeli, kuvvetli bir manyetik alan içerisine yerleştirilen bir molekülde bulunan manyetik olarak aktif çekirdeklerin radyo frekansındaki elektromanyetik ışınları absorbe etmesi üzerine kurulmuş olmasına dayanmaktadır. NMR, molekülün iskeleti hakkında bilgi verir. NMR, biyolojik olarak ilişkili biyomoleküler kompleksleri doğru bir şekilde tanımlayabilir (Erten ve ark. 2009; Pazos ve ark. 2008; Grigoriev 2001). NMR tipik olarak 30 kDa'dan daha küçük proteinler için uygundur, fakat özellikle membrandaki etkileşim yüzeylerinin tanımlanmasında önemli rol oynamaktadır. NMR Spektroskopisi molekülü parçalamadığından örnekler tekrar kullanılabilir. X-ışını kristalografisi ve nükleer manyetik rezonans (NMR) spektroskopisi, ayrıntılı bir yapısal bilgiye sahip olmak için yaygın olarak kullanılan deneysel yöntemlerdir.

3.2.4. Elektron Mikroskobu (EM)

Genellikle geçici, büyük çoklu protein sinyal aktarma merkezlerinin anlaşılması bunların kristalize olmamalarından dolayı zor olmaktadır. Bundan dolayı kristalografi ile birlikte kullanılacak diğer metotlara ihtiyaç duyulmaktadır. Bu tamamlayıcı metotlardan en

temel olanı EM'dur. EM'ları 500 kDa'dan büyük yapılar için uygundur ve moleküler-altı (5-15Å çözünürlüklü) haritalama sağlayabilirler (Lander ve ark. 2012). EM'daki gelişmeler, oldukça simetrik olan virüslerin atomik modellerini direkt olarak EM'larında yapılmasına olanak sağlamıştır (Grigorieff ve Harrison 2011). EM'daki son gelişmeler moleküler çözünürlükte protein komplekslerinin direkt görüntülenebilmesi fırsatını vermiştir.

3.3. Protein-Protein Etkileşim Çalışmalarında Kullanılacak Yapıların Belirlenmesi

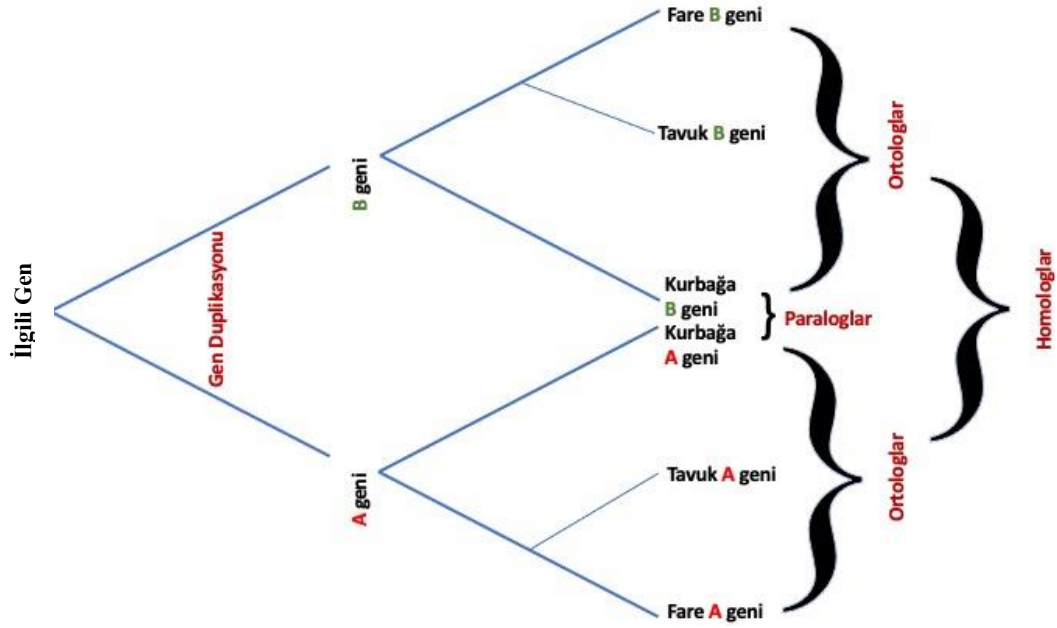
Bu kısımda; protein-protein kompleks yapılarının analizinde önemli yeri olan protein homolojisi, paralog-ortolog genler, sekans aynılığı konularına değinilmiştir

3.3.1. Protein Homolojisi

İki proteinin sekanslarının amino asitleri %30 ve üzeri oranda aynı dizilime sahipse, aynı protein ailesinin üyeleri olarak kabul edilirler. Sekans aynılığı %15'e kadar düşse bile, işlevleri ve üç boyutlu yapıları oldukça benzerse bu proteinler yine aynı protein ailesi üyesi sayılabilirler. Örneğin globinler aynı işlevi sergiledikleri, oldukça benzer yapılara ve sekans aynılığına sahip olduklarından protein ailesidirler. Protein domeyni ("protein domain") "bir protein dizisinde ya da yapısında özgün ve kararlı bir üçüncül yapısı olup bağımsız olarak evrimlenebilen işlevsel bölge" olarak tanımlanır (TÜBA, 2013). Domeyn ise ya kendi kendine evrim geçirebilen, işlevsel olabilen, katlanabilen ya da bir zincirin komşu kısımlarından kompaktlığı sayesinde ayrılmasıyla tanımlanabilen bir polipeptit zinciri parçasıdır. Domeyn protein işlevinin en küçük birimidir. Sıklıkla, her bir domeyn bir proteinde, liganda bağlanmak, aktif bölge içermek (enzimlerdeki katalitik domeynler), bir nükleotide bağlanmak (transkripsiyon faktörlerinde DNA/RNA bağlanma domeynleri) ya da diğer proteinlere bağlanmak (substrat bağlama domeynleri) için yüzey oluşturma etkisi gibi modüler bir işleve sahiptir. Tek bir protein zinciri bir, ya da daha fazla sayıda, çoğunlukla ise birkaç tane domeyn içerebilir. Evrimsel süreç sonucunda aynı domeyn farklı proteinlerde de görülebilir (Bahar ve ark. 2017). Globin protein ailesi, örneğin, bütün miyogloblin ve hemogloblinleri içerir diğer taraftan globin domeyni bir miyoglobindeki tek domeyne, bir hemoglobindeki dört domeyne işaret etmektedir.

3.3.2. Paralog ve Ortolog Genler

Farklı genomlar arasındaki ilişki hem ortolog hem de paralogları içeren homolog ailelerin sistemi olarak ifade edilir. Ortologlar türleşmede ortak bir atadan evrimleşen farklı türlerdeki genlerdir. Paraloglar ise bir genom içindeki duplikasyonlarla ilişkili genlerdir. Normal olarak ortologlar evrimleşme sırasında aynı işlevlerini sürdürürken, paraloglar yeni işlevlere evrilirler (Şekil 3. 7). Bundan dolayı yeni sekanslanan genomlarda gen işlevlerinin güvenilir bir şekilde tahmini için ortologların tanımlanması kritiktir. Bu filogenetik analizler için eşit derece önemlidir çünkü yorumlanabilir filogenetik ağaçlar sadece ortolog setleri dahilinde inşa edilir. Ortologların tam bir listesinin oluşmuş olması genom organizasyonunun herhangi bir anlamlı kıyaslanması için ön şarttır. Kabaca ifade edilirse, bir genomda bulunan bir gen, en yüksek sekans benzerliği olan diğer genomdaki genle ortologtur (Bahar ve ark. 2017).



Şekil 3.7. Homolog Genler

3.3.3. Sekans Aynılığı

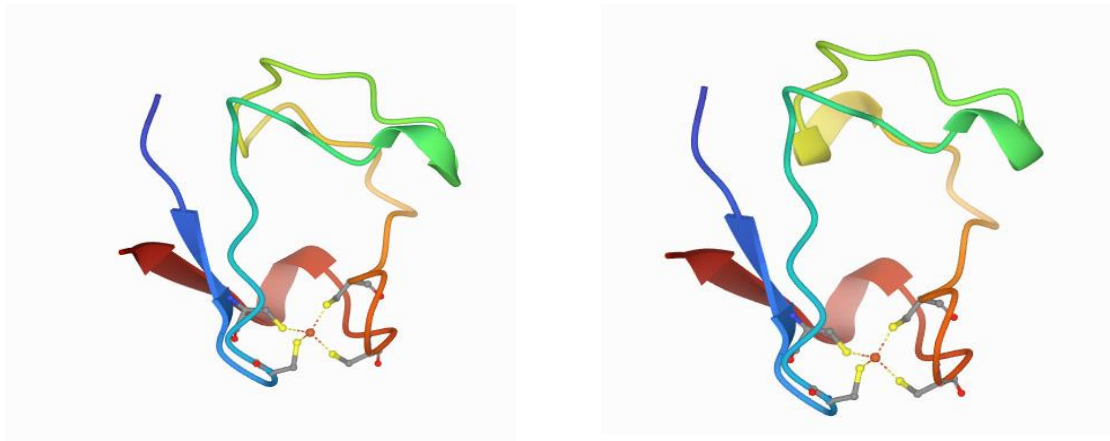
Protein sekans aynılığı bu sekans dizilerinin ne kadarının değişmez, aynı, olduğunu belirtmektedir. Sekans aynılığı aşağıdaki gibi belirlenir:

$$\%Id = 100 \times \frac{N_{ayni}}{\max(L_{tekil}, L_{kompleks})}$$

N_{ayni} , BLAST sonuçlarına göre tekil protein halinde ve kompleks halinde iken onun yakın homoloğu ile aynı olan amino asit rezidü sayısını, L_{tekil} , tekil proteinin uzunluğu (rezidülerin toplam sayısı) ve $L_{kompleks}$, kompleks homolog proteinin uzunluğunu ifade etmektedir. ComSin veri tabanında (Lobanov ve ark. 2010) tekil ve kompleks halindeki zincirler arasındaki sekans aynılığı bilgisi de mevcuttur. Bu benzerlik %0 ila %100 arasında belirtilmiştir. Aşağıda sekans aynılığı %0-50-100 olan bazı proteinlerin tekil ve kompleks halindeki yapıları hakkında FASTA formatlı amino asit sekans dizileri ve 3B'lu yapı görüntüleri verilmiştir (Şekil 3.8-10).

Sekans aynılığı %100 olan tekil haldeki 1RDV ve kompleks halindeki 2RDV PDB kodlu Rubredoksin proteinlerine ait formatındaki amino asit sekans dizileri (Şekil 3.8)

```
>1RDV_1|Chain A|RUBREDOXIN|Desulfovibrio vulgaris str. 'Miyazaki F' (883)
MKKYVCTVCGY EYDPAEGDPDNGVKPGTAFEDVPADWVCPICGAPKSEFEPA
>2RDV_1|Chains A,B,C|RUBREDOXIN|Desulfovibrio vulgaris str. 'Miyazaki F' (883)
MKKYVCTVCGY EYDPAEGDPDNGVKPGTAFEDVPADWVCPICGAPKSEFEPA
```



1RDV PDB kodlu Tekil Haldeki Rubredoksin Proteini (Sehnal ve ark. 2021)

2RDV PDB kodlu Kompleks Halindeki Rubredoksin Proteini (Sehnal ve ark. 2021)

Şekil 3.8. Rubredoksin Proteini

$$\%Id = 100x \frac{N_{ayni}}{\max(L_{tekil}, L_{kompleks})} = 100x \frac{52}{\max(52,52)} = 100x \frac{52}{52} = \%100$$

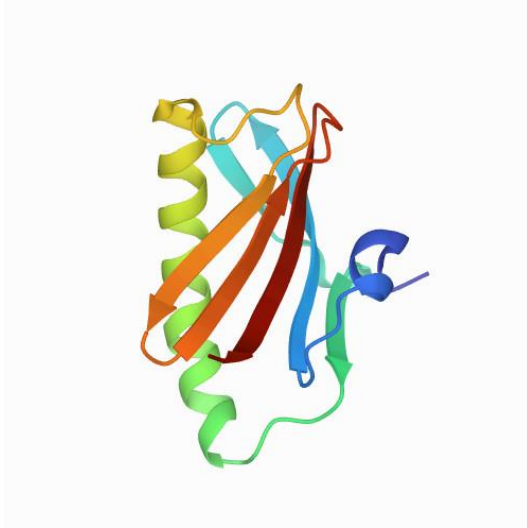
Sekans aynılığı %50 olan tekil haldeki 2EA9 ve kompleks halindeki 2H28 PDB kodlu Hipotetik proteinlerine ait FASTA formatındaki amino asit sekans dizileri (Şekil 3.9)

>2EA9_1|Chain A|Hypothetical protein yfjZ|*Escherichia coli* (562)

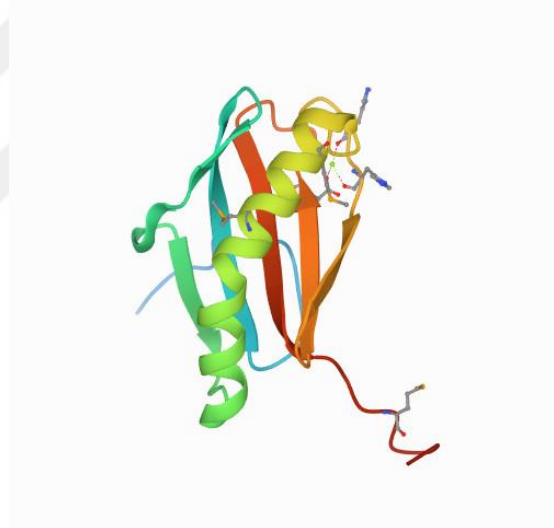
MSNTTWGLQRDITPRLGARLVQEGNQLHYLADRASITGKFSDAECPKLDVVFPHFIS
QIESMLTTGELNPRHAQCVTLYHNGFTCEADTLGSCGYVYIAVYPTQR

>2H28_1|Chains A,B|Hypothetical protein yeeU|*Escherichia coli* (562)

MSDTLPGTTLPDDNHDRPWWGLPCTVTPCFGARLVQEGNRLHYLADRAGIRGLFSD
ADAYHLDQAFPLMKQLELMLTSGELNPRHQHTVTLYAKGLTCKADTLSSCDYVYL
AVYPTPEMKNLEHHHHHHH



2EA9 PDB kodlu Tekil Haldeki Hipotetik Proteini (Sehnal ve ark. 2021)



2H28 PDB kodlu Kompleks Haldeki Hipotetik Proteini (Sehnal ve ark. 2021)

Şekil 3.9. Hipotetik Protein

$$\%Id = 100x \frac{N_{ayni}}{\max(L_{tekil}, L_{kompleks})} = 100x \frac{65}{\max(105,130)} = 100x \frac{65}{130} = \%50$$

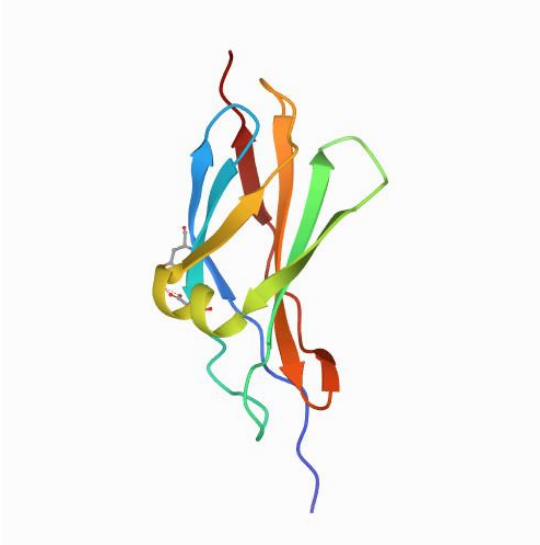
Sekans aynılığı %0 olan tekil haldeki 2HAZ PDB kodlu Nöral hücre adhezyon ve kompleks halindeki 1EBA PDB kodlu Eritropoietin reseptör proteinlerine ait FASTA formatındaki amino asit sekans dizileri (Şekil 3.10)

>2HAZ_1|Chain A|Neural cell adhesion molecule 1|Homo sapiens (9606)

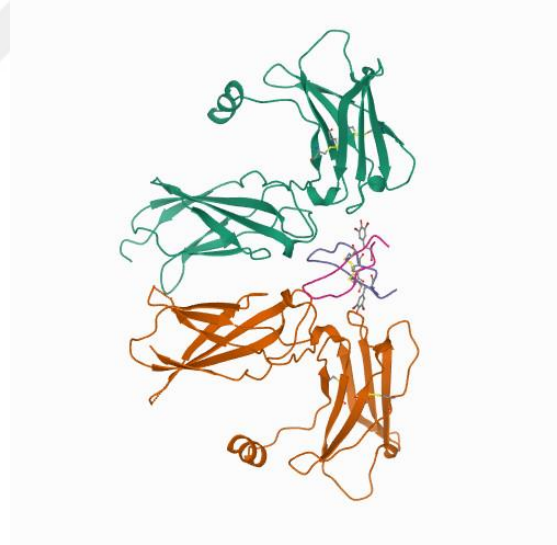
GSHMDTPSSPSIDQVEPYSSTAQVQFDEPEATGGVPILKYKAEWRAVGEEVWHSKW
YDAKEASMEGIVTIVGLKPETTYAVRLAALNGKGLGEISAASEFKTQPV

>1EBA_1|Chains A,B|Protein (Erythropoietin Receptor)|Homo sapiens (9606)

KFESKAALLAARGPEELLCFTERLEDLVCFWEEAASAGVPGNYSFSYQLEDEPWKL
CRLHQAPTARGAVRFWCSLPTADTSSFVPLELRVTAASGAPRYHRVIHINEVVLLDAP
VGLVARLADESGHVVLRLWLPPEPMTSHIRYEVDVSAGNGAGSVQRVEILEGRTEC
VLSNLRGRTRYTFVRARMAEPSFGGFWSAWSEPVSLLTPSDL



2HAZ PDB kodlu Tekil Haldeki
Nöral Hücre Adhezyon Proteinini
(Sehnal ve ark. 2021)



1EBA PDB kodlu Kompleks Haldeki
Eritropoietin Reseptör Proteinini
(Sehnal ve ark. 2021)

Şekil 3.10. Hipotetik Proteinini

$$\%Id = 100x \frac{N_{ayni}}{\max(L_{tekil}, L_{kompleks})} = 100x \frac{0}{\max(105, 215)} = 100x \frac{0}{215} = \%0$$

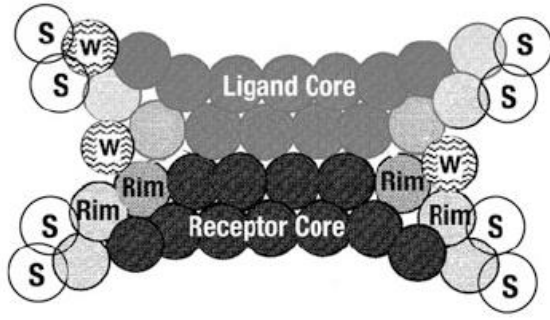
3.4. İkili Protein Komplekslerinin Sınıflandırılması

Etkileşim halinde olan domeyn-çiftleri tersinir ya da tersinmez olarak sınıflandırılırlar. Proteozom ya da ribozom gibi büyük protein kompleksleri yüksek afiniteye sahiptirler ve hasar görene ya da hücre kendilerini kullanmayınca kadar etkileşim halinde kalmayı sürdürdüklerinden tersinmez kompleksler olarak sınıflandırılırlar.

Diğer taraftan tersinir kompleksler belirli bir afinite aralığına ve farklı yaşam ömürlerine sahiptirler. Bazı kompleksler zayıfça bir araya gelirler ve geçici olarak birleşirler. Diğer tersinir sinyal kompleksleri sadece kısıtlı bir süre için daha yüksek afiniteye sahiptirler. Fakat er ya da geç, ayrılacaklardır (Hunter 2012). Bazı kompleksler güçlü afiniteye sahip olsalar bile kendi yapılarına bağlı olan sayaçları vardır. Sayaç bittiğinde ayrışırlar (Wolfenson ve ark. 2013). Antikor-antijen ve proteaz-inhibe edici komplekslerin yaşam süreleri saatlerden günlere kadar değişebilir. Bunlardan bazıları biyolojik sürecin zamanına göre tersinmez olarak da sınıflandırılabilirler (Rawlings ve ark. 2004).

3.5. Domeyn-Domeyn Arayüzlerinin Yapısı

Birçok domeyn-domeyn arayüzü standart yama'yı tanımlayan tipik bir tasarıma sahiptir. Merkezi; çözücü-içermeyen "kor", su-aracılı etkileşimleri içeren kısmi olarak gömülmüş "çember" tarafından çevrilmiştir. Kor, ortalama, gömülü yüzey alanının %75'ini ve bağlanma enerjisinin büyük kısmını oluşturur. Kor genel olarak yüzeyden daha hidrofobik fakat proteinin iç kısmından daha az hidrofobiktir. Çember protein yüzeyine benzer bir içeriğe sahiptir. Çember gömülü yüzeyin %25'ini oluşturur. Temel görevinin kordan su moleküllerinin uzak tutulmasını sağlama olduğu düşünülmektedir (Bogan ve Thorn, 1998). Şekil ve yük açısından iki molekülün birbirini tamamlayıcı (komplementer) oluşu, paketlenme yoğunluğunun proteinin iç kısmından farklı olmaması için önemlidir. Birçok arayüz oldukça düzdür. Arayüz yama çeşitleri büyüklük bakımından farklılık gösterirler.



Şekil 3.11. Standart Yama ve Arayüz Büyüklüğü (Meyerkord ve Fu 2015)

3.6. Arayüzlerde Bulunan Rezidüer

Rezidü kompozisyonunun farklı tip komplekslerde, örneğin tersinir-tersinmez kompleksler arasında ve homodimerik-heterodimerik kompleksler arasında, sistematik şekilde farklı olduğunun farkına varılmıştır. Hesaplamalı çalışmalar sayesinde arayüzleri tanımlamanın en az üç farklı yolu olduğu bulunmuştur.

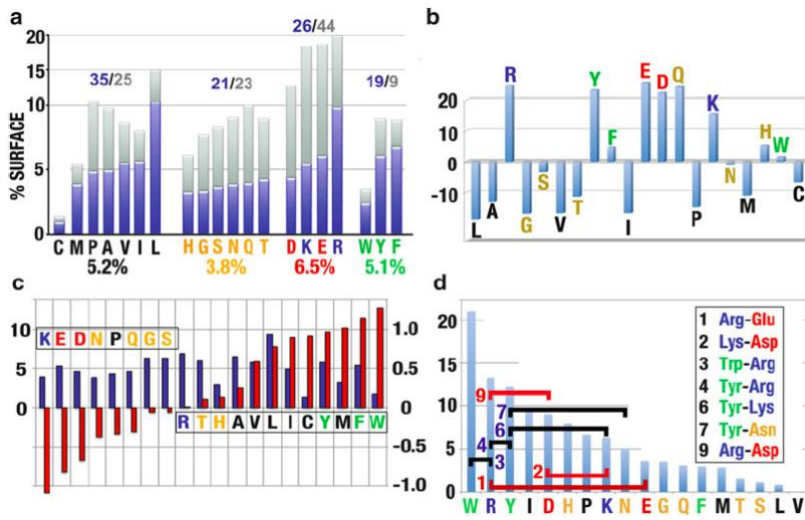
3.6.1. Rezidü Sayımı

Bir arayüzü tanımlamanın en kolay yolu belirlenen tipteki rezidü sayısını saymaktır ve onları toplamın yüzdesi olarak ifade etmektir. Bazen rezidü sayımında rezidülerin büyüklükleri (gömülü yüzeye \AA^2 olarak alan katkıları) göz önüne alınarak hesaplama yapılır. İnsana, maya mantarına ve *E. coli*'ye ait protein-protein yüzeylerini içeren bir çalışmada bulunan bazı sonuçlar şunlardır; Leu genel olarak arayüzlerde en sık bulunan rezidü iken bunu takip eden Arg olmuştur. Yüklü rezidüer polar rezidüerden daha sık bulunmuşken, Arg ve His hariç, her ikisi de genel olarak arayüzde oldukça boldurlar. Aromatikler yüzeylerde daha düşük sıklıktayken, Trp hariç arayüzlerde oldukça çokturlar. Hidrofobik rezidüer arayüzlerde yüksek sıklıktayken, yüzeylerde oldukça az bulunurlar. Cys yüzey ve arayüzlerde oldukça nadirdir (Levy 2010).

Tersinir arayüzler tüm yüklü rezidüer, polar Gln ve aromatik rezidüer (özellikle Tyr) açısından daha zenginken, bütün hidrofobik ve küçük polar rezidüer bakımından yoksundurlar (Ofra ve Rost 2003).

3.6.2. Rezidü Çiftleri

Diğer metot arayüz boyunca etkileşen rezidülerin sıklığına odaklanmaktadır (Glaser ve ark. 2001). Tersinir hetero-dimerik komplekslerle yapılan bir çalışmada en çok bulunan on rezidü çifti Çizelge 3.2’de listelenmiştir. Yüklü-yüklü çiftlerden Arg-Glu ve Lys-Asp hem çemberde hem de korda oldukça çokken, Arg-Asp ve Lys-Glu çemberde çokken, korda azdır. Geriye kalan 6 çift rezidü, aromatik rezidüleri içermektedir. Bunların ikisi sadece hidrofobik yan-zincir içermekte ve sadece korda bulunurken, diğer dördü polar etkileşimlerdir (Şekil 3.12) (Headd ve ark. 2007).



Şekil 3.12. Arayüzde rezidü kullanımı (MeyerKord ve Fu 2015)

3.6.3. Hot Spot’lar

Arayüzlerin son tanımlayıcısı etkileşimin enerjeliğine odaklanır ve alanin mutasyon sonuçlarını temel alır. Arayüzdeki birçok mutasyon kısıtlı bağlanma etkisine sahiptir, “hot spots” olarak bilinen kimi kümeler büyük/baskın bağlanma etkisine sahiptir (Bogan ve Thorn 1998). Hot spots neredeyse her zaman korun merkezinin yakınlarında gömülüdür. Ortolog kompleksler arasındaki sekans korunumu hot spots’da en yüksektir (Keskin ve ark. 2005). Trp, Arg ve Tyr rezidüleri hot spots’da en yaygın olan rezidülerdir. Bu üç rezidü de hidrofobik, aromatik ve polar etkileşim kurma açısından oldukça çok yönlüdür. Arayüzdeki en yaygın rezidü olan Leu hot spots’da oldukça nadir bulunurken, Ile ise bol bulunur (Crowley ve Golovin 2005; Keskin ve ark.2005).

Çizelge 3.2. Tersinir hetero-dimerik komplekslerinde en çok bulunan 10 rezidü çifti

Korda bulunma Sıklığı	Arayüzde bulunma Sıklığı	Rezidü Çiftleri
1	1	Arg-Glu
2	3	Lys-Asp
3	9	Trp-Arg
4	5	Trp-Arg
5	-	Phe-Leu
6	7	Tyr-Lys
7	6	Tyr-Asn
8	-	Trp-Ile
9	2	Arg-Glu
-	4	Lys-Glu

4. PROTEİN ETKİLEŞİMLERİ İÇİN BİYOİNFORMATİK YAKLAŞIMLAR

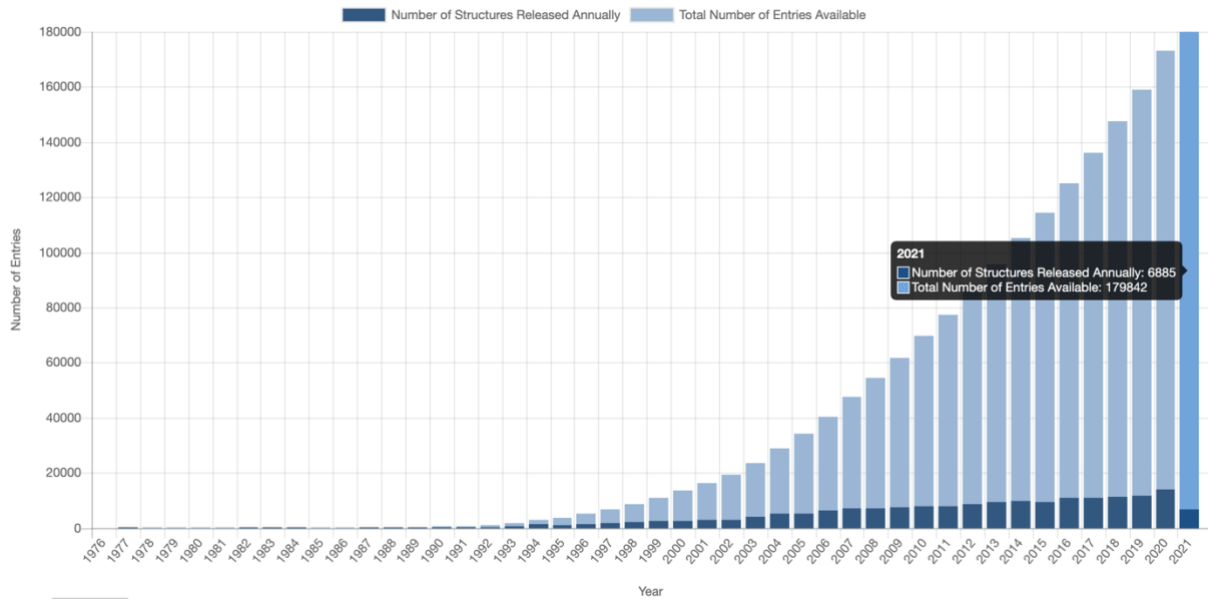
Hücredeki protein molekülleri arasındaki etkileşimler oldukça spesifikdir ve sadece seçilmiş moleküller birbirine bağlanır. Bu etkileşim yüzeylerinin geometrik uyumluluğu molekülün seçici bağlanması için temel gereksinimlerden birisidir. Protein-protein arayüzlerinin geometrik uyumluluğunu tahmin etmek için çok sayıda metot geliştirilmiştir

Bu bölümde veri tabanları olan Protein Data Bank ve ComSin hakkında bilgi verildikten sonra protein arayüz tahmini, teselasyon yöntemleri ve tetrahedronların geometrik karakterizasyonuna örnekler başlıklı konulara değinilmiştir.

4.1. Protein Veri Bankası

Protein Data Bank biyolojik yapıların dünya geneli bilgi havuzu olarak 1971'de kurulmuştur. Research Collaboratory for Structural Bioinformatics (RCSB) tarafından yönetilen büyük ve halka açık domeyn deposudur. PDB'de güncel olarak (Temmuz 2021 itibariyle) 179842 biyolojik yapı bulunmaktadır ve bunların yaklaşık %93'ü proteinlere aittir. Bu yapıların büyük çoğunluğu X-ray kristalografisi (%88) tarafından belirlenmiştir. NMR ile belirlenenlerin oranı %8 iken elektron mikroskopuyla belirlenenlerin oranı %4'tür (Şekil 4. 1) (Berman ve ark. 2000).

Bu veri tabanındaki her bir protein PDB ID olarak adlandırılan 4'lü kodla isimlendirilir. Yüksek kaliteli standart yürütülmesi çabalarına rağmen PDB'deki yapıların bazıları eksik atomlu ya da farklı çözünürlük seviyeleriyle hatalı verileri içermektedir.



Şekil 4.1. PDB’de yıllık olarak depolanan toplam protein yapıları (mavi) ve yıl başına yeni protein yapıları (lacivert)

4.1.1. PDB Formatı

Protein yapılarına ilişkin bilgiler PDB dosyalarında tutulur. PDB dosyaları bahsi geçen satırın gerçek içeriklerinin eşlik ettiği her bir satırda ne çeşit bilginin sağlandığını belirten bazı standart ana sözcükleri içeren çoklu satırlara sahip yazı dosyalarıdır.

ATOM	288	N	ALA	A	56	-4.088	40.315	75.585	1.00	6.08	N
ATOM	289	CA	ALA	A	56	-5.339	41.053	75.670	1.00	9.34	C
ATOM	290	C	ALA	A	56	-5.906	41.161	77.082	1.00	7.31	C
ATOM	291	O	ALA	A	56	-6.564	42.135	77.425	1.00	7.38	O
ATOM	292	CB	ALA	A	56	-6.359	40.663	74.598	1.00	7.86	C
ATOM	293	N	AHIS	A	57	-5.626	40.117	77.818	0.30	12.85	N
ATOM	294	N	BHIS	A	57	-5.691	40.105	77.871	0.70	16.57	N
ATOM	295	CA	AHIS	A	57	-5.987	39.854	79.197	0.30	13.41	C
ATOM	296	CA	BHIS	A	57	-6.156	39.958	79.263	0.70	15.45	C
ATOM	297	C	AHIS	A	57	-5.333	40.844	80.145	0.30	12.74	C
ATOM	298	C	BHIS	A	57	-5.378	41.006	80.034	0.70	17.07	C

Şekil 4.2. 1trm PDB’sine ait kompleks reseptör molekülünden PDB dosyasından bir kesit. Amino asitlere (yeşil renk) ait atomların (mor renk) x,y ve z koordinatlarını (kırmızı renk) gösterilmektedir.

Protein yapıları sıklıkla makalelerle yayımlandığından PDB dosyaları yazarların isimlerini, keşfedilme metodunu ve tarihini, literatür bilgilerini, biyolojik yapılarını vb. detaylandırır. PDB dosyası ayrıca molekülün bilinen tüm atomlarının listesini içerir. Atomlara ait; proteindeki atomun ardışık sayısı, kimyasal elementi, rezidü adı, zincir adı, rezidü sayısı,

3B'lu koordinatları, “occupancy” değeri ve sıcaklık faktörü bilgileri her satırda yazılır (Şekil 4. 2).

Proteinlerin PDB (Berman ve ark. 2000) formatlarında yer alan bilgilerden en önemlilerinden olan REMARK bölümünde bu proteinlere ait eksik rezidü, eksik atom gibi bilgiler yer almaktadır (Şekil 4. 3). Tetrahedronların hacmini bulmak için atom düzeyinde hesaplamalar yapıldığından kullanılan protein PDB dosyalarında bu bilgiler büyük önem

```
REMARK 465
REMARK 465 MISSING RESIDUES
REMARK 465 THE FOLLOWING RESIDUES WERE NOT LOCATED IN THE
REMARK 465 EXPERIMENT. (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN
REMARK 465 IDENTIFIER; SSEQ=SEQUENCE NUMBER; I=INSERTION CODE.)
REMARK 465
REMARK 465 M RES C SSEQI
REMARK 465 LEU A 299
REMARK 465 HIS A 300
REMARK 465 LEU A 301
REMARK 465 ASP A 302
REMARK 465 GLN A 303
REMARK 465 THR A 304
REMARK 465 PRO A 701
REMARK 465 LEU D 299
REMARK 465 HIS D 300
REMARK 465 LEU D 301
REMARK 465 ASP D 302
REMARK 465 GLN D 303
REMARK 465 PRO D 701

REMARK 470
REMARK 470 MISSING ATOM
REMARK 470 THE FOLLOWING RESIDUES HAVE MISSING ATOMS (M=MODEL NUMBER;
REMARK 470 RES=RESIDUE NAME; C=CHAIN IDENTIFIER; SSEQ=SEQUENCE NUMBER;
REMARK 470 I=INSERTION CODE):
REMARK 470 M RES CSSEQI ATOMS
REMARK 470 PRO A 305 CG CD
REMARK 470 ARG A 307 CG CD NE CE NH1 NH2
REMARK 470 GLU A 466 CG CD OE1 OE2
REMARK 470 THR D 304 OG1 CG2
REMARK 470 PRO D 305 CG CD
REMARK 470 SER D 306 OG
REMARK 470 ARG D 307 CG CD NE CE NH1 NH2
REMARK 470 GLU D 465 CG CD OE1 OE2
REMARK 470 GLU D 466 CG CD OE1 OE2
REMARK 470 ARG D 470 CG CD NE CE NH1 NH2
REMARK 470 HIS D 471 CG ND1 CD2 CE1 NE2
REMARK 470 ASN D 473 CG OD1 ND2
REMARK 470 ARG D 474 CG CD NE CE NH1 NH2
REMARK 470 LYS D 475 CG CD CE NZ
```

Şekil 4.3. 1hp2 PDB dosyasındaki eksik rezidüleri (solda) ve eksik atomları (sağda)

taşımaktadır.

PDB dosyalarında her bir atomun koordinatları deneysel olarak bulunmaktadır. Bazı atomlar her zaman aynı koordinatta bulunmayabilir. Atomların her zaman aynı koordinatlarda bulunup bulunmadığı PDB kodlarında “occupancy” değeri ile gösterilmektedir. “occupancy” değeri 1.0 olması o atomun her zaman o koordinatta bulunduğunu ifade ederken, aynı atoma ait farklı koordinatlar varsa bunlar 0.3, 0.5, 0.7 gibi “occupancy” değerleriyle ifade edilir (Şekil 4. 4). Bu değerler atomun o koordinatlarda görülme sıklığını ifade etmektedir. Değerin 1.0’dan farklı olması bu atomun en az 2 farklı yerde görüldüğünü belirtmektedir. Farklı yerlerdeki görülme sıklıkları toplamı her zaman 1.0 olmalıdır.

ATOM	288	N	ALA	A	56	-4.088	40.315	75.585	1.00	6.08	N
ATOM	289	CA	ALA	A	56	-5.339	41.053	75.670	1.00	9.34	C
ATOM	290	C	ALA	A	56	-5.906	41.161	77.082	1.00	7.31	C
ATOM	291	O	ALA	A	56	-6.564	42.135	77.425	1.00	7.38	O
ATOM	292	CB	ALA	A	56	-6.359	40.663	74.598	1.00	7.86	C
ATOM	293	N	AHIS	A	57	-5.626	40.117	77.818	0.30	12.85	N
ATOM	294	N	BHIS	A	57	-5.691	40.105	77.871	0.70	16.57	N
ATOM	295	CA	AHIS	A	57	-5.987	39.854	79.197	0.30	13.41	C
ATOM	296	CA	BHIS	A	57	-6.156	39.958	79.263	0.70	15.45	C
ATOM	297	C	AHIS	A	57	-5.333	40.844	80.145	0.30	12.74	C
ATOM	298	C	BHIS	A	57	-5.378	41.006	80.034	0.70	17.07	C

Şekil 4.4. 1trm PDB kodlu proteine ait “Occupancy” değerleri

4.2. ComSin Veri Tabanı

Bu tezde kullanılacak proteinleri belirlemek için ComSin veri tabanında (“Database of protein structures in bound (Complex) and unbound (Single) states in relation to their intrinsic disorder: COMplexSINgle”) bulunan verilerden yola çıkılmıştır. ComSin veri tabanı protein yapılarının bağlı (kompleks) ve bağımsız (tekil) yapılarının PDB kodlarını içermektedir (Çizelge 4.1). Bu veri tabanının kullanım alanı özünde düzensiz protein bölgelerinin (“intrinsically disordered protein regions”) ve bunların komplekslerinin araştırılmasıyla ilişkili çalışmalarla sınırlı değildir. ComSin proteinlerin bağlı ve bağımsız durumlardaki yapısal farklarının analiz edilmesinde, örneğin protein bağlanmasıyla indüklenen değişimleri keşfetmede ve diğer amaçlar için de kullanılabilir. Bu veri tabanında 1035 protein ailesi ve 24.910 protein-protein etkileşimi bulunmaktadır.

4.2.1. ComSin’in Oluşturulmasında Kullanılan Veri Tabanları

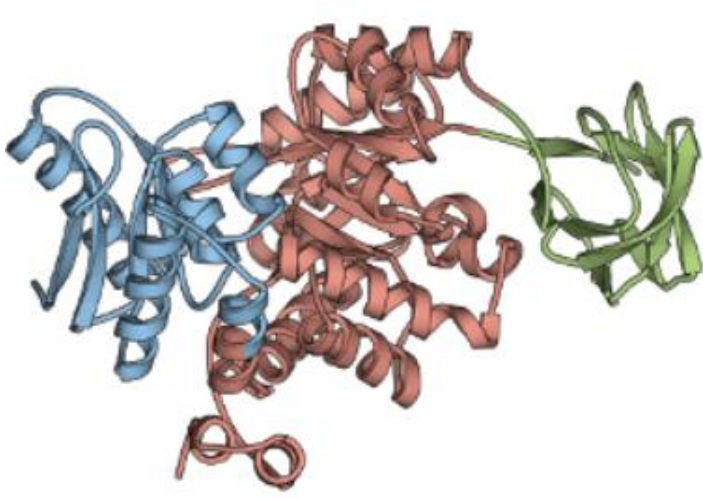
ComSin veri tabanı oluşturulurken elde edilen proteinler farklı veri tabanlarından alınmıştır. Bunlar arasında CDD, Pfam, SMART ve COG veri tabanları bulunmaktadır.

4.2.1.1. CDD

ComSin veri tabanı CDD (“Conserved Domain Database”, Korunmuş Domeyn Veri Tabanı) temel alınarak oluşturulmuştur. CDD protein bölgelerini çoklu sekans dizileme modelleri ile elde eder (Şekil 4.5). CDD, dış kaynaklı veri tabanlarından (Pfam, SMART ve COG) alınan domeyn modellerinin yanı sıra domeyn sınırlarını tanımlamak ve doğru hizalamaları sağlamak için 3B’lu yapı bilgilerinin kullanılmasıyla elde edilmiş domeynleri de içermektedir (Bahar ve ark. 2009).

4.2.1.2. Pfam

Pfam (“Protein Families Database”, Protein Aileleri Veri Tabanı) veri tabanı protein ailelerinin ve domeynlerinin kapsamlı bir bütünü içerir. Pfam’de her bir girdi altı farklı tipten birisiyle Aile, Domeyn, Motif, Tekrar ya da Düzensiz olarak etiketlenir. Bütün girdilerin %97,2’sini aile ve domeyn oluşturmaktadır. 2021 Mart’ta Pfam 34 versiyonu, 19.179 Pfam ailesi içermektedir (Mistry ve ark. 2020).



Şekil 4.5. Pirüvat kinaz; C-Terminal (mavi), N-Terminal (kırmızı) ve PK domeyn (yeşil) rezidülerinden oluşmuş üç domeyn içermektedir (Bahar ve ark. 2017)

4.2.1.3. SMART

SMART (“Simple Modular Architecture Research Tool”, Basit Modüler Yapı Arama Aracı) protein domeyn yapılarının analizinde ve protein domeynlerinin açıklanması ve tanımlanması için hazırlanmış internet erişimli bir yapı arama aracıdır (<http://smart.embl.de>). SMART’ın 9. versiyonu manuel olarak düzenlenmiş olan 1300’den fazla domeyn modeli içermektedir. SMART, Uniprot ve Ensembl protein veri tabanlarıyla senkronizedir (Letunic ve ark. 2020).

4.2.1.4. COG

COG (“Cluster of Ortholog Groups”, Ortolog Grup Kümeleri) ortolog grupların kümelerinin veri tabanına aktarılması ile meydana gelmiştir. SMART, Pfam ya da CDD gibi protein domeyn veri tabanlarının aksine COG veri tabanındaki girdilerin çoğunluğu bütün (“full-length”) proteinleri içermektedir (Galperin ve ark. 2021).

4.2.2. ComSin Veri İeriđi

ComSin veri tabanındaki kayıtlardan (“records”) bir kısmı ile izelge 4.1 hazırlanmıřtır. Burada bulunan bilgiler ile ComSin veri ieriđi ařađıda anlatılmıřtır. İlk stun gncel listedeki iftlerin numaralandırılmasını gstermektedir. Sonraki iki stun hem tekil hem de kompleksteki homolog proteinin ait olduđu CDD domeyn aileleri hakkında bilgi iermektedir. Drdnc stun tekil protein ile onun kompleksteki homologu arasındaki sekans aynılıđını gstermektedir (3.2.3. numaralı altblmde detaylı bir řekilde anlatılmıřtır). Sonraki 3 stun ise sadece tekil formdaki yapıyla alakalıdır. Bunlarda sırasıyla: zincirin PDB kodu, zincirin uzunluđu ve zincirdeki znde dzensiz rezidlerin sayısı hakkında bilgiler bulunmaktadır. Benzer řekilde sonraki 3 stun homolog yapının yer aldıđı kompleksin PDB kodu, kompleksteki homolog yapının zincir uzunluđu ve znde dzensiz rezidlerinin sayısı bilgilerini ierir. Sonraki 2 stun btn kompleksle ilgilidir. İlk stun kompleksin tipini (homo ya da hetero-oligomer) tanımlar.

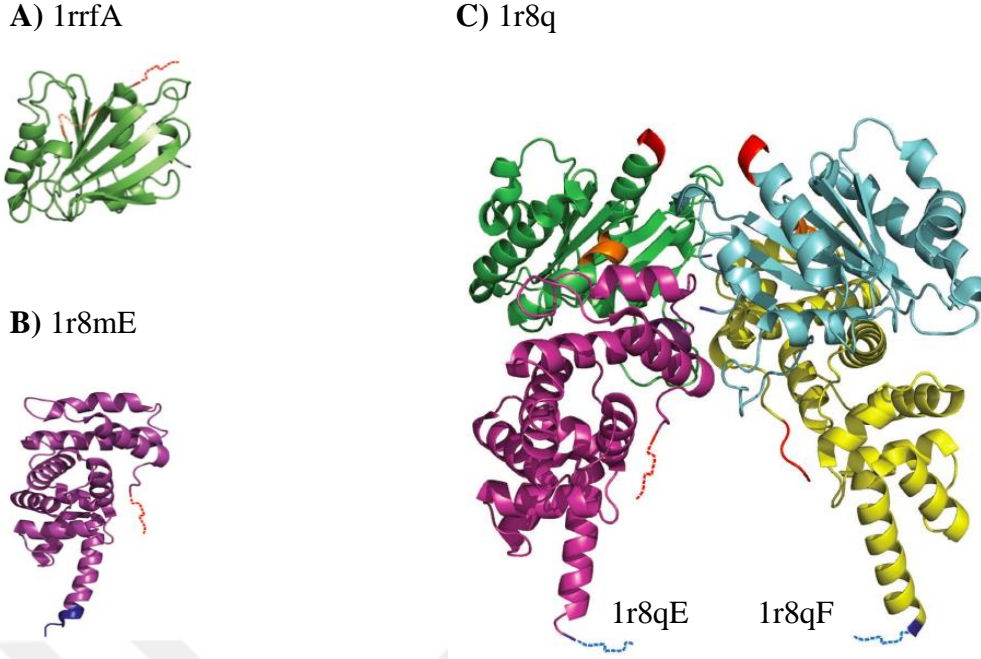
Protein-protein etkileřiminde etkileřen domeyn arayznn homolog proteinlerde korunması biyolojik olarak byk neme sahiptir. Korunmuř bađlanmanın (“Conserved Binding”) etkileřen domeyn ailelerinin farklı yeleri aynı řekilde kenetlendiđinde (“dock”) var olduđu belirtilir (Shoemaker ve ark. 2006). Bu bađlanmaların tm ile bilinmeyen bir bađlanma yzeyinin zelliklerini anlamak iin molekler modellemede kullanılabilen bir kenetleme řablonları ktphanesi elde edilmiřtir. Aynı zamanda, CBM’ler (“Conserved Binding Modes”) ile domeyn etkileřim ađları oluřturulmaktadır.

izelge 4.1’de, C stununda (+) bulunması kompleksin CBM dođrulamasının var olduđunu belirtir. P stunundaki (+) ise tekil ve kompleks durumlarındaki PISA (“Protein Interfaces Surfaces and Assemblies”) deđerlendirilmesinin yapıldıđını gsterir. PISA olası stabil komplekslerin tahminini ve verilen yapının analizini yapmak iin kullanılır.

Çizelge 4.1. ComSin veri tabanında bulunan proteinlere ait bilgiler

		Tekil					Kompleks					
		Protein Ailesi	%Id	PDB	L	ödrs	PDB	L	ödrs	tip	c	p
2918	36	cd00171	100	1r8mE	203	7	1r8qE	203	12	het	+	+
2919	36	cd00171	100	1r8mE	203	7	1r8qF	203	9	het	+	+
24196	995	smart00177	100	1rrfA	181	6	1r8qA	181	2	het	+	+
24197	995	smart00177	100	1rrfA	181	6	1r8qB	181	2	het	+	+

Bunu bir örnekle açıklamak gerekirse 2918-2919-24196 ve 24197 sayılarını içeren ilk sütun protein çiftlerinin ComSin'deki sıra numarasını vermektedir. İkinci ve üçüncü sütunlarda yer alan protein çiftlerinin ait olduğu CDD domeyn aileleri hakkında bilgi vermektedir. İkinci sütunda yer alan 36 ve 995 sayıları, cd00171 ve smart00177 (üçüncü sütun) protein ailesi kodlarının artan sıralama şeklinde numaralandırılması sonucunda elde edilmiş sayılardır. Dördüncü sütun ise tekil protein olan 1r8mE (Şekil 4.6A) ve 1rrfA (Şekil 4.6B) zincirlerinin homolog kompleksleri olan 1r8qA-B-E-F (Şekil 4.6C) ile sekans aynılığını göstermektedir. Buradaki aynılık 4 protein çifti için %100'dür. Sonraki 3 sütunda yer alan bilgiler tekil haldeki PDB kodu, zincir büyüklüğü ve özünde düzensiz rezidü sayısı hakkında bilgi vermektedir. Beşinci sütundaki tekil haldeki 1r8mE ve 1rrfA PDB kodlarını vermektedir. 1r8m PDB kodu "ARNO" ("The Arf Exchange Factor ARNO") proteini ifade ederken, 1rrf ise ARF'ı ("ADP-Ribosylation Factor-1") ifade etmektedir. Altıncı sütundaki 203-181 sayıları proteinlerin zincir uzunluğunu vermektedir. Yedinci sütunda ise bu tekil haldeki proteinlerde bulunan özünde düzensiz rezidülerin sayısı verilmektedir. 1r8mE'de 7, 1rrfA'da ise 6 tane özünde düzensiz rezidü bulunmaktadır. Sekizinci, dokuzuncu ve onuncu sütunlarda bilgiler ise kompleks haldeki homolog yapıyla ilgili PDB kodu, zincir uzunluğu ve özünde düzensiz rezidüler hakkındadır. On birinci sütun ise kompleksin tipinin homo- (hom) ya da hetero-oligomer (het) olup olmadığı hakkında bilgi vermektedir. Bu örneklerde kompleks tiplerinin hepsi hetero-oligomerdir. C sütunu kompleksin CBM doğrulamasını belirtirken, P sütunu tekil ve kompleks durumlarındaki PISA değerlendirilmesini belirtmektedir. Tablodaki örneklerin hem CBM hem de PISA değerlendirilmesi yapılmıştır.



Şekil 4.6. Tekil (A ve B; sırasıyla PDB kodları 1RRF ve 1R8M) ve kompleks (C; PDB kodu 1R8Q) yapıdaki ADP-ribozilasyon faktör (yeşil ve mavi zincirler) ve değişim faktörü ARNO (mavi ve sarı zincirler). Kesik çizgiler özünde düzensiz rezidüleri ifade etmektedir.

4.3. Protein Arayüz Tahmininde Hesaplamalı Yaklaşımlar

Proteinin fonksiyonun belirlenmesi için proteinin çeşitli ölçekte tüm organizasyon ve işlevlerinin örneğin amino asit dizilimi, katlanmaları, protein yüzeyi ve protein arayüzünün tümünün rolü vardır. Proteinin amino asit dizilimi onun nasıl katlanacağı hakkında bilgiye sahiptir. Protein katlanmaları sonucunda benzer katlanmaya sahip olan proteinlerden bazılarının farklı işleve sahip olabildikleri ya da aksine farklı katlanmaya sahip olanların aynı işlevlere sahip oldukları bilinmektedir. Ayrıca, protein arayüzünün, protein yüzeyine göre daha korunmuş olduğu bilinmektedir.

Proteinlerin nasıl etkileştiğini belirlemek için hesaplamalı yaklaşımlar etkileşim arayüzlerinin tahminine veya protein yapılarının kenetlenme modellerine yer vermektedir. Etkileşim arayüzleri hakkında bilgi sahip olmanın etkileşen proteinlerin konformasyon tahminini iyi bir şekilde geliştirebileceği gösterilmiştir. Protein protein arayüz tahmini birçok farklı yaklaşıma sahiptir. Bu yaklaşımlar yapının ya da dizinin yapı-içi özelliklerini, evrimsel ilişkilerini ya da referans şablon olarak var olan kompleksleri kullanabilirler.

Arayüz tahminleri yapı-içi ve kalıp temelli yaklaşımlar olarak iki gruba bölünebilir (Esmailbeiki ve ark. 2016).

4.3.1. Yapı-içi Temelli Yaklaşım

Yapı-içi temelli yaklaşımı protein yapısı ya da protein dizisi içindeki özelliklere odaklanır. Yapı-içi temelli yaklaşımlar tahmin girdisi olarak ya sekans bilgilerini ya da yapısal bilgileri kullanır.

4.3.1.1. Sekans Temelli Yaklaşım

Sekans temelli arayüz tahminlerinde arayüzleri belirlemek için sorgulanan (“query”) proteinlerin sadece sekans bilgileri kullanılır. Arayüz rezidüleri proteinin yüzeyinin geri kalanlarından daha çok korunmuş olduğundan, bu evrimsel bilgilerle birleştirilerek arayüz tahminleri yapılmıştır. Bu yolla yapılmış ilk arayüz tahmininde %64'lük doğruluk elde edilmiştir (Reš ve ark. 2005). Bu yaklaşımda evrimsel bilgilere ek olarak, ikincil yapı ve yüzey erişilebilirliği gibi tahmin edilebilir yapısal bilgilerin (“predicted structural information”) avantajından da yararlanılmıştır. Bu bilgilerin dahil edilmesiyle yapılan çalışmalarda arayüz tahmin doğruluğunun artabileceği kanıtlanmıştır (Ofran ve Rost 2007; Murakami ve Mizuguchi 2010). Mevcut olan bilgilerin birleştirilmesiyle yapılan çalışmaların artık arayüz tahmin doğruluğunu arttırmamasından dolayı, sekans temelli yaklaşımların limitlerini doldurmuşlardır.

4.3.1.2. Yapısal Temelli Yaklaşımlar

Yapısal özellikler protein arayüz tahmini için önemli ayırım yapabilen özelliklerdir (attributes). Bu özellikler proteinlerin atomik koordinatlarıyla ilişkilidir. Çözümlemiş yapıların sayısının artması, 3B tabanlı arayüz tahminlerinin gelişmesine olanak sağlamıştır. Bu tahminlerde sorgulanan 3B'lu yapı ya birbirlerine yakın uzaklıktaki arayüz rezidülerinin tanımlanmasında kullanılır ya da arayüz rezidülerinin bulunması için yapısal özellikler olarak kullanılır. Yapısal temelli yaklaşımlar ya deneysel veriden ya da homoloji modellemesinden gelen yapısal bilgiyi kullanırlar. Önceki çalışmalar tahminin kalitesinin yapının kalitesine bağlı olduğunu ve homoloji modellerinin daha az doğru olan tahminler ürettiğini göstermiştir (Zhang ve ark. 2017). Yapısal yaklaşımlardan birisi protein yüzeylerinin yamalara bölünmesini ve etkileşim yüzeylerinin tahmin edilmesi için bu yamaların kullanılmasını içermektedir. Bu metotlar için yama büyüklüğü önceden tanımlanır ve değişmezdir. Bu da çoklu partnerleri olan proteinlerinin arayüzlerinin tahmininde ya da belirlenen yüzey yamasının gerçek arayüz büyüklüğünü doğru bir şekilde yansıtmaması problemlerine sebep olur (Jones ve

Thornton 1997). Birçok öngörü partnerleri göz ardı etseler de bu partnerlerin kullanılmasının tahminleri geliştiği gösterilmiştir.

4.3.2. Kalıp Temelli Yaklaşım

Mevcut yapısal komplekslerin sayısının artışı doğru arayüz şablonlarının tanımlanmasına yardımcı olmaktadır. Yapı temelli yaklaşımlar arayüzlerin korunmuş olduğu homolog komplekslere ihtiyaç duymaktadır (Jones ve ark. 2003). Ancak bu tarz homolog yapılar her zaman mevcut olmadığından, evrimsel olarak olmasa bile yapısal olarak benzer kompleks şablonlar bu çalışmalarda kullanılabilir. Kalıp temelli yaklaşım yapıları bilinen protein komplekslerinin veri tabanlarını tarar ve bu arayüzleri tahminler yapmak için kullanılır. Bu yaklaşım ilgilenilen proteinlerin önceden çalışılmış yapısal bilgilerine ihtiyaç duyar.

4.3.2.1. Homolog Şablon Temelli Yaklaşımlar

Bu metotlar etkileştiği bilinen protein kompleksindeki partnerlerden birisinin sorgulanan proteinin homoloğu olduğu kompleksleri kullanırlar. Homolog protein etkileşimlerinin arayüze karşılık geldiği yerde bir belirteç olduğu kabul edilen arayüz sorgulanan proteinde bulunabilir. Bu yaklaşım, homolog proteinlerin benzer bir oryantasyonla partnerleriyle etkileşme eğiliminde olduklarından ve her bir ailede bağlanma bölgeleri genellikle korunmuş olduğundan olasıdır. Ayrıca, arayüz rezidülerinin fizikokimyasal özellikleri, homolog olmayan proteinlere kıyasla, homolog olanlarda daha fazla benzerliğe sahiptir.

Bu yaklaşımda, tahminlerin kalitesine katkıda bulunduğu görülen temel özellikler yapı temelli çoklu sekans dizilemeleri (ÇSD) ve bağlanma partneri bilgileridir. Homolog şablon tabanlı yaklaşımlar, yapı içi temelli metotlara kıyasla tahminleri geliştirmiş olmasına rağmen, bu yaklaşımlar homolog kompleks yapıları mevcut olan proteinlerle sınırlıdır ve sorgulanan proteinin homolog kompleksinin bulunmadığı durumlarda başarısızdırlar.

4.3.2.2. Yapısal Komşu Temelli Yaklaşımlar

Sorgulanan proteinle benzer katlanmaları paylaşan proteinler, evrimsel olarak ilişkili olmasalar bile, homologlarına benzer tahmini bilgiler sunabilirler. Dahası, düşük sekans aynılığı olan benzer katlanmalara sahip olan proteinler aynı yerlerden partnerleriyle etkileşme eğilimindedirler. Bu tür yapısal komşular homoloji temelli metotlar için sorun yaratabilecek

düşük şablon kapsamıyla başa çıkabilmeye yardımcı olmak için arayüz tahmini için şablonlar olarak kullanır. Yapısal komşu temelli yaklaşımlarda %81'lik doğruluk başarılmıştır (Zhang ve ark. 2011; Zhang ve ark. 2010).

4.3.2.3. Partner-spesifik Arayüz Tahmini

Partner spesifik arayüz tahminleri çok yaygın değildir ancak belirli proteinler arasındaki tamamlayıcılığı dikkate alma avantajına sahiptir. Partner spesifik öngörüler her bir protein için etkileşen arayüz tahmininde etkileştiği varsayılan iki proteinin sekanslarını ya da yapılarını kullanır. Partner spesifik yaklaşım kullanıcının moleküler tanımda merkezi bir rol oynayan tamamlayıcılığı değerlendirmesine olanak tanır. Birden çok partnere rastgele bağlanan proteinler, arayüzleri tahmin etmek için benzersiz bir zorluk oluşturur. Bu çoklu bağlanma partnerlerinin hepsi aynı bölgeden bağlanabilir ya da protein yüzeyindeki birçok bölgeden bağlanabilir. Son çalışmalar bireysel proteinlerde bağlanma rezidülerinden bağlanmayan rezidüleri ayırma yeteneğini vurgulamakla birlikte, bu tahminler partner spesifik etkileşim bölgelerini ayırmada başarısızlık gösterir.

4.4. Voronoi/Delaunay Teselasyonu

Biyoinformatik yaklaşımları içinde önemli bir yer teşkil eden geometrik tabanlı çalışmalar Delaunay ve Voronoi teselasyonlarını kullanmaktadır.

Voronoi/Delaunay teselasyonları protein yapılarının çalışılmasında güçlü geometrik araçlardır. Kompleks atomik yapılarla başa çıkmakta kullanılmaktadırlar ve metodolojik olarak geliştirilmektedirler. Tanımlanan birkaç enteresan yeni uygulamalarıyla ümit vadeden sonuçlar elde edilmiştir (Şekil 4.7) (Poupon 2004).

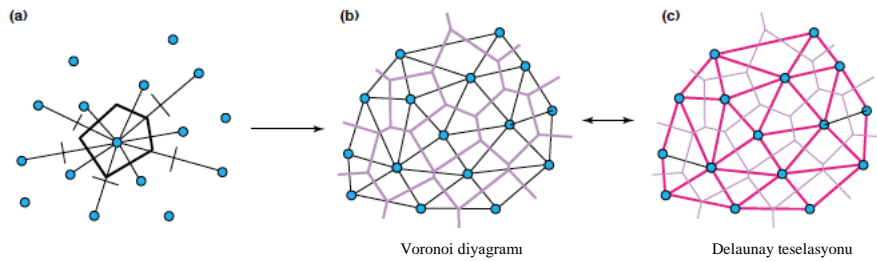
Rastgele atom sistemlerinin yapısının tanımlanmasında temel yaklaşım Voronoi çokgenlerine ve Delaunay üçgenlerine dayanmaktadır.

4.4.1. Voronoi Teselasyonu

Literatürde Dirichlet, Thiessen veya Wigner-Seitz diyagramı olarak da bilinen Voronoi diyagramı, Georgy Voronoi tarafından önerilmiş bir veri parçalama yöntemidir. Düzlemde yer alan sonlu nokta kümesine ait herhangi bir noktaya kümedeki diğer noktalardan daha yakın konumda bulunan düzlem noktalarının geometrik yerine o noktanın Voronoi Çokgeni

(poligonu) denilmektedir. Kümedeki tüm noktaların Voronoi çokgenlerinin birleşimi, o kümenin Voronoi diyagramını oluşturur.

Voronoi hücreleri ligandların ceplere sığması ve ligand bağlanma bölgelerinin belirlenmesi ve modellenmesi için de kullanılmıştır (Crippen 1987; Bradley ve Crippen 1993). Bir proteinin yüzeyinde veya içinde bulunan ve bir ligandın bağlanması için uygun özelliklere sahip olan boşluklar, genellikle bir bağlama cebi olarak adlandırılırlar (Stank ve ark. 2016). Voronoi teselasyonunun ayrıca protein-protein ve protein-DNA komplekslerindeki iki protein arasındaki ve protein-DNA arasındaki arayüz çalışmalarında kullanışlı olduğu bulunmuştur (Nadassy ve ark. 1999; Nadassy ve ark. 2001).



Şekil 4.7. Voronoi teselasyonunun ve Voronoi'den elde edilen yapılar. a) Bir noktanın Voronoi hücresi b) Voronoi hücrelerinin tekrarlanması sonucu elde edilen Voronoi teselasyonu. c) Ortak bir Voronoi yüzü paylaşan noktalar arasındaki bütün segmentlerin çizimiyle elde edilen Delaunay teselasyonu (Poupon 2004).

4.4.2. Delaunay Teselasyonu

1934 yılında Boris Delaunay tarafından ortaya atılan, Delaunay üçgenleri Voronoi diyagramının tamamlayıcısıdır. Voronoi çokgenleri bir teselasyonu tanımlar. İlgili Delaunay teselasyonu Voronoi çokgenlerinin ortak yüzeye sahip olan bütün noktalar arasındaki köşelerinin izlenmesiyle elde edilir. Her bir Voronoi diyagramı bir tek Delaunay teselasyonu ile örtüşmektedir (Poupon 2004). Delaunay üçgenlerinin elde edilmesinde, üç noktadan geçecek ve içerisinde başka herhangi bir nokta bulunmayacak biçimde daireler seçilir; bu dairenin üzerindeki noktalara ait üçgenlerin her biri bir Delaunay üçgenidir.

Protein zincirlerindeki amino asit rezidülerinin $C\alpha$ 'larca ifade edilmesiyle, proteinler üç boyutlu uzayda nokta setleri olarak tanımlanabilirler. Bir proteinin Delaunay teselasyonu irregüler tetrahedronlar kümesini ya da Delaunay simplekslerini oluşturur. Delaunay

teselasyonunu protein yapısının analizi için ilk uygulayanlar Singh ve ark.'nın çalışmasıdır (Singh ve ark. 1996).

4.5. Tetrahedronların Geometrik Karakterizasyonlarına Örnekler

Son zamanlarda algoritmaların ürettiği kaliteli tetrahedronluğa artan bir ilgi vardır. Şekil 4.8.'de bazı yaygın “kötü” şekilli tetrahedronların konfigürasyonlarını göstermektedir. Araştırmacılar, geçmişte, tetrahedron şeklini karakterize etmek için çeşitli ölçümler önermişlerdir (Parthasarathy 1994).

Caendish ve ark. bir tetrahedronu dış küre çapının iç küre çapına oranıyla karakterize etmiştir (Caendish ve ark. 1985). Baker ve ark. ince ve yassı tetrahedronları kategorize etmek için maksimum kenar uzunluğunun iç küre çapı ile maksimum kenar uzunluğunun minimum kenar uzunluğuna oranlarının kombinasyonunu önermiştir (Baker 1989). Dannenlogue ve Tanguy bir tetrahedronu karakterize etmek için tetrahedronun hacmi ve 6 köşenin ortalama kenar uzunluğuna oranının yeterli olabileceğini belirtmişlerdir (Dannelongue ve Tanguy 1991).



Şekil 4.8. Bazı yaygın “kötü” şekilli tetrahedronlar (Parthasarathy 1994)

4.5.1. Tetrahedronluk Özelliği (“Tetrahedrality”)

Medvedev ve ark. tarafından önerilmiş ve günümüzde de kullanılmakta olan Tetrahedronluk, ideal tetrahedrondan yola çıkılarak Delaunay simplekslerinin bozulma derecesinin nicel ölçülmesidir. Bir simpleksin tetrahedronluğunun sıfır olması onun düzenli bir tetrahedron olduğunu belirtir (Medvedev ve ark.1988; Vaisman ve ark.1998; Medvedev, 1987; Anikeenko ve ark. 2008).

$$T = \sum_{i>j} (l_i - l_j)^2 / 15\bar{l}$$

l_i i'ninci kenarın uzunluğu ve \bar{l} verilen simpleksteki kenarların ortalama uzunluğunu ifade etmektedir.

4.5.2. Küresellik (“Sphericity”)

Bir nesnenin ne kadar küresel olduğu küresellik (“sphericity”) ile hesaplanmaktadır (Wadell 1935). Bir nesnenin tıkHzlık (compactness) ölçümü küresellikle yapılmaktadır. Bir nesnenin küreselliği (Ψ), nesne ile aynı hacme (V_N) sahip bir kürenin yüzey alanı ile nesnenin yüzey alanının (A_N) oranına eşittir. Nesnenin küreselliği arttıkça bu oran 1'e yaklaşır.

Nesne ile aynı hacme (V_N) sahip kürenin yüzey alanı A_K ise:

$$A_K = 4\pi r^2 \quad (4.1)$$

$$V_K = \frac{4}{3}\pi r^3 = V_N \quad (4.2)$$

Eşitlik 4.1 ve 4.2 kullanılarak, Eşitlik 4.3 elde edilir.

$$A_K^3 = (4\pi r^2)^3 = 4^3 \pi^3 r^6 = 4\pi(4^2 \pi^2 r^6) = 4\pi 3^2 \left(\frac{4^2 \pi^2}{3^2} r^6\right) = 36\pi \left(\frac{4}{3}\pi r^3\right)^2 = 36\pi V_N^2 \quad (4.3)$$

Eşitlik 4.3'den yola çıkılarak böylece yüzey alanı bulunur.

$$A_K = (36\pi V_N^2)^{\frac{1}{3}} = 36^{\frac{1}{3}} \pi^{\frac{1}{3}} V_N^{\frac{2}{3}} = 6^{\frac{2}{3}} \pi^{\frac{1}{3}} V_N^{\frac{2}{3}} = \pi^{\frac{1}{3}} (6V_N)^{\frac{2}{3}} \quad (4.4)$$

Eşitlik 4.4 kullanılmasıyla küresellik bulunur.

$$\text{Küresellik } (\Psi) = \frac{A_K}{A_N} = \frac{\pi^{\frac{1}{3}} (6V_N)^{\frac{2}{3}}}{A_N} \quad (4.5)$$

Düzgün bir tetrahedronun küreselliği hesaplanırken nesne düzgün bir tetrahedron olduğundan “N” yerine tetrahedronu ifade etmesi açısından “tet” kullanılmıştır.

$$A_{\text{tet}} = \sqrt{3}a^2 \quad (4.6)$$

$$V_{\text{tet}} = \frac{a^3\sqrt{2}}{12} \quad (4.7)$$

Bu iki eşitlik 4.6 ve 4.7’yi, 4.5 de yerlerine koyarsak 4.8 elde edilir.

$$\text{Küresellik}_{\text{tet}} = \left(\frac{\pi}{6\sqrt{3}}\right)^{1/3} \cong 0,671 \quad (4.8)$$



5. MATERYAL VE METOT

5.1. Materyal

Bu çalışmada kullanılan proteinler belirlenirken ComSin veri tabanından (Lobanov ve ark. 2009) yararlanılmıştır. Proteinler belirlendikten sonra ilgili proteinlere ait PDB dosyaları indirilmiş, gerekli düzenlemeler yapılarak çalışmaya ait veri seti oluşturulmuştur.

5.1.1. Proteinlerin Belirlenmesi

Çalışmada kullanılacak proteinler belirlenirken öncelikle ComSin veri tabanında bulunan protein-protein etkileşimi tekil ve kompleks halinde sekans aynılığı %100 (yani aynı) olan proteinler seçilmiştir. Bu etkileşimli proteinlerin sayısının oranı 2448/24910 iken, protein aile sayısının oranı ise 352/1035'tir (Bahar ve ark. 2010). Daha sonra proteinler tekil ve kompleks halindeyken özünde düzensiz rezidüleri sahip proteinler çıkarılmıştır. Bunun sonucunda elde edilen etkileşen proteinlerin sayısı 696'ya düşmüştür. Bunların arasından toplamda 347 adet olan dimer seçilmiştir (trimer, tetramer ve diğer multimerler elenmiştir). ComSin veri tabanındaki heterodimerlerden hepsi elenmiştir çünkü bu komplekslerin yapılarında bulunan iki zincirin her ikisinin tekil yapıları bulunmamaktadır. Bu durumda etkileşen protein sayısı 318'a ve kompleks sayısı bunun yarısı olan 159'a düşmüştür (Şekil 5.1).

Proteinlerin PDB formatlarında yer alan bilgilerden en önemlilerinden olan REMARK bölümünde bu proteinlere ait eksik rezidü, eksik atom gibi bilgiler yer almaktadır. Tetrahedronların hacmini bulmak için atom düzeyinde hesaplamalar yapıldığından kullanılan protein PDB dosyalarında bu bilgiler büyük önem taşımaktadır. Seçilen PDB dosyalarında eksik rezidü ve atom olup olmadığının belirlenmesi için shell'de komutlar yazılmıştır. Bunlar aşağıda gösterilmiştir.

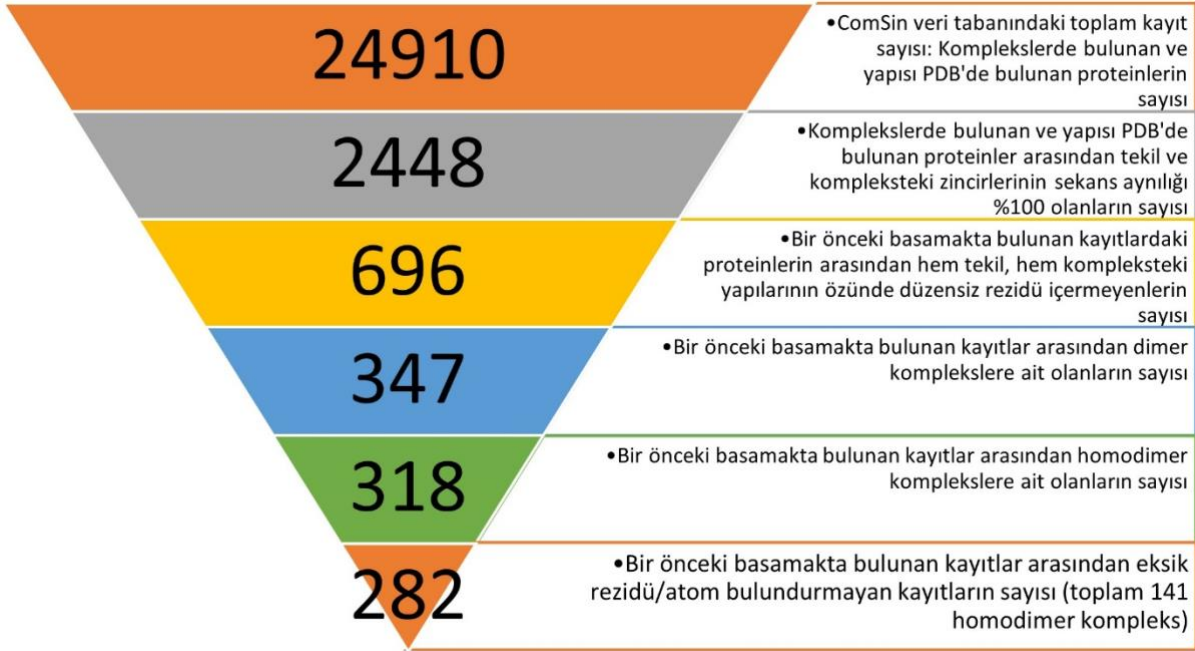
Eksik atomların bulunması için shell'de yazılan komut satırı;

```
>% sudo grep -irl --null "REMARK 470" . | xargs -0 sh -c 'mv "$@" ./moved' sh
```

Eksik rezidülerin bulunması için shell'de yazılan komut satırı;

```
>% sudo grep -irl --null "REMARK 465" . | xargs -0 sh -c 'mv "$@" ./moved' sh
```

Bu komut satırları sonucunda eksik rezidüye sahip proteinler hiç bulunmamış, eksik atoma sahip olan 18 tane protein bulunmuştur. Bu PDB’ler veri seti listesinden çıkarılmıştır. Bunun sonucunda etkileşen protein sayısı 282’ye ve kompleks sayısı da yarısı olan 141’e düşmüştür (Şekil 5.1). Bu proteinlerin PDB kodları Çizelge 5.1’de sunulmuştur. Proteinlerin sınıflandırılması yapılırken SCOP veri tabanı (Andreeva ve ark. 2019) kullanılmıştır.



Şekil 5.1. Protein veri setinin oluşturulması

PDB dosya formatında, protein ve nükleik asit zincirlerini ayırmak için TER kayıtları kullanılır. Zincirler birbiri ardına PDB dosyasına dahil edilirken, zincirlerin fiziksel olarak birbirine bağlı olmadığını belirtmek için TER kaydı yazılır. Tekil proteinlerde 1 tane olması beklenen sonlandırıcı olan TER sayısı; 2qhe’de 6 tane, 1bbc’de 5 tane, 1bio’da 5 tane, 1toj’da 7 tane, 1akb’de 6 tane olarak bulunmuştur. Ayrıca bu PDB dosyalarında TER’den sonra gelen rezidü sayıları artan şekilde ardışık devam etmemektedir. Sonuç olarak bu tekil proteinler ve kompleksleri de protein veri setinden çıkarılarak tezde çalışılacak olan protein sayısı 136 düşürülmüştür.

Çizelge 5.1. Çalışmada kullanılacak proteinlere ait bilgiler

PDB Kodu (Tekil)	Rezidü Sayısı	Protein	Ligand	PDB Kodu (Kompleks)	Protein	Ligand	Organizma
1f9h	158	Hppk(R92a)	APC, PH2, CL, MG	1g4c	Hppk(R92a)	CL, MG	<i>E. coli</i>
1aaw	396	Aspartat Aminotransferaz	PLP	1arg	Aspartat Aminotransferaz	PPD	<i>E. coli</i>
1aaw	396	Aspartat Aminotransferaz	PLP	1asl	Aspartat Aminotransferaz	PLA	<i>E. coli</i>
1aaw	396	Aspartat Aminotransferaz	PLP	1asm	Aspartat Aminotransferaz	MAE, PLP	<i>E. coli</i>
1aaw	396	Aspartat Aminotransferaz	PLP	1asn	Aspartat Aminotransferaz	SO4, PLP	<i>E. coli</i>
1noa	113	Apo-Neocarzinostatin		1nco	Antitümör Proteini	CHR, MRD	<i>S. carzinostaticus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	1b2k	Lizozim	IOD	<i>G. gallus</i>
1dkj	129	Lizozim		1dkk	Lizozim	NO3	<i>C. virginianus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	1hf4	Yumurta Akı Lizozimi	CL, NA	<i>G. gallus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	1lcn	Yumurta Akı Lizozimi	SCN	<i>G. gallus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	1lj3	Lizozim	NO3	<i>G. gallus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	1lj4	Lizozim	NO3	<i>G. gallus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	1lje	Lizozim	NO3	<i>G. gallus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	1ljf	Lizozim	NO3	<i>G. gallus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	1ljg	Lizozim	NO3	<i>G. gallus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	1ljh	Lizozim	NO3	<i>G. gallus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	1lji	Lizozim	NO3	<i>G. gallus</i>

193l	129	Yumurta Akı Lizozimi	CL, NA	1ljj	Lizozim	NO3	<i>G. gallus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	1ljk	Lizozim	NO3	<i>G. gallus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	1lkr	Yumurta Akı Lizozimi	IOD	<i>G. gallus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	1lys	Yumurta Akı Lizozimi		<i>G. gallus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	1rem	Lizozim	ACY	<i>G. gallus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	1uco	Yumurta Akı Lizozimi		<i>G. gallus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	1v7t	Lizozim	NO3, NA	<i>G. gallus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	2d4i	Yumurta Akı Lizozimi	NO3, NA	<i>G. gallus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	2d4k	Yumurta Akı Lizozimi	CL, NA	<i>G. gallus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	3lyt	Lizozim		<i>G. gallus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	4lyt	Lizozim		<i>G. gallus</i>
193l	129	Yumurta Akı Lizozimi	CL, NA	5lym	Yumurta Akı Lizozimi	NO3	<i>G. gallus</i>
1gqn	252	3-Dehydroquinaz		1qfe	Type I 3-Dehydroquinaz Dehydrataz	DHS	<i>S. enterica subsp. enterica serovar typhi</i>
3ezm	101	Cyanovirin-N		115b	Cyanovirin-N Dimer	NHE, NA	<i>N. elliposporum</i>
3ezm	101	Cyanovirin-N		1m5j	Cyanovirin-N	NHE, MG	<i>N. elliposporum</i>
3ezm	101	Cyanovirin-N		1m5m	Cyanovirin-N	NHE, MG	<i>N. elliposporum</i>
1cpq	129	Sitokrom C'	HEM	1nbb	Sitokrom C'	HEM NBN	<i>R. capsulatus</i>
1cpq	129	Sitokrom C'	HEM	1rcp	Sitokrom C'	HEM	<i>R. capsulatus</i>
1drf	186	Dihidrofolat Redüktaz	FOL, SO4	2c2t	Dihidrofolat Redüktaz	NDP, 39B, 39E, GOL	<i>H. sapiens</i>

1tjd	216	The Reduced Disulphide Bond İzomeraz, Dsbc		1eej	The Proteini Disülfit Bond İzomeraz, Dsbc	MES	<i>E. coli</i>
1ecy	142	Proteaz İnhibitor Ecotin	GLC, BGC	1ecz	Proteaz İnhibitor Ecotin	BOG	<i>E. coli</i>
1ejc	419	Mura (Type2)	PO4, GOL	1ejd	Unliganded Mura (Type1)	HAI, PO4	<i>Enterobacter cloacae</i>
1uz0	131	Karbohidrat Bağlama Modülü	GOL, CA, CL	1uxz	Karbohidrat Bağlama Modülü		<i>C. mixtus</i>
2cbr	136	Retinoik Asit Bağlama Proteini I	A80	1cbi	Retinoik Asit Bağlama Proteini		<i>M. musculus</i>
1fe3	131	Yağ Asidi Bağlama Proteini	OLA	1fdq	Yağ Asidi Bağlama Proteini	HXA	<i>H. sapiens</i>
1tvq	125	Yağ Asidi Bağlama Proteini		1tw4	Yağ Asidi Bağlama Proteini	CHD	<i>G. gallus</i>
1fkb	107	The Rapamycin Human Immunophilin Fkbp-12 Complex	RAP	1a7x	Fkbp12-Fk1012 Complex	FKA	<i>H. sapiens</i>
1fkb	107	The Rapamycin Human Immunophilin Fkbp-12 Complex		1fki	Fkbp Ligands Complexes	SB1	<i>H. sapiens</i>
2aae	104	Histidin-40 In Ribonucleaz T1 Catalysis	PO4, CA	2aad	Histidin-40 In Ribonucleaz T1 Catalysis	2GP, CA	<i>A. oryzae</i>
1vd5	377	Unsaturated Glucuronyl Hidrolaz	DTT, MPD, GLY	2d5j	Glucuronyl Hidrolaz		<i>Bacillus sp. GL1</i>
1ex7	186	Guanylate Kinaz	5GP, SO4	1ex6	Guanylate Kinaz		<i>S. cerevisiae</i>
1c81	191	Fruktoz-2,6-Bifosfataz	FDQ	1c7z	Fruktoz-2,6-Bifosfataz	G3H, PO4	<i>R. norvegicus</i>
1c81	191	Fruktoz-2,6-Bifosfataz	FDQ	1c80	Fruktoz-2,6-Bifosfataz	GTP, PO4	<i>R. norvegicus</i>

1c81	191	Fruktoz-2,6-Bifosfataz	FDQ	1tip	Bifosfataz Domeyn	F6P, NEP	<i>R. norvegicus</i>
1ooi	124	LUSH		1oof	Odorant Bağlama Proteini LUSH	ACT, EOH	<i>D. melanogaster</i>
1ooi	124	LUSH		1oog	Odorant Bağlama Proteini LUSH	POL, ACT	<i>D. melanogaster</i>
1ooi	124	LUSH		1t14	LUSH	ACT	<i>D. melanogaster</i>
1ooi	124	LUSH		2gte	OBP LUSH	VA, PO4	<i>D. melanogaster</i>
1v7r	186	Nükleotit Trifosfat Pirofosfataz	CIT	2dvn	PH1917 Proteini	IMP, SO4, GOL	<i>P. horikoshii</i>
1rpj	288	D-Alloz Bağlama Proteini	ALL, SO4, ZN	1gud	D-Alloz Bağlama Proteini	ZN	<i>E. coli</i>
1sn1	64	Nörotoksin Bmk M1		1djt	Alpha-Like Toksin Bmk M1		<i>M. martensii</i>
1utt	159	MMP-12	CP8, HAE, ZN, CA	1utz	MMP-12	PF3, HAE, ZN, CA	<i>H. sapiens</i>
1hrc	104	Sitokrom C	HEM	1crc	Sitokrom C	HEM	<i>E. caballus</i>
5cyt	103	Sitokrom C	HEM	1i54	Sitokrom C	ZNH, HEM	<i>T. alalunga</i>
5cyt	103	Sitokrom C	HEM	1i55	Sitokrom C	ZNH, HEM	<i>T. alalunga</i>
5cyt	103	Sitokrom C	HEM	1lfm	Sitokrom C	COH	<i>T. alalunga</i>
5cyt	103	Sitokrom C	HEM	3cyt	Sitokrom C	HEM	<i>T. alalunga</i>
1cd5	266	Glukozamin-6-Fosfat Deaminaz		1dea	Glukozamin-6-Fosfat Deaminaz	PO4	<i>E. coli</i>
1cd5	266	Glukozamin-6-Fosfat Deaminaz		1fqo	Glukozamin-6-Fosfat Deaminaz	F6R	<i>E. coli</i>
1cd5	266	Glukozamin-6-Fosfat Deaminaz		1frz	Glukozamin-6-Fosfat Deaminaz	16G	<i>E. coli</i>
1cd5	266	Glukozamin-6-Fosfat Deaminaz		1fs5	Glukozamin-6-Fosfat Deaminaz	16G, TLA	<i>E. coli</i>
1cd5	266	Glukozamin-6-Fosfat Deaminaz		1hor	Glukozamin-6-Fosfat Deaminaz	AGP, PO4	<i>E. coli</i>

1qw6	420	Nitrik Oksit Sentaz Oksijenaz Domeyn	HEM, H4B, 3AR, ZN	1zvl	Nitrik Oksit Sentaz Oksijenaz Domeyn	HEM, H4B, ARG, ZN	<i>R. norvegicus</i>
1job	162	Murine Olfactory Marker Proteini	ZN	1f35	Murine Olfactory Marker Proteini	CAC, ZN	<i>M. musculus</i>
1kwn	207	Thaumatın	TLA	1pp3	Thaumatın		<i>T. daniellii</i>
1atk	215	Katepsin K	E64	1nlj	Katepsin K	2CA	<i>H. sapiens</i>
1nqc	217	Katepsin S İnhibitor Complexes	C4P	1npz	Katepsin S İnhibitor Complexes	C1P	<i>H. sapiens</i>
1atk	215	Katepsin K	E64	1tu6	Katepsin K	FSP, SO4	<i>H. sapiens</i>
1atk	215	Katepsin K	E64	2ftd	Katepsin K	ILI	<i>M. mulatta</i>
1yx9	326	Porcine Pepsin	DMS	1psa	Pepsin/Renin İnhibitor	OZL	<i>S. scrofa</i>
1hpb	238	Histidin-Bağlama Proteini	HIS	1hsl	Histidin-Bağlama Proteini	HIS, CD	<i>E. coli</i>
1rzl	91	Lipit Transfer Proteini	CXS, SO4	1uvc	Lipit Transfer Proteini -1	STE	<i>O. sativa</i>
2acg	125	Profilin Ii		1f2k	Profilin Ii		<i>A. castellanii</i>
1dua	242	Exfoliative Toksin A		1agj	Epidermolitik Toksin		<i>S. aureus</i>
1gim	431	Adenilosuksinat Sentetaz	GDP, IMP, HDA, NO3, MG	1ade	Adenilosuksinat Sentetaz		<i>E. coli</i>
1gim	431	Adenilosuksinat Sentetaz	GDP, IMP, HDA, NO3, MG	1adi	Adenilosuksinat Sentetaz		<i>E. coli</i>
1gim	431	Adenilosuksinat Sentetaz	GDP, IMP, HDA, NO3, MG	1hon	Adenilosuksinat Sentetaz	GDP, IMP, HDA, NO3, MG	<i>E. coli</i>

1gim	431	Adenilosuksinat Sentetaz	GDP, IMP, HDA, NO3, MG	1hoo	Guanine Nucleotide (Gppcp)	GNP, GNH	<i>E. coli</i>
1gim	431	Adenilosuksinat Sentetaz	GDP, IMP, HDA, NO3, MG	1hop	Guanine Nükleotit (Gppcp)	GCP	<i>E. coli</i>
1zp8	99	HIV Proteaz	AB2	2o4k	HIV-1 Proteaz	DR7, CL	<i>HIV 1</i>
1zp8	99	HIV Proteaz	AB2	2o4p	HIV-1 Proteaz	TPV, GOL	<i>HIV 1</i>
1zp8	99	HIV Proteaz	AB2	2o4s	HIV-1 Proteaz	AB1, GOL, CL	<i>HIV 1</i>
1zp8	99	HIV Proteaz	AB2	2p3b	HIV Proteaz	3TL	<i>HIV 1</i>
1deo	233	Rhamnogalacturonan Asetilesteraz	NAG, SO4	1pp4	Rhamnogalactur onan Asetilesteraz	NAG	<i>A. aculeatus</i>
2jhy	138	Rhogdi E155h		2jhx	Rhogdi E155h		<i>H. sapiens</i>
1aqp	124	Ribonükleaz	CU	1a2w	Pankreatik Ribonükleaz A	SO4, CL	<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	1afk	Ribonükleaz A	PAP	<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	1jn4	Ribonükleaz A	139	<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	1o0f	RNAz A	A3P	<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	1o0h	Ribonükleaz A	ADP	<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	1o0m	Ribonükleaz A	U2P	<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	1o0n	Ribonükleaz A	U3P	<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	1o0o	Ribonükleaz A	A2P	<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	1qhc	Ribonükleaz A	PUA	<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	1rbb	Ribonükleaz B		<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	1u1b	Pankreatik Ribonükleaz A	PAX	<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	1w4o	Ribonucleaz A	UA3	<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	1w4q	Ribonucleaz A	UMF	<i>B. taurus</i>

1aqp	124	Ribonükleaz	CU	1wbu	Fragment Bazd Lead Discovery	WBU	<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	1xps	Ribonükleaz A		<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	1xpt	Ribonükleaz A		<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	1z6d	Ribonükleaz A	IMP	<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	1z6s	Ribonükleaz A	AMP	<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	2g4w	Ribonükleaz A	SO4, CL	<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	2g8q	RNAz A		<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	2g8r	RNAz A	N3E	<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	8rsa	Ribonükleaz A	ADT	<i>B. taurus</i>
1aqp	124	Ribonükleaz	CU	9rsa	Ribonükleaz A	ADU	<i>B. taurus</i>
1yp5	58	Myo5 SH3 Domeyn		1zuy	Myo5 SH3 Domeyn	IPA	<i>S. cerevisiae</i>
1b2l	254	Alkol Dehidrogenaz	NDC, DTT, CYH, CA	1a4u	Alkol Dehidrogenaz		<i>S. lebanonensis</i>
1b2l	254	Alkol Dehidrogenaz	NDC, DTT, CYH, CA	1b14	Alkol Dehidrogenaz	NAD	<i>S. lebanonensis</i>
1b2l	254	Alkol Dehidrogenaz	NDC, DTT, CYH, CA	1b15	Alkol Dehidrogenaz T	NAE	<i>S. lebanonensis</i>
1b2l	254	Alkol Dehidrogenaz	NDC, DTT, CYH, CA	1b16	Alkol Dehidrogenaz T	NAQ	<i>S. lebanonensis</i>
1b2l	254	Alkol Dehidrogenaz	NDC, DTT, CYH, CA	1sby	Alkol Dehidrogenaz	NAD, EFT	<i>S. lebanonensis</i>
1qke	62	Erabutoksin	SO4	1qkd	Erabutoksin		<i>L. semifasciata</i>
1ajm	264	Thymidylate Sentaz		1aiq	Thymidylate Sentaz	CB3, UMP	<i>E. coli</i>
1ffl	264	Apo-Thymidylate Sentaz		1fwm	Thymidylate Sentaz	CB3, SO4	<i>E. coli</i>
2cwy	94	TTHA0068		2cxd	TTHA0068		<i>T. thermophilus</i>

1ubi	76	Ubikütin		1aar	Diubikütin Conjugate		<i>B. taurus</i>
1udc	338	Udp-Galahtoz-4- Epimeraz	NAD, UFM, PGE, EDO, NA	2udp	Udp-Galahtoz-4- Epimeraz	NAD, UPP, EDO, NA	<i>E. coli</i>
1utg	70	Uteroglobin		2utg	Uteroglobin		<i>O. cuniculus</i>
1q7a	121	Russell's Viper Fosfolipaz A2	OPB, SO4, MOH	1cl5	Fosfolipaz A2		<i>D. russelii pulchella</i>
1q7a	121	Russell's Viper Fosfolipaz A2	OPB, SO4, MOH	1fb2	Fosfolipaz A2		<i>D. russelii pulchella</i>
1q7a	121	Russell's Viper Fosfolipaz A2	OPB, SO4, MOH	1fv0	Fosfolipaz A2	9AR, SO4, GOL, DIO, ACT	<i>D. russelii pulchella</i>
1q7a	121	Russell's Viper Fosfolipaz A2	OPB, SO4, MOH	1kpm	Fosfolipaz A2	VIT, ACY	<i>D. russelii pulchella</i>
1q7a	121	Russell's Viper Fosfolipaz A2	OPB, SO4, MOH	1ox1	Fosfolipaz A2 Dimer	IDA, SO4, CO3	<i>D. russelii pulchella</i>
1q7a	121	Russell's Viper Fosfolipaz A2	OPB, SO4, MOH	1tp2	Fosfolipaz A2 Dimer	TDA, SO4, ACY	<i>D. russelii pulchella</i>
1q7a	121	Russell's Viper Fosfolipaz A2	OPB, SO4, MOH	1y38	Fosfolipaz A2 Dimer	G3P, PGE, SO4, NA	<i>D. russelii pulchella</i>
2bch	123	Fosfolipaz A2	MPD, CA, CL	2bd1	Fosfolipaz A2	MPD, CA	<i>B. taurus</i>

5.1.2. Protein Veri Setinde Yapılan Bazı Manipülasyonlar

Tezde çalışılacak 136 protein komplekslerine ait PDB dosyalarında tetrahedron sayılarını ve buna bağlı olarak tetrahedron hacimleri ile küresellik değerlerini etkileyebilecek bazı durumlar düzenlenmiştir. Bu düzenlemeler “Occupancy” değeri ve” HETATM” bilgileridir.

PDB kodları incelendiğinde bazı proteinlerinde olması gereken rezidü sayısından daha fazla rezidü olduğu dikkati çekmiştir. PDB kodlarında her bir atomun koordinatları deneysel olarak bulunup, işlenmektedir. Bu koordinatlarının yerlerinin doğruluğunun belirlenmesi için her bir atomun koordinatları girildiğinde her zaman orada olup olmadıklarına bakılmaktadır. Bazı atomlar her zaman aynı koordinatta bulunmayabilir. Atomların her zaman aynı koordinatlarda bulunup bulunmadığı PDB dosyalarında kodlarında “occupancy” değeri ile gösterilmektedir. “Occupancy” değeri 1.0 olması o atomun her zaman o koordinatta bulunduğunu ifade ederken, aynı atoma ait farklı koordinatlar varsa bunlar 0.3, 0.5, 0.7 gibi “occupancy” değerleriyle ifade edilir. Bu değerler atomun o koordinatlarda görülme sıklığını ifade etmektedir. Değerin 1.0’dan farklı olması bu atom en az 2 farklı yerde görüldüğünü belirtmektedir. Farklı yerlerdeki görülme sıklıkları toplamı her zaman 1.0 olmalıdır. Örneğin bir atom 15.132, -8.198, 32.480 x, y, ve z koordinatlarında 0.5 oranında bulunurken aynı atom 15.099, -8.195, 32.492 x, y, z koordinatlarında 0.5 oranında bulunmuştur. Bu iki oranın toplamı 1.0’dır. Dolayısıyla bir PDB dosyasında “occupancy” değeri 1.0’dan farklı olan C α atomları ile Delaunay teselasyonu tetrahedronlar oluşturulduğunda tetrahedron hacim hesaplamalarında yanlışlıklara sebep olacaktır. Bu çalışmada kullanılacak olan PDB dosyalarından “occupancy” değeri 1.0’dan farklı olan C α atomları PyMOL programında yazılan bir script ile her bir rezidü için bir tane C α atomu olacak şekilde düzeltilmiştir (Schrodinger 2010). Scriptten önce bir *.pml* dosyası oluşturulmuştur. Bu dosyayı oluşturmak için bir script daha yazılmıştır.

```
> load 1a4u.pdb
remove not alt "+A
alter all, alt=""
save 1a4u_alter.pdb
delete all
```

Hazırlanan bu *.pml* dosyası PyMOL programında, çalıştırılarak “occupancy” değerleri 1.0’dan farklı olan PDB dosyalarında birden fazla C α atom koordinatları bulunanların sayısı bire indirgenmiştir.

PyMOL>@DimerOccAltering.pml komut ile PyMOL programında çalıştırılmıştır.

Veri seti ile ilgili bütün PDB dosyaları incelendiğinde bazı dosyalarda bazı rezidülere ait atom bilgilerinin ATOM, bazılarının da HETATM olarak işlendiği görülmüştür. ATOM; protein, DNA ya da RNA'nın standart rezidülerindeki atomları içermektedir. HETATM; karbonhidratlar, substratlar, ligandlar, solventler ve metal iyonlarının yanı sıra DNA, RNA ya da proteinlerin standart olmayan rezidülere uygulanmaktadır. Bilinen 20 amino aside ek olarak genetik olarak kodlanan selenosistein ve pirolizin amino asitler standart rezidüler olarak bilinirler. Bunlar dışında kalanlar is standart olmayan rezidülerdir.

Veri setinde yer alan 1aiq, 1ajm, 1ejc, 1ejd, 1f35, 1fwm, 1ffl, 1tip, 1tok, 1toj, 2cxd, 2cwy, 2p3b PDB kodlu dosyalarda HETATM'lar ATOM olarak değiştirilmiştir (Şekil 5.2). Bu HETATM'lar ve değiştirilen HETATM sayıları Çizelge 5.2'de gösterilmiştir.

Çizelge 5.2. İlgili PDB dosyalarında değiştirilen HETATM'lar ve sayıları

PDB	HETATM Adı	HETATM Kısaltması	HETATM Sayısı
1aiq	N-karboksümetiyonin	CXM	2
1ajm	N-karboksümetiyonin	CXM	1
1ejc	Beta-L-aspartik asit	IAS	1
1ejd	Beta-L-aspartik asit	IAS	2
1f35	Selenometiyonin	MSE	10
1fwm	N-karboksümetiyonin	CXM	2
1ffl	N-karboksümetiyonin	CXM	1
1tip	N1-fosfonohistidin	NEP	2
1tok	((2S)-2-amino-6-[[3-hidroksi-2-metil-5-(fosfonooksimetil)piridin-4-yl]metilideneamino]heksanoik asit)	LLP	2
1toj	((2S)-2-amino-6-[[3-hidroksi-2-metil-5-(fosfonooksimetil)piridin-4-yl]metilideneamino]heksanoik asit)	LLP	1
2cxd	Selenometiyonin	MSE	2
2cwy	Selenometiyonin	MSE	1
2p3b	Selenometiyonin	CME	1

Kalsiyum elementi de PDB dosyasında HETATM olarak elementin simgesi olan CA (Ca) şeklinde kaydedildiğinden, her rezidüde bulunan ve CA olarak kaydedilen C α (alfa karbon) ile karışıklığına sebebiyet verdiğinden bu HETATM’lar PDB dosyasından silinmiştir.

HETATM	1	N	CXM	A	1	22.186	20.264	49.286	1.00	25.89	N
HETATM	2	CA	CXM	A	1	23.510	20.742	49.650	1.00	26.72	C
HETATM	3	CB	CXM	A	1	24.527	20.351	48.562	1.00	31.15	C
HETATM	4	CG	CXM	A	1	24.568	18.854	48.263	1.00	30.44	C
HETATM	5	SD	CXM	A	1	25.569	18.462	46.833	1.00	31.15	S
HETATM	6	CE	CXM	A	1	24.388	18.725	45.476	1.00	29.10	C
HETATM	7	C	CXM	A	1	23.473	22.236	49.790	1.00	29.28	C
HETATM	8	O	CXM	A	1	22.746	22.912	49.065	1.00	27.78	O
HETATM	9	CN	CXM	A	1	21.411	18.980	49.749	1.00	42.20	C
HETATM	10	ON1	CXM	A	1	22.006	18.090	50.381	1.00	28.21	O
HETATM	11	ON2	CXM	A	1	20.201	18.833	49.421	1.00	29.05	O
ATOM	12	N	LYS	A	2	24.258	22.761	50.714	1.00	25.45	N
ATOM	13	CA	LYS	A	2	24.280	24.191	50.905	1.00	21.24	C
ATOM	14	C	LYS	A	2	24.203	24.944	49.574	1.00	20.13	C
ATOM	15	O	LYS	A	2	23.280	25.708	49.366	1.00	20.42	O
ATOM	16	CB	LYS	A	2	25.494	24.636	51.722	1.00	26.59	C
ATOM	17	CG	LYS	A	2	25.894	26.091	51.482	1.00	58.47	C
ATOM	18	CD	LYS	A	2	25.800	26.925	52.755	1.00	75.58	C
ATOM	19	CE	LYS	A	2	24.419	26.816	53.404	1.00	100.00	C
ATOM	20	NZ	LYS	A	2	23.307	27.050	52.437	1.00	89.71	N
ATOM	21	N	GLN	A	3	25.157	24.702	48.671	1.00	18.44	N
ATOM	22	CA	GLN	A	3	25.204	25.402	47.381	1.00	21.13	C
ATOM	23	C	GLN	A	3	23.910	25.319	46.565	1.00	31.39	C
ATOM	24	O	GLN	A	3	23.459	26.316	45.975	1.00	23.34	O

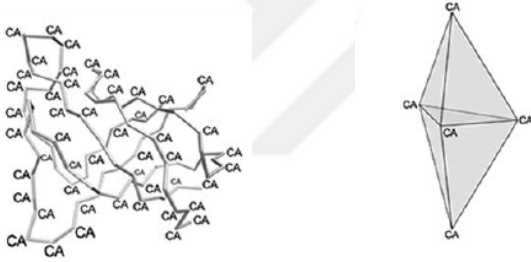
Şekil 5.2. 1aiq PDB kodlu proteine ait “CXM” HETATM’ları

5.2. Metot

Protein yüzeyleri ve protein komplekslerinin arayüzlerinin karakterizasyonu için ilgili proteinlerle ilgili PDB dosyaları kullanılarak öncelikle Delaunay teselasyonu ile tetrahedronlar elde edildikten sonra bu tetrahedronların hacimleri ve küresellik değerleri hesaplanmıştır. Daha sonra elde edilen bu tetrahedronlardan proteinin yüzeyini oluşturanlar ve yüzeyde olan tetrahedronlarda bulunan rezidülerin protein komplekslerinin arayüzüne geçenleri bulunmuştur. Bütün hesaplamalar R programlama dilinde yazılmıştır (R Core Team 2020).

5.2.1. Tetrahedronların Hacimlerinin Hesaplanması

Tetrahedronların oluşturulmasında Delaunay teselasyonu kullanılmıştır. Delaunay teselasyonu sonucunda üç boyuttaki (3B) noktaları köşe olarak alan her bir tetrahedronu çevreleyen küre içinde nokta bulunmamalıdır. Bu çalışmada, noktaların koordinatları bir proteindeki C α (alfa karbon) atomlarının 3B'lu koordinatlarıdır (Şekil 5.3).



Şekil 5.3. Bir proteinden tetrahedronların elde edilmesi (Aydinkal ve Bağcı 2018)'den modifiye edilmiştir

Delaunay teselasyonu ile tetrahedronların elde edilmesinde ilk basamak, proteinlere ait olan PDB dosyalarının <https://www.rcsb.org/pdb> sitesinden indirilmesidir (R kodları eğik yazı ile gösterilmiştir) (Berman ve ark. 2000).

```
> mypdbdata <-read.pdb(file="d3n3.pdb", ATOM = TRUE)
```

İndirilen PDB dosyasında rezidülere ait C α atomları da bulunmaktadır. İlk olarak bu atomlar elde edilmiştir. Elde edilen bu C α atomların koordinatları bulunmuştur.

```
> CAcoords <-coords(subset(mypdbdata,elename=='CA')$atoms)
```

Koordinatlar matrise dönüştürülerek *mycoordata* değişkenine atanmıştır.

```
> mycoordata <- as.matrix(CAcoords)
```

Elde edilen bu koordinatlardan Delaunay teselasyonu ile tetrahedronlar oluşturulmuştur. Bir tetrahedron için 4 tane rezidünün C α atomlarına ihtiyaç vardır. Bunların koordinatları kullanılarak 4 tane indis elde edilmiş ve *tc* değişkenine atanmıştır. Her bir *tc* satırı 1 tetrahedronu oluşturan C α atomlarının indislerine sahiptir.

```
> tc <- delaunayn(mycoordata)
```

tc değişkenindeki her C α 'ya isim verilmiştir (*tc1*, *tc2*, *tc3*, *tc4*). Bu işlem *tc*'deki satır sayısı [*jjj*] kadar tekrar edilmiştir. *tc1* tetrahedronun birinci elemanını, *tc2* ikinci, *tc3* üçüncü, *tc4* ise dördüncü elemanını oluşturmuştur.

```
> for (jjj in 1:nrow(tc)) {  
  tc1 <- tc[jjj,1]  
  tc2 <- tc[jjj,2]  
  tc3 <- tc[jjj,3]  
  tc4 <- tc[jjj,4]  
}
```

Her bir *tc* elemanının x, y ve z koordinatları bulunmuştur (Çizelge 5.3).

```
tc1x <- mycoordata[tc1,1]  
tc1y <- mycoordata[tc1,2]  
tc1z <- mycoordata[tc1,3]  
tc2x <- mycoordata[tc2,1]  
tc2y <- mycoordata[tc2,2]  
tc2z <- mycoordata[tc2,3]  
tc3x <- mycoordata[tc3,1]  
tc3y <- mycoordata[tc3,2]  
tc3z <- mycoordata[tc3,3]  
tc4x <- mycoordata[tc4,1]  
tc4y <- mycoordata[tc4,2]  
tc4z <- mycoordata[tc4,3]
```

Çizelge 5.3. Tetrahedronu oluşturan C α (alfa karbon) atomlarına ait x, y ve z koordinatları

	x	y	z
tc1	mycoordata [tc1,1]	mycoordata [tc1,2]	mycoordata [tc1,3]
tc2	mycoordata [tc2,1]	mycoordata [tc2,2]	mycoordata [tc2,3]
tc3	mycoordata [tc3,1]	mycoordata [tc3,2]	mycoordata [tc3,3]
tc4	mycoordata [tc4,1]	mycoordata [tc4,2]	mycoordata [tc4,3]

Bir tetrahedronun hacmi (V), karşılık olan bir paralel yüzünün 1/6'sı oranındadır (Şekil 5.4). Paralel yüzünün hacmi $|\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})|$ ile bulunmuştur (Şekil 5.5). Burada \mathbf{a} , \mathbf{b} ve \mathbf{c} paralel yüzü oluşturan vektörlerdir.

\mathbf{a} , \mathbf{b} ve \mathbf{c} vektörlerinin koordinatlarını aşağıdaki gibi kabul edersek

$$\mathbf{a} = x_3\mathbf{i} + y_3\mathbf{j} + z_3\mathbf{k} \quad (5.1)$$

$$\mathbf{b} = x_1\mathbf{i} + y_1\mathbf{j} + z_1\mathbf{k} \quad (5.2)$$

$$\mathbf{c} = x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k} \quad (5.3)$$

$$V = |\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})| / 6 \quad (5.4)$$

$$\mathbf{b} \times \mathbf{c} = (x_1\mathbf{i} + y_1\mathbf{j} + z_1\mathbf{k}) \times (x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k}) \quad (5.5)$$

$$\mathbf{b} \times \mathbf{c} = (y_1z_2 - z_1y_2)\mathbf{i} - (x_1z_2 - z_1x_2)\mathbf{j} + (x_1y_2 - y_1x_2)\mathbf{k} \quad (5.6)$$

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = (x_3\mathbf{i} + y_3\mathbf{j} + z_3\mathbf{k}) \cdot ((y_1z_2 - z_1y_2)\mathbf{i} - (x_1z_2 - z_1x_2)\mathbf{j} + (x_1y_2 - y_1x_2)\mathbf{k}) \quad (5.7)$$

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = x_3(y_1z_2 - z_1y_2) - y_3(x_1z_2 - z_1x_2) + z_3(x_1y_2 - y_1x_2) \quad (5.8)$$

$$V = |x_3(y_1z_2 - z_1y_2) - y_3(x_1z_2 - z_1x_2) + z_3(x_1y_2 - y_1x_2)| / 6 \quad (5.9)$$

olarak bulunmuştur.

Hacim için gerekli olan \mathbf{a} , \mathbf{b} ve \mathbf{c} vektörleri aşağıdaki şekilde bulunmuştur;

$$\mathbf{a} = x_3\mathbf{i} + y_3\mathbf{j} + z_3\mathbf{k} \quad \text{için } (x_3 : \text{vec}21x, y_3 : \text{vec}21y, z_3 : \text{vec}21z)$$

$$\begin{aligned} \text{vec}21x &<- tc2x-tc1x \\ \text{vec}21y &<- tc2y-tc1y \\ \text{vec}21z &<- tc2z-tc1z \\ \text{vec}21 &<- c(\text{vec}21x, \text{vec}21y, \text{vec}21z) \end{aligned}$$

$$\mathbf{b} = x_1\mathbf{i} + y_1\mathbf{j} + z_1\mathbf{k} \quad \text{için } (x_1 : \text{vec}31x, y_1 : \text{vec}31y, z_1 : \text{vec}31z)$$

$$\begin{aligned} \text{vec}31x &<- tc3x-tc1x \\ \text{vec}31y &<- tc3y-tc1y \\ \text{vec}31z &<- tc3z-tc1z \\ \text{vec}31 &<- c(\text{vec}31x, \text{vec}31y, \text{vec}31z) \end{aligned}$$

$$\mathbf{c} = x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k} \quad \text{için } (x_2 : \text{vec}41x, y_2 : \text{vec}41y, z_2 : \text{vec}41z)$$

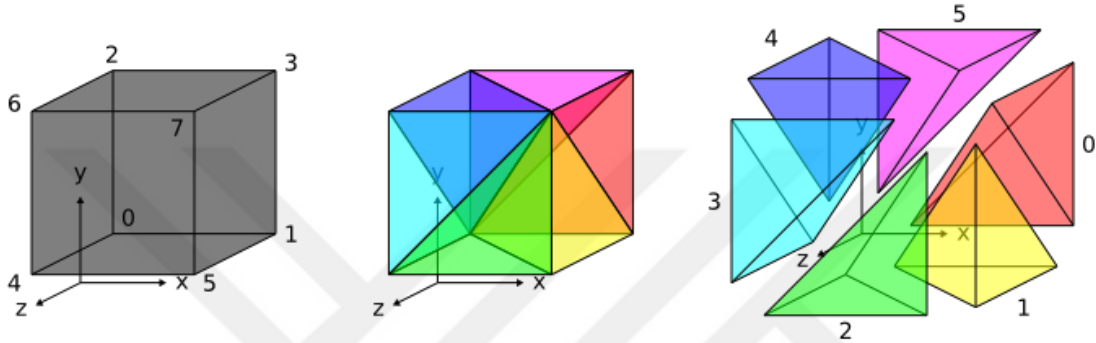
$$\begin{aligned} \text{vec}41x &<- tc4x-tc1x \\ \text{vec}41y &<- tc4y-tc1y \\ \text{vec}41z &<- tc4z-tc1z \\ \text{vec}41 &<- c(\text{vec}41x, \text{vec}41y, \text{vec}41z) \end{aligned}$$

bxc için

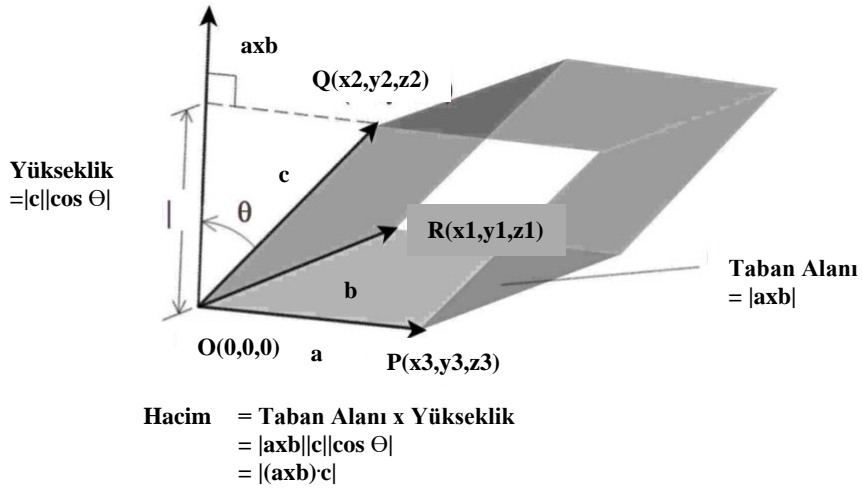
```
> crossvecfn <- function(vector1,vector2) {
  crossvectorx <- vector1[2] * vector2[3] - vector1[3] * vector2[2]
  crossvectory <- vector1[3] * vector2[1] - vector1[1] * vector2[3]
  crossvectorz <- vector1[1] * vector2[2] - vector1[2] * vector2[1]
  c(crossvectorx,crossvectory,crossvectorz)}
```

$V = |\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})| / 6$ işlemi için:

```
crossvector <- crossvecfn(vec31,vec41)
V <- abs(vec21%*%crossvector)/6
```



Şekil 5.4. Bir paralel yüzünün içerisine yerleşen 6 adet tetrahedronun şematik gösterimi (Dune-Project.org)



Şekil 5.5. Paralel yüzünün hacminin hesaplanması

5.2.2. Tetrahedronların Küresellik Değerlerinin Hesaplanması

Tetrahedronun küreselliğinin hesaplanması için, tetrahedronun hacmi ile tetrahedronun yüzey alanının bilinmesi gerekmektedir. Tetrahedronun hacmini bulunması bir önceki konuda anlatılmıştır. Tetrahedronun yüzey alanının belirlenmesi için tetrahedronu oluşturan üçgenlerin, üçgenlerin belirlenmesi için üçgenleri oluşturan kenarların, kenarların da belirlenmesi için tetrahedronu oluşturan dört tane rezidüye ait x, y ve z koordinatlarının bulunması gerekmektedir.

Rezidülere ait x, y ve z koordinatları tetrahedronun hacmi hesaplanırken kullanılmıştı. Bu koordinatlar aşağıdaki gibi kullanılarak tetrahedrona ait 6 kenar aşağıda ifade edildiği şekilde hesaplanmıştır.

```
< a1 <- sqrt( (tc1x - tc2x)^2 + (tc1y - tc2y)^2 + (tc1z - tc2z)^2 ) # 1. kenar uzunluğu
< a2 <- sqrt( (tc1x - tc3x)^2 + (tc1y - tc3y)^2 + (tc1z - tc3z)^2 ) # 2. kenar uzunluğu
< a3 <- sqrt( (tc1x - tc4x)^2 + (tc1y - tc4y)^2 + (tc1z - tc4z)^2 ) # 3. kenar uzunluğu
< a4 <- sqrt( (tc2x - tc3x)^2 + (tc2y - tc3y)^2 + (tc2z - tc3z)^2 ) # 4. kenar uzunluğu
< a5 <- sqrt( (tc2x - tc4x)^2 + (tc2y - tc4y)^2 + (tc2z - tc4z)^2 ) # 5. kenar uzunluğu
< a6 <- sqrt( (tc3x - tc4x)^2 + (tc3y - tc4y)^2 + (tc3z - tc4z)^2 ) # 6. kenar uzunluğu
```

Bulunan kenarlar tetrahedronu oluşturan 4 üçgenin alanını bulmakta kullanılmıştır.

```
< s1 <- sqrt( (a1+a2+a4)*(a2+a4-a1)*(a4+a1-a2)*(a1+a2-a4) )/4 #1. üçgenin alanı
< s2 <- sqrt( (a1+a3+a5)*(a3+a5-a1)*(a5+a1-a3)*(a1+a3-a5) )/4 #2. üçgenin alanı
< s3 <- sqrt( (a2+a3+a6)*(a3+a6-a2)*(a6+a2-a3)*(a2+a3-a6) )/4 #3. üçgenin alanı
< s4 <- sqrt( (a4+a6+a5)*(a6+a5-a4)*(a5+a4-a6)*(a4+a6-a5) )/4 #4. üçgenin alanı
```

4 üçgenin alanının toplamı ile tetrahedronun yüzey alanı bulunmuştur.

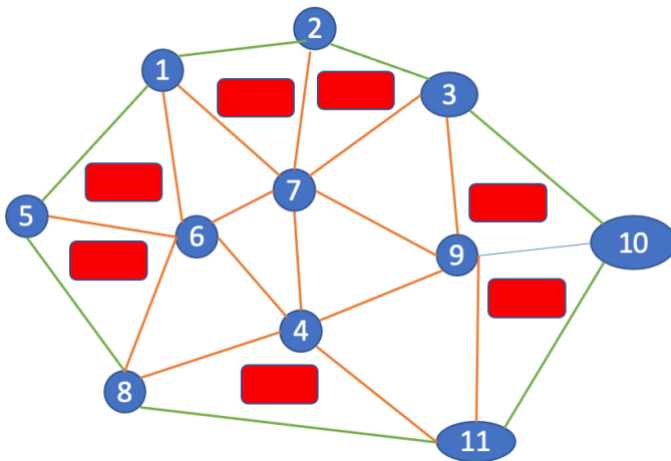
```
< s <- s1+s2+s3+s4 #Tetrahedronun yüzey alanı
```

Tetrahedronlara ait hacim ve yüzey alanı bulunduktan sonra tetrahedronun küresellik değeri aşağıdaki komut satırı ile hesaplanmıştır.

```
<sphericity[jjj] <- ((6*Vjjj)^(2/3)*pi^(1/3))/s
```

5.2.3. Protein Yüzeyinin Belirlenmesi

Proteinler 3B'lu yapıya sahiptirler ve proteini oluşturan rezidülerin bir kısmı en üst kısımda kalarak proteinin yüzeyini oluştururlarken bir kısmı da iç kısımda bulunurlar. Proteinler kompleks oluştururken genellikle proteinlerin yüzeyinde bulunan rezidüler etkileşerek protein arayüzlerini oluştururlar. Bundan dolayı hangi rezidülerin yüzeyde bulunduğu bilinmesi protein-protein etkileşimlerini anlaşılması için oldukça önemlidir. 4 rezidüden oluşturulmuş tetrahedronların 4 tane üçgen yüzeyi bulunmaktadır. Eğer bir tetrahedrona ait bu üçgen yüzeyler, başka bir tetrahedrona ait bir üçgene ait değil ise bu üçgene ait tetrahedron yüzey tetrahedronudur. Bu durumu 2B'da gösterebilmek için Şekil 5.6 çizilmiştir. Çok sayıda birbirine komşu olan üçgenler kümesinden oluşmuş bir yapıda hangi üçgenlerin yapının yüzeyini hangilerinin yapının iç kısmını oluşturduğu Şekil 5.6'da görülmektedir. Üçgenlerin 3 kenarı bulunmaktadır. Eğer üçgenin en az bir kenarı diğer üçgenlere ait değil ise bu üçgen yapının yüzeyini oluşturur. Bunun aksine eğer üçgenin üç kenarı da diğer başka bir üçgene de ait ise bu üçgen yapının içerisinde yer alır. Şekil 5.6'da yeşil renkle ifade edilen kenarlar en dıştaki kenarlar sadece bir üçgene ait bir kenarı oluşturmakta ve diğer üçgenler ile herhangi bir kesişimi bulunmamaktadır. Dolayısıyla bu kenara sahip olan üçgenin yapının dışında olduğu yani yapının yüzeyini oluşturduğu söylenebilir. Turuncu ile gösterilmiş olan kenarlar ise diğer üçgenler ile kesiştiğinden turuncu kenarları olan üçgenler yapının içerisindedir. Üçgen içerisindeki kırmızı dikdörtgenler bu üçgenlerin yapının yüzeyinde bulduklarını ifade etmektedir.



Şekil 5.6. 2B'de çok sayıda birbirine komşu olan üçgenler kümesinden oluşmuş bir yapıda yüzey üçgenlerinin belirlenmesi

1ejc PDB kodlu proteine ait Delaunay sonucu elde edilmiş birkaç tetrahedrona ait veriler satırlar şeklinde Çizelge 5.4’de gösterilmiştir. Satırlar incelendiğinde tetrahedronu oluşturan rezidülerin sıra sayılarının genellikle artan bir şekilde yazılmadığı görülmüştür.

```
> tc<-delaunayn(mycoordata)
```

Çizelge 5.4. 1ejc PDB kodlu proteine ait başlangıç tetrahedronları

	[,1]	[,2]	[,3]	[,4]
[1,]	113	418	419	140
[2,]	156	154	131	155
[3,]	330	368	333	114
[4,]	157	156	183	155
[5,]	157	121	131	155
[6,]	157	156	131	155
[7,]	133	154	131	155
[8,]	85	113	418	140
[9,]	366	368	333	114
[10,]	366	368	113	114
[11,]	336	113	340	348
[12,]	336	346	348	340
[13,]	336	330	333	114

11. ve 12. satırlarda belirtilen tetrahedronlara bakıldığında 11. tetrahedrona ait rezidülerin 336-113-340-348, 12. tetrahedronunkiler ise 336-346-348-340 olduğu görülmüştür.

Bu tetrahedronlara ait üçgenler oluşturulmak istendiğinde;

```
> mymatrix<-matrix(c(tc[11,1],tc[11,2],tc[11,3],1),ncol=1)
```

```
> mymatrix<-cbind(mymatrix,matrix(c(tc[11,1],tc[11,2],tc[11,4],1),ncol=1))
```

```
> mymatrix<-cbind(mymatrix,matrix(c(tc[11,1],tc[11,3],tc[11,4],1),ncol=1))
```

```
> mymatrix<-cbind(mymatrix,matrix(c(tc[11,2],tc[11,3],tc[11,4],1),ncol=1))
```

```
tc[11,1] = 336
```

```
tc[11,2] = 113
```

```
tc[11,3] = 340
```

```
tc[11,4] = 348
```

```
#336-113-340-348: 336-113-340, 336-113-348, 336-340-348, 113-340-348
```

```

> mymatrix<-matrix(c(tc[12,1],tc[12,2],tc[12,3],1),ncol=1)
> mymatrix<-cbind(mymatrix,matrix(c(tc[12,1],tc[12,2],tc[12,4],1),ncol=1))
> mymatrix<-cbind(mymatrix,matrix(c(tc[12,1],tc[12,3],tc[12,4],1),ncol=1))
> mymatrix<-cbind(mymatrix,matrix(c(tc[12,2],tc[12,3],tc[12,4],1),ncol=1))

```

tc[12,1] = 336

tc[12,2] = 346

tc[12,3] = 348

tc[12,4] = 340

#336-346-348-340: 336-346-348, 336-346-340, **336-348-340**, 346-348-340 üçgenleri elde edilmiştir.

Yukarıda koyu şekilde ifade edilen üçgenler aynı rezidüler ile oluşturulmasına rağmen ikinci ve üçüncü rezidülerin sıralaması farklı olduğundan, algoritma bu üçgenlerin aynı olduğunu ayırt edemeyecektir. Üçgenlerin ayırt edilememesi yüzey tetrahedronlarının belirlenmesinde hatalı sonuçlar elde edilmesine neden olacaktır. Bu hatanın giderilmesi için algorithmada her bir tetrahedronu oluşturan 4 rezidünün küçükten büyüğe doğru sıralanması gerçekleştirilmiştir.

```

> for(j in 1:nrow(tc)){
>   sortedindicis<-sort.int(c(tc[j,1],tc[j,2],tc[j,3],tc[j,4]),decreasing = F)
>   tc[j,1]<-sortedindicis[1]
>   tc[j,2]<-sortedindicis[2]
>   tc[j,3]<-sortedindicis[3]
>   tc[j,4]<-sortedindicis[4]
}

```

Bir sonraki aşamada aynı sıralamaya sahip rezidülerin oluşturduğu üçgenlerin silinip sadece farklı olanların tutulacağı bir fonksiyon yazılmıştır.

```

> functioncheck<-function(mymatrix,newtriangle){
>   ncolmymatrix<-ncol(mymatrix)
>   for (i in 1:ncolmymatrix){
>

```

```

>if(newtriangle[1,1]==mymatrix[1,i]      &      newtriangle[2,1]==mymatrix[2,i]      &
newtriangle[3,1]==mymatrix[3,i]){
  mymatrix<-mymatrix[-i]
  return(mymatrix)
}
}
> mymatrix<-cbind(mymatrix,newtriangle)
> return(mymatrix)}

```

Fonksiyon üçgenlerin aynı olup olmadıklarını kıyaslayabilmek için ilk tetrahedrondan elde edilmiş üçgenler kullanılmıştır.

```

>mymatrix<-matrix(c(tc[1,1],tc[1,2],tc[1,3],1),ncol=1)
>mymatrix<-cbind(mymatrix,matrix(c(tc[1,1],tc[1,2],tc[1,4],1),ncol=1))
>mymatrix<-cbind(mymatrix,matrix(c(tc[1,1],tc[1,3],tc[1,4],1),ncol=1))
>mymatrix<-cbind(mymatrix,matrix(c(tc[1,2],tc[1,3],tc[1,4],1),ncol=1))

```

İlk tetrahedronun üçgenleri ile aynı olan üçgenler silinmiş, farklı olanlar ise matris olacak şekilde *mymatrix* değişkenine atanmıştır. Diğer tetrahedronlara ait üçgenler *mymatrix* değeri içinde yer alan bütün yüzey üçgenleriyle kıyaslanmış ve farklı olanlar bu değışkende birikmeye devam etmiştir. Böylece yüzeyde bulunan üçgenler *mymatrix* ile belirlenmiştir. Bu üçgenlerin hangi tetrahedrona ait olduğunu belirtmek için her matris değerinin sonuna *ii* (yukarıda koyu ile 1 yazılmıştır çünkü ilk tetrahedronla işlem yapılmıştır) ile ifade edilen *tc*'nin satır numarası yazdırılmıştır.

```

>for (ii in 2:nrow(tc)){
> newtriangle <- matrix(c(tc[ii,1],tc[ii,2],tc[ii,3],ii),ncol=1)
> mymatrix<-functioncheck(mymatrix,newtriangle)
> newtriangle <- matrix(c(tc[ii,1],tc[ii,2],tc[ii,4],ii),ncol=1)
> mymatrix<-functioncheck(mymatrix,newtriangle)
> newtriangle <- matrix(c(tc[ii,1],tc[ii,3],tc[ii,4],ii),ncol=1)
> mymatrix<-functioncheck(mymatrix,newtriangle)
> newtriangle <- matrix(c(tc[ii,2],tc[ii,3],tc[ii,4],ii),ncol=1)
> mymatrix<-functioncheck(mymatrix,newtriangle)
>}

```

Bazı tetrahedronların birden fazla üçgeni yüzeyde bulunduğundan bu tetrahedronların hacimleri hesaplanırken hata oluşmaması için *tc* satır numaraları sadece bir defa tekrar ettirilmiş (unique fonksiyonu ile) ve *surfacetetrahedra* değişkene atanmıştır. “*surfacetetrahedra*” bulunan *tc* satır numaraları kullanılarak proteine ait yüzeyde bulunan tetrahedronları oluşturan rezidüler *SurMat* değişkenine atanmıştır.

```
>surfacetetrahedra<-unique(mymatrix[4,])
>SurMat<-tc[surfacetetrahedra,]
```

SurMat değişkeni kullanılarak yüzeyi oluşturan tetrahedronların hacim, küresellik değerleri hesaplanmıştır.

5.2.4. Protein Arayüzünün Belirlenmesi

Bir tetrahedronun protein etkileşimlerinin arayüzünde kabul edilmesi için en az bir rezidüsünün proteinin farklı zincirinde bulunması gerekmektedir. *Residue_number* değişkeni homo-dimer proteinin monomerinin sahip olduğu rezidü sayısını ifade etmektedir. Rezidü sayısı *Residue_number*'a eşit veya küçükse homo-dimer proteinin ilk zincirini, büyükse proteinin ikinci zincirini, en az bir rezidüsü farklı zincirde ise protein arayüzünü oluşturmaktadır.

```
>if (tc[jjj,1]<=Residue_number && tc[jjj,2]<=Residue_number &&
tc[jjj,3]<=Residue_number && tc[jjj,4]<=Residue_number) {
    arraytcV1<-c(arraytcV1, tc[jjj,1],tc[jjj,2],tc[jjj,3],tc[jjj,4])}

>else if
(tc[jjj,1]>Residue_number&&tc[jjj,2]>Residue_number&&tc[jjj,3]>Residue_number&&tc[
jjj,4]>Residue_number) {
    arraytcV2<-c(arraytcV2, tc[jjj,1],tc[jjj,2],tc[jjj,3],tc[jjj,4])}

>if
((tc[jjj,1]>Residue_number||tc[jjj,2]>Residue_number||tc[jjj,3]>Residue_number||tc[jjj,4]>
Residue_number)&&(tc[jjj,1]<=Residue_number||tc[jjj,2]<=Residue_number||tc[jjj,3]<=Re
sidue_number ||tc[jjj,4]<=Residue_number)) {
    arraytcInt<-c(arraytcInt,tc[jjj,1],tc[jjj,2],tc[jjj,3],tc[jjj,4]) }
```


Arayüzde bulunan rezidüler *tcInt* değişkenine matris şeklinde atanmıştır.

```
>tcInt<-matrix(arraytcInt,ncol=4, byrow=TRUE )
```

Arayüzde bulunan her bir tetrahedrona (*tcInt*) ait rezidüler, proteinin yüzeyini oluşturan her bir üçgende (*mymatrix*) bulunan rezidülerle aynı ise bu yüzey rezidülerinin arayüz rezidüsüne dönüştüğünün bir göstergesi kabul edilmiştir. *tcInt*, kompleks proteine ait olduğundan içerdiği rezidü sayısı, tekil proteininden elde edilen *mymatrix*'deki rezidü sayısının iki katıdır. *Residue_number* tek bir zincirde bulunan rezidü sayısını ifade etmektedir. Dolayısıyla *Residue_number* 'a göre modu (*%%Residue_number*) alındığında kompleksin ikinci zincirinde bulunan rezidülerin sıra sayısı, tekil halde bulunan rezidünün sıra sayısına eşit hale getirilmiştir. *mymatrix*'in ilk üç satırında üçgenlere ait rezidü bilgileri bulunurken, dördüncü satırında (*mymatrix[4,ggg]*) ise bu üç rezidünün bulunduğu tetrahedronun *tc* değişkenindeki satır numarası bulunmaktadır. Böylece dördüncü satır bilgileri kullanılarak ilgili tetrahedronlar bulunarak hacim, küresellik hesapları yapılmıştır.

```
>for (mmm in 1:nrow(tcInt)) {  
> for(ggg in 1:ncol(mymatrix))  
> if(tcInt[mmm,1]==mymatrix[1,ggg]||  
tcInt[mmm,1]==mymatrix[2,ggg]||  
tcInt[mmm,1]==mymatrix[3,ggg]||  
tcInt[mmm,2]==mymatrix[1,ggg]||  
tcInt[mmm,2]==mymatrix[2,ggg]||  
tcInt[mmm,2]==mymatrix[3,ggg]||  
tcInt[mmm,3]==mymatrix[1,ggg]||  
tcInt[mmm,3]==mymatrix[2,ggg]||  
tcInt[mmm,3]==mymatrix[3,ggg]||  
tcInt[mmm,4]==mymatrix[1,ggg]||  
tcInt[mmm,4]==mymatrix[2,ggg]||  
tcInt[mmm,4]==mymatrix[3,ggg]||  
tcInt[mmm,1]%%Residue_number==mymatrix[1,ggg]||  
tcInt[mmm,1]%%Residue_number==mymatrix[2,ggg]||  
tcInt[mmm,1]%%Residue_number==mymatrix[3,ggg]||  
tcInt[mmm,2]%%Residue_number==mymatrix[1,ggg]||
```

```

tcInt[mmm,2]%%Residue_number==mymatrix[2,ggg]||
tcInt[mmm,2]%%Residue_number==mymatrix[3,ggg]||
tcInt[mmm,3]%%Residue_number==mymatrix[1,ggg]||
tcInt[mmm,3]%%Residue_number==mymatrix[2,ggg]||
tcInt[mmm,3]%%Residue_number==mymatrix[3,ggg]||
tcInt[mmm,4]%%Residue_number==mymatrix[1,ggg]||
tcInt[mmm,4]%%Residue_number==mymatrix[2,ggg]||
tcInt[mmm,4]%%Residue_number==mymatrix[3,ggg]){
> tcRowNumber<-c(mymatrix[4,ggg],tcRowNumber)}
>}

```

tcRowNumber değişkeni, yüzeyde bulunan üçgenlerdeki rezidülerden arayüze geçenlerin *tc*'deki tetrahedron satır bilgilerini barındırmaktadır. Bu satır bilgileri kullanılarak *Interface* değişkeninde proteinin arayüzünü oluşturan tetrahedronu oluşturan 4 rezidü bilgisi bulunmaktadır. Bu arayüzde bulunan 4 rezidü kullanılarak hacim ve küresellik değerleri hesaplanmıştır.

```
>Interface<-tc[tcRowNumber,]
```

5.2.5. Protein Yüzeyinde ve Arayüzünde Bulunan Keskin Tetrahedronların Belirlenmesi

Protein yüzeyini oluşturan tetrahedronlar incelendiğinde az sayıda da olsa bazı tetrahedronların iki tane üçgen yüzünün proteinin yüzeyinde yer aldığı görülmüştür (aşağıda duplicated fonksiyonu ile elde edilmişlerdir). Tetrahedronun iki yüzeyinin protein yüzeyinde bulunması protein yüzeyinin bazı yerlerde dışarıya doğru çıkıntı yaptığının bir göstergesi olduğu düşünülmüş ve bu şekilde olan tetrahedronların arayüzde bulunma sıklıklarının yüksek ihtimal olduğu düşünülmüştür. Protein yüzeyinde bu şekilde olan tetrahedronlar bulunmuş ve bunlara ait hacim, küresellik değerleri hesaplanmıştır.

```

>surfacetetrahedra<-mymatrix[4,]
>surfacetetrahedra<-surfacetetrahedra[duplicated(surfacetetrahedra)]
>mymatrix<-mymatrix[,duplicated(mymatrix[4,])]
>SurMat<-tc[surfacetetrahedra,]

```

6. ARAŞTIRMA BULGULARI

6.1. Tekil ve Kompleks Yapılarda Tetrahedron Sayıları, Ortalama Hacim ve Ortalama Küresellik Değerlerindeki Değişim

Tekil proteinler homo-dimer kompleksleri oluşturduklarında; tetrahedron sayıları ortalama olarak 2,17 katına çıkmıştır. Kompleksteki tetrahedronların ortalama hacimleri ortalama olarak 1,26 katına çıkmıştır (Çizelge 6.1).

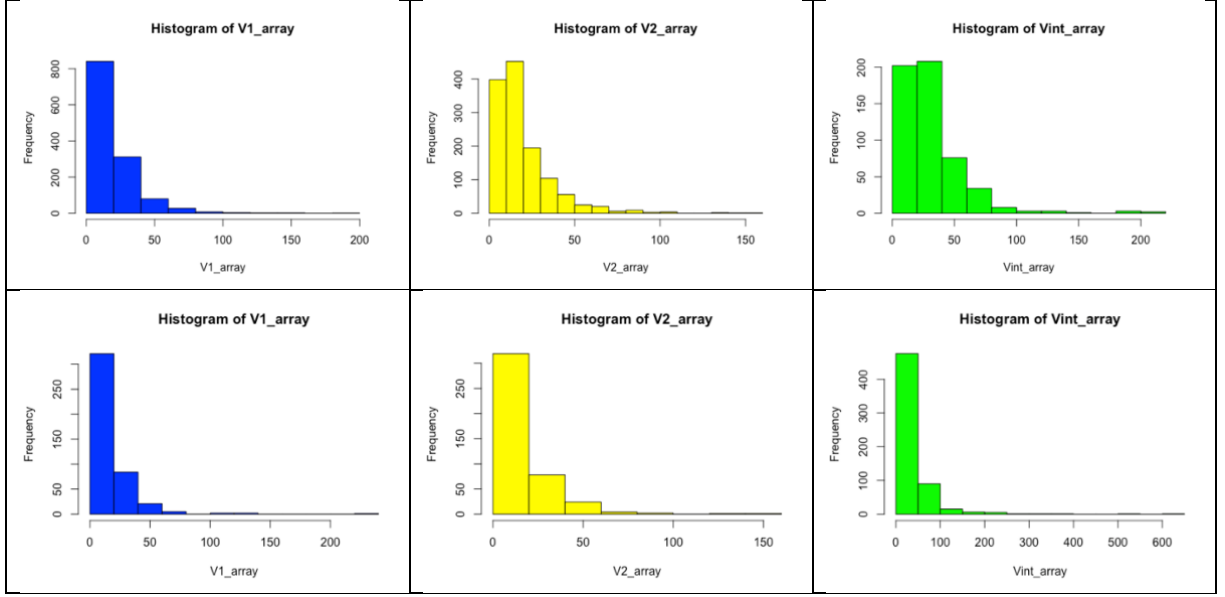
Komplekslerin arayüzdeki tetrahedronları incelendiğinde; sayılarının tüm tetrahedronların ortalama %18'ini oluşturduğu bulunmuştur. Arayüz tetrahedronlarının ortalama hacimleri diğer tetrahedronların ortalama hacimlerinin 2,17 katıdır (Şekil 6.1). Bu tetrahedronların ortalama küresellik değerleri diğer tetrahedronların değerlerinden düşüktür (0,61 oranında).

Çizelge 6.1. Homo-dimer kompleksleri oluşturan zincirler ve arayüzlere ait bilgiler

PDB	1. Zincir			2. Zincir			Arayüz		
	Tet. Sayısı	Tet. Hacim Ort. (Å ³)	Tet. Hacim Std. Sapma	Tet. Sayısı	Tet. Hacim Ort. (Å ³)	Tet. Hacim Std. Sapma	Tet. Sayısı	Tet. Hacim Ort. (Å ³)	Tet. Hacim Std. Sapma
1a2w	436	17.64	18.33	429	16.86	15.18	597	38.07	50.78
1a4u	1278	19.63	17.45	1277	19.61	17.09	540	31.08	26.46
1a7x	505	17.07	14.60	498	17.39	15.03	215	45.26	45.10
1aar	302	16.58	13.70	308	17.12	14.91	206	31.93	26.66
1ade	2365	22.23	22.67	2373	22.11	20.83	781	47.05	52.22
1adi	2367	22.63	22.32	2379	22.81	23.51	779	47.75	53.40
1afk	604	21.49	24.44	595	21.10	26.23	241	54.41	48.82
1agj	1368	20.16	17.63	1354	21.65	21.23	260	158.51	168.25
1aiq	1360	21.50	19.70	1378	21.40	19.81	525	41.48	41.19
1aka	2147	21.89	22.51	2152	21.65	22.27	800	48.62	55.88
1arg	2114	21.15	21.77	2110	21.22	21.96	756	46.15	51.05
1asl	2130	20.78	20.76	2101	21.06	20.85	768	45.27	49.67
1asm	2124	21.00	20.85	2124	20.99	20.37	759	45.69	50.37
1asn	2128	21.68	21.91	2121	21.82	21.73	755	47.54	51.57
1b14	1284	18.91	14.96	1283	18.91	15.00	545	29.85	27.81
1b15	1285	19.40	16.19	1283	19.44	16.05	536	30.63	27.06
1b16	1293	19.37	16.00	1275	19.60	15.93	528	31.25	27.06
1b2k	640	18.39	13.99	627	17.04	12.64	233	62.15	55.54
1c7z	966	18.48	14.91	960	18.66	15.08	360	45.29	37.48
1c80	965	18.71	15.12	966	18.79	15.34	362	47.09	42.63
1cbi	675	18.91	20.07	663	18.68	18.83	279	43.48	45.11
1cl5	584	17.97	14.30	575	17.51	13.76	291	45.97	42.07
1crc	499	17.42	17.65	509	17.39	16.53	187	62.20	46.18
1dea	1501	20.79	16.84	1497	20.73	16.99	330	109.84	125.73
1dfp	1254	19.60	15.50	1247	19.74	15.65	329	51.04	46.20
1djt	262	14.99	10.86	252	15.16	10.75	148	33.00	31.42
1dkk	627	17.97	14.67	628	18.51	14.65	230	65.32	64.63

lecz	574	18.99	16.20	581	18.91	16.17	543	58.52	58.76
leej	1129	21.94	21.67	1139	22.96	24.40	386	87.12	126.30
lejd	2384	21.96	25.74	2402	22.01	27.35	506	121.13	142.87
lex6	964	24.89	38.47	961	26.25	45.79	322	65.76	71.83
1f2k	610	16.76	11.94	613	16.85	11.79	232	46.07	38.99
1f35	782	21.67	20.15	806	21.93	20.84	363	90.81	96.55
1fb2	603	17.03	13.83	586	17.19	13.22	271	49.62	49.34
1fdq	647	20.46	22.07	646	20.03	21.35	236	51.19	43.69
1fki	514	17.20	16.11	524	18.30	17.13	212	44.37	48.00
1fqo	1506	20.87	17.16	1503	20.80	17.27	323	112.68	126.68
1frz	1516	20.70	17.24	1502	20.77	17.14	327	111.42	127.20
1fs5	1504	20.53	17.20	1501	20.69	17.53	340	106.41	124.88
1fv0	585	17.60	14.48	578	17.35	13.67	281	46.46	42.56
1fwm	1366	21.82	20.22	1381	21.55	19.65	530	41.01	40.92
1g4c	777	21.23	22.46	744	17.93	13.95	357	47.14	43.84
1gud	1587	22.75	28.86	1564	21.23	22.40	415	75.82	85.04
1hf4	625	17.72	14.48	638	18.07	14.02	232	63.74	60.84
1hon	2348	22.91	27.68	2386	22.43	23.62	782	47.80	52.05
1hoo	2369	21.88	23.62	2366	22.51	23.21	780	47.32	51.17
1hop	2354	22.73	25.48	2367	22.98	23.47	784	48.16	54.32
1hor	1507	20.72	16.90	1493	20.72	17.12	333	108.91	125.02
1hsl	1325	21.08	19.38	1311	20.76	17.28	309	87.68	104.05
1i54	484	17.47	16.61	485	17.40	16.72	205	45.29	32.44
1i55	486	17.37	17.04	483	17.52	17.25	205	45.48	32.38
1jn4	599	21.32	24.36	590	20.98	25.33	252	50.81	44.37
1kpm	578	17.93	14.85	582	17.36	14.23	277	48.43	45.49
1l5b	261	17.59	22.40	261	17.82	24.22	658	27.95	30.08
1lcn	624	17.50	13.30	647	18.48	14.15	231	64.08	55.96
1lfm	489	17.22	16.95	486	17.33	16.95	204	45.23	32.49
1lj3	621	18.00	14.41	636	18.21	14.61	237	62.80	68.95
1lj4	625	17.93	14.53	639	18.18	14.16	232	63.89	61.46
1lje	621	18.18	14.66	644	18.14	14.25	233	64.14	70.69
1ljf	623	18.05	14.59	637	18.30	14.73	233	64.52	70.38
1ljg	624	17.99	14.69	637	18.21	14.31	236	63.00	60.27
1ljh	627	17.91	14.53	638	18.24	14.54	232	64.24	62.70
1lji	622	17.95	14.56	638	18.13	14.18	235	62.96	62.31
1ljj	625	17.90	14.41	643	18.06	14.29	227	65.48	67.30
1ljk	621	18.18	14.59	634	18.52	14.65	233	64.40	62.89
1lkr	625	17.32	12.93	643	18.70	14.69	226	65.52	57.97
1lys	615	17.84	13.84	635	18.53	13.91	239	60.23	63.01
1m5j	270	17.03	22.07	269	16.39	21.82	637	28.43	32.05
1m5m	267	18.09	24.08	263	18.26	27.04	641	27.67	30.72
1nbb	620	18.19	17.68	621	18.65	18.92	265	48.99	82.39
1nco	516	16.07	13.00	526	15.57	12.39	250	56.51	67.04
1nlj	1132	19.05	15.18	1153	19.32	15.39	301	69.19	66.28
1npz	1146	18.74	14.05	1157	19.15	14.99	339	57.37	58.84
1o0f	595	21.63	24.48	596	21.32	27.27	251	49.91	46.13
1o0h	598	21.57	24.33	593	21.08	25.13	250	50.30	47.05
1o0m	595	21.70	24.10	596	20.97	25.30	241	51.72	48.66
1o0n	597	21.69	24.38	594	21.01	25.44	247	51.52	48.74
1o0o	603	21.41	23.90	598	20.98	26.29	244	52.14	47.85
1oof	610	18.04	15.33	611	18.04	15.20	199	41.83	34.59
1oog	608	18.10	14.62	608	18.12	14.63	199	41.76	34.66
1oxl	588	18.03	15.14	579	17.52	14.17	273	48.27	45.52
1poe	622	16.89	12.21	628	16.82	12.40	240	62.90	65.96
1pp3	1099	20.73	20.93	1095	20.42	21.29	331	88.88	108.93
1pp4	1208	19.14	14.31	1230	19.35	16.13	386	55.38	53.20
1psa	1789	20.77	18.07	1783	20.86	17.32	486	56.04	57.78

lqfe	1386	18.78	15.03	1378	18.90	15.05	361	48.46	52.98
lqhc	598	21.54	24.75	593	20.93	24.03	249	51.86	44.16
lqkd	238	15.72	13.84	244	15.22	12.96	205	41.43	41.87
lrbb	598	20.58	25.20	617	20.91	25.10	235	56.08	49.51
lrem	628	18.27	14.36	628	18.43	14.37	285	43.26	40.17
lrep	624	18.28	17.30	625	17.77	16.23	259	48.42	78.55
lsby	1291	19.26	15.94	1266	19.55	15.97	527	31.08	26.87
lt14	610	18.07	15.04	610	18.00	15.03	196	42.77	35.75
ltip	975	18.81	15.28	960	19.07	15.36	359	46.08	38.42
ltok	2109	20.94	19.68	2106	20.83	20.26	750	45.40	50.42
ltp2	595	17.09	13.35	575	17.58	13.67	279	49.34	47.49
ltu6	1129	18.31	13.47	1123	18.77	13.81	368	52.24	47.49
ltw4	616	19.74	22.16	619	20.39	22.19	229	62.14	58.75
lu1b	596	19.62	20.96	603	20.67	22.78	249	63.06	85.00
luco	622	17.84	14.15	634	18.31	14.37	233	63.80	65.70
lutz	805	19.43	14.35	800	19.59	14.51	289	39.91	31.06
luvc	391	16.82	13.00	401	16.10	11.41	220	27.86	22.04
luxz	630	17.63	15.22	626	17.81	15.16	231	39.03	33.43
lv7t	631	17.89	13.34	626	17.84	13.07	215	66.64	63.44
lw4o	599	21.42	25.04	601	20.85	25.22	257	48.58	43.42
lw4q	601	21.47	24.29	590	21.47	25.07	250	52.06	48.31
lwbu	604	21.27	25.12	596	20.87	26.29	246	52.86	47.97
lxps	613	21.74	23.72	607	19.82	16.50	198	71.02	81.20
lxpt	622	21.60	24.57	612	19.52	17.01	204	72.56	79.68
ly38	591	17.18	12.98	575	17.56	13.45	279	49.33	46.73
lz6d	602	21.43	24.46	596	21.04	25.43	246	51.69	48.58
lz6s	603	21.48	24.02	596	21.06	26.36	242	53.77	52.24
lzuy	215	15.48	11.31	223	15.17	10.86	158	27.72	21.71
lzvl	2259	23.43	22.80	2257	23.57	23.77	835	65.66	120.78
2aad	515	18.90	15.15	515	18.76	15.52	183	65.91	57.42
2bd1	597	17.63	13.01	613	17.06	12.71	255	49.73	35.41
2c2t	992	21.26	18.80	982	21.56	19.22	312	72.28	62.65
2cxd	426	17.41	13.01	400	17.06	12.41	187	41.97	34.75
2d4i	628	17.95	14.43	642	18.21	14.07	233	64.00	62.34
2d4k	621	17.94	14.64	638	18.53	14.76	232	62.68	66.96
2d5j	2169	20.53	17.65	2163	20.51	17.24	408	71.30	72.58
2dvn	965	19.80	16.91	923	20.14	17.56	367	51.10	50.76
2ftd	1159	19.04	15.27	1147	19.17	15.12	356	63.14	59.87
2g4w	601	21.37	24.30	599	21.02	25.91	248	50.33	43.75
2g8q	602	21.31	24.03	598	20.83	24.65	249	50.96	47.70
2g8r	598	21.63	24.65	600	20.83	25.27	250	50.08	47.72
2gte	600	18.28	14.89	595	18.06	14.89	223	42.07	34.77
2jhx	675	22.17	21.23	666	21.12	19.11	289	52.25	54.23
2o4k	378	16.28	13.50	378	16.56	14.01	354	22.82	19.50
2o4p	377	16.16	13.49	378	16.69	14.24	360	22.28	18.32
2o4s	371	16.47	13.58	381	16.51	14.24	351	23.04	19.02
2p3b	386	16.15	13.84	378	16.70	14.24	337	24.82	20.35
2qhd	599	17.29	13.32	591	17.65	14.34	272	51.60	50.01
2udp	1898	20.10	16.75	1912	20.12	16.91	437	73.80	86.37
2utg	247	12.99	8.76	250	12.63	8.38	248	26.04	20.11
3cyt	486	17.40	17.71	483	17.29	17.12	205	45.49	33.32
3lyt	632	17.44	13.24	647	17.77	13.59	215	67.89	67.53
4lyt	628	17.78	14.05	629	18.47	14.30	226	66.47	66.69
5lym	612	18.24	14.32	639	18.20	14.22	229	64.95	63.10
8rsa	580	18.33	14.89	571	18.60	15.21	307	43.73	64.92
9rsa	572	18.35	14.83	566	18.53	15.20	306	43.52	65.18



Şekil 6.1. 1a4u ve 1a2w pdb kodlu protein komplekslerine ait tetrahedronların hacimlerinin ortalamalarına ait histogram grafikleri. Sol taraftaki histogram kompleksin birinci zincirine, ortadaki ikinci zincirine, sağdaki ise arayüzüne aittir. Üsteki protein kompleksi 1a4u pdb kodlu protein kompleksine aitken, alttakiler 1a2w pdb kodlu protein kompleksini göstermektedir

6.2. Tekil Proteine Ait Tetrahedron Sayılarının Protein Yüzeyindeki ve Arayüzündeki Değişimleri

Her bir protein kompleksi için tüm tetrahedronların ve tetrahedronların tekil proteinde, komplekste, yüzey ve arayüzde bulunma durumları Çizelge 6.2’de ve Şekil 6.2’de sunulmuştur. Tekil proteinlerin ortalama tetrahedron sayısı 971,56 iken yüzeyi oluşturan tetrahedronların ortalaması 71,99 olarak bulunmuştur. Tek tek hesaplandığında ise, tekil proteinlerin yüzey tetrahedronlarının sayısının tüm tetrahedronların ortalamada %9’ünü oluşturduğu bulunmuştur. Her bir protein için yüzeyi oluşturan tetrahedronların kaçınının bir veya daha rezidüsünün komplekste arayüzde olduğu bulunmuştur ve yüzde oran hesaplanmıştır. Tüm proteinlerde bu oranların ortalaması %94’tür. Keskin tetrahedronlar için hesap yapıldığında, oran %96 olarak bulunmuştur. Fark az olmasına rağmen keskin tetrahedronların önemli bir bulgu içerdiği Çizelge 6.2’de görülmektedir. Protein yüzeyi 67 komplekste, yani tüm komplekslerin %50’sinde tamamen arayüze dönüşmüştür, keskin tetrahedronlar ise 96 protein kompleksinde (toplam protein komplekslerinin %71’i) tamamen protein arayüzüne geçmiştir.

Çizelge 6.2. Tekil, yüzey ve arayüzdeki tetrahedron değerleri

Tekil	Kompleks	Tekil Protein	Protein Yüzeyi	Protein Yüzeyinden Protein Arayüze	Protein Yüzeydeki Keskin Tetrahedron	Arayüzdeki Keskin Tetrahedron
PDB	PDB	Tetrahedron Sayısı	Tetrahedron Sayısı	Tetrahedron Sayısı	Tetrahedron Sayısı	Tetrahedron Sayısı
1f9h	1g4c	854	79	79	21	21
1aaw	1arg	2458	88	72	10	9
	1asl	2458	88	72	10	10
	1asm	2458	88	72	10	10
	1asn	2458	88	72	10	10
1noa	1nco	609	56	47	12	11
1931	1b2k	682	69	69	11	11
	1hf4	682	69	69	11	11
	1lcn	682	69	69	11	11
	1lj3	682	69	69	11	11
	1lj4	682	69	69	11	11
	1lje	682	69	69	11	11
	1ljf	682	69	69	11	11
	1ljg	682	69	69	11	11
	1ljh	682	69	69	11	11
	1lji	682	69	69	11	11
	1ljj	682	69	69	11	11
	1ljk	682	69	69	11	11
	1lkr	682	69	69	11	11
	1lys	682	69	69	11	11
	1rcm	682	69	66	11	10
	1uco	682	69	69	11	11
	1v7t	682	69	69	11	11
	2d4i	682	69	69	11	11
	2d4k	682	69	69	11	11
3lyt	682	69	69	11	11	
4lyt	682	69	69	11	11	
5lym	682	69	69	11	11	
1dkj	1dkk	672	65	65	19	19
1gqn	1qfe	1482	89	72	13	13
3ezm	1l5b	543	52	52	6	6
	1m5j	543	52	52	6	6
	1m5m	543	52	52	6	6
1cpq	1nbb	690	77	65	13	11
	1rcp	690	77	60	13	11
1drf	2c2t	1046	75	74	15	15
1tjd	1eej	1234	87	72	23	17
1ecy	1ecz	805	51	51	11	11
1ejc	1ejd	2493	129	115	25	24
1uz0	1uzz	672	83	82	17	17
1fe3	1fdq	695	67	56	5	3

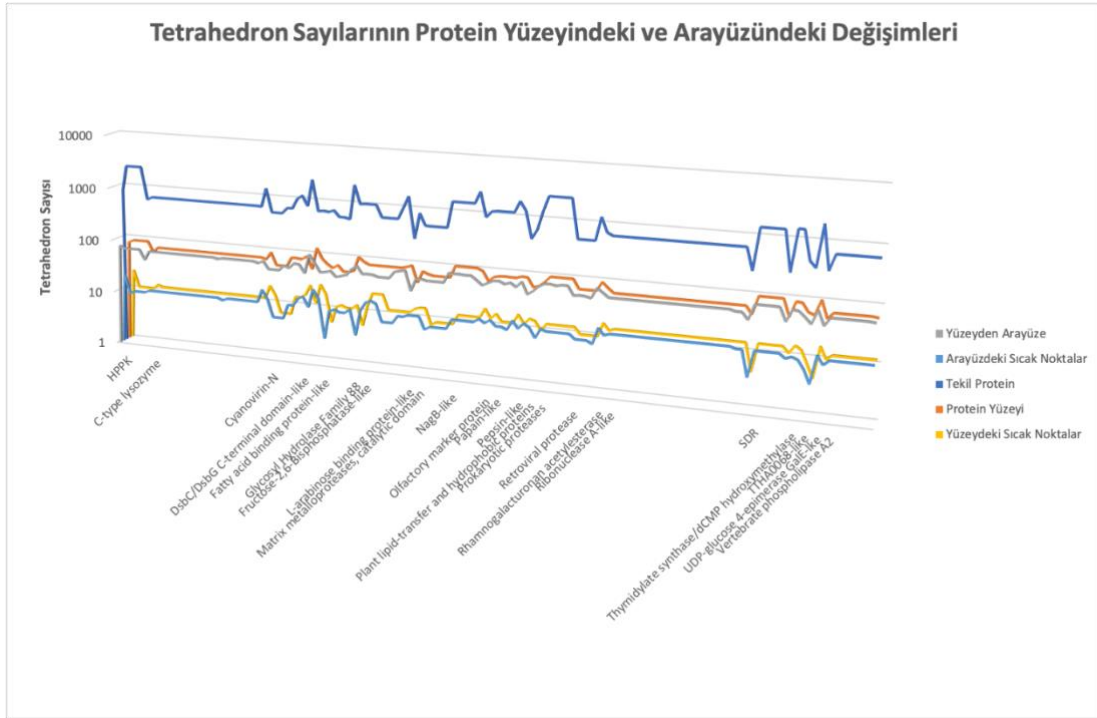
1tvq	1tw4	674	58	58	10	10
2cbr	1cbi	733	67	62	11	11
1fkb	1a7x	566	52	49	10	10
	1fki	566	52	52	10	10
2aae	2aad	537	56	56	12	12
1vd5	2d5j	2280	103	73	5	4
1ex7	1ex6	1042	86	86	12	12
1c81	1c7z	1071	77	62	21	17
	1c80	1071	77	63	21	19
	1tip	1071	77	62	21	17
1ooi	1oof	642	77	58	11	8
	1oog	642	77	58	11	8
	1t14	642	77	58	11	8
	2gte	642	77	77	11	11
1v7r	2dvn	1028	81	81	11	11
1rpj	1gud	1692	89	86	13	12
1sn1	1djt	297	44	37	14	12
1utt	1utz	872	72	63	14	12
1hrc	1crc	530	65	63	7	7
5cyt	1i54	525	62	60	8	8
	1i55	525	62	60	8	8
	1lfm	525	62	60	8	8
	3cyt	525	62	60	8	8
1cd5	1dea	1566	102	91	12	12
	1fgo	1566	102	90	12	12
	1frz	1566	102	90	12	12
	1fs5	1566	102	88	12	12
	1hor	1566	102	89	12	12
1qw6	1zvl	2563	92	75	18	14
1job	1f35	920	62	62	12	12
1kwn	1pp3	1182	75	70	15	14
1atk	1nlj	1225	79	78	11	11
	1tu6	1225	80	79	11	11
	2ftd	1225	79	71	11	10
1nqc	1npz	1233	79	77	16	15
1yx9	1psa	1961	85	66	11	11
1hpb	1hsl	1373	84	84	14	14
1rzl	1uvc	446	57	51	13	12
2acg	1f2k	658	61	60	9	8
1dua	1agj	1410	74	74	12	12
1gim	1ade	2623	94	84	12	11
	1adi	2623	94	85	12	11
	1hon	2623	94	81	12	11
	1hoo	2623	94	85	12	11
	1hop	2623	94	85	12	11
1zp8	2o4k	511	63	58	9	9
	2o4p	511	63	60	9	9
	2o4s	511	63	59	9	9

	2p3b	511	63	56	9	8
1deo	1pp4	1337	92	87	16	16
2jhy	2jhx	737	74	73	12	12
1aqp	1a2w	657	61	61	13	13
	1afk	657	61	61	13	13
	1jn4	657	61	61	13	13
	1o0f	657	61	61	13	13
	1o0h	657	61	61	13	13
	1o0m	657	61	61	13	13
	1o0n	657	61	61	13	13
	1o0o	657	61	61	13	13
	1qhc	657	61	61	13	13
	1rbb	657	61	61	13	13
	1u1b	657	61	61	13	13
	1w4o	657	61	61	13	13
	1w4q	657	61	61	13	13
	1wbu	657	61	61	13	13
	1xps	657	61	61	13	13
	1xpt	657	61	61	13	13
	1z6d	657	61	61	13	13
	1z6s	657	61	61	13	13
	2g4w	657	61	61	13	13
	2g8q	657	61	61	13	13
	2g8r	657	61	61	13	13
8rsa	657	61	57	13	12	
9rsa	657	61	57	13	12	
1yp5	1zuy	263	44	43	4	4
1b2l	1a4u	1467	93	81	13	12
	1b14	1467	93	81	13	12
	1b15	1467	93	81	13	12
	1b16	1467	93	81	13	12
	1sby	1467	93	80	13	12
1qke	1qkd	280	46	46	10	10
1ajm	1aiq	1551	86	76	14	11
1ffl	1fwm	1548	84	73	12	10
2cwy	2cxd	469	59	59	7	7
1ubi	1aar	362	54	46	4	4
1udc	2udp	2028	101	81	15	13
1utg	2utg	339	49	45	9	9
1q7a	1cl5	657	63	63	11	11
	1fb2	657	63	63	11	11
	1fv0	657	63	63	11	11
	1kpm	657	63	63	11	11
	1oxl	657	63	63	11	11
	1tp2	657	63	63	11	11
	1y38	657	63	63	11	11
2bch	2bd1	658	61	61	11	11

136 protein kompleksinin 121'inde yüzeydeki keskin tetrahedronların, diğer yüzey tetrahedronlarına göre aynı veya daha yüksek oranda arayüze geçtiği gözlemlenmiştir. Sadece 15 etkileşimde daha az geçtiği bulunmuştur. Bu 136 kompleksten 6 tanesi keskin tetrahedronların önemini gösteren durumlar olarak seçilmiş ve sonuçları Çizelge 6.3'de ayrıntılı olarak verilmiştir. Bunlardan sadece 1eej PDB kodlu olan komplekste keskin tetrahedronların arayüzde temsili azalmıştır (%83'ten %74'e düşmüştür), diğerlerinde ise keskin tetrahedronların arayüzde temsili artmıştır. Bu sebeple keskin tetrahedronlardaki rezidülerin ("kilit noktalar") protein-protein kompleksi oluşumunda önemli oldukları düşünülmektedir. Örneğin, 1gqn PDB kodlu proteinin homodimer olmasında önemli olacak rezidülerin sayısı tüm yüzey rezidü sayısı olan 97 değil, arayüzdeki "kilit noktalar" oldukları düşünülen rezidü sayısı olan 38'dir.

Çizelge 6.3. Anlamlı bir biçimde önemli olduğu düşünülen protein çiftlerinde "kilit noktaların" bulunması

Tekil PDB	Komp PDB	Tetrahedron Sayıları					Rezidü Sayıları							%	
		Tekil	Protein Yüzeği	Yüzeyden Arayüze	Yüzeydeki Keskin Tet	Arayüzdeki Keskin Tet	Tekil	Yüzey	Arayüz	Yüzeydeki Kilit Noktalar	Arayüzdeki Kilit Noktalar				
1gqn	1qfe	1482	89	72	13	13	81% 72/89	100% 13/13	252	97	86	38	38	15	+
1tjd	1eej	1234	87	72	23	17	83% 72/87	74% 17/23	216	92	85	49	43	20	-
1ejc	1ejd	2493	129	115	25	24	89% 115/129	96% 24/25	419	138	132	62	62	15	+
1vd5	2d5j	2280	103	73	5	4	71% 73/103	80% 4/5	377	112	87	19	15	4	+
1c81	1c80	1071	77	63	21	19	82% 63/77	90% 19/21	191	84	73	46	41	21	+
1yx9	1psa	1961	85	66	11	11	78% 66/85	100% 11/11	326	93	79	31	31	10	+



Şekil 6.2. Tetrahedron sayıları

6.3. Tekil Proteine Ait Tetrahedron Hacim Ortalamalarının Protein Yüzeyindeki ve Arayüzündeki Değişimleri

Her protein için tekil, yüzey ve arayüzdeki tetrahedronlara ait ortalama hacim değerleri hesaplanmış ve Çizelge 6.4'te sunulmuştur (Şekil 6.3). Bir proteinin tetrahedronlarının hacim ortalaması tüm proteinlerde hesaplanmıştır ve değerlerin ortalaması bulunmuştur. Elde edilen ortalama değerler tekil proteinlerinde $20,01 \text{ \AA}^3$, protein yüzeyinde $14,46 \text{ \AA}^3$, arayüze dönüşen protein yüzeylerinde $14,60 \text{ \AA}^3$, yüzeydeki keskin tetrahedronlarda $6,37 \text{ \AA}^3$ ve arayüze dönüşmüş keskin tetrahedronlarda $6,48 \text{ \AA}^3$ tür.

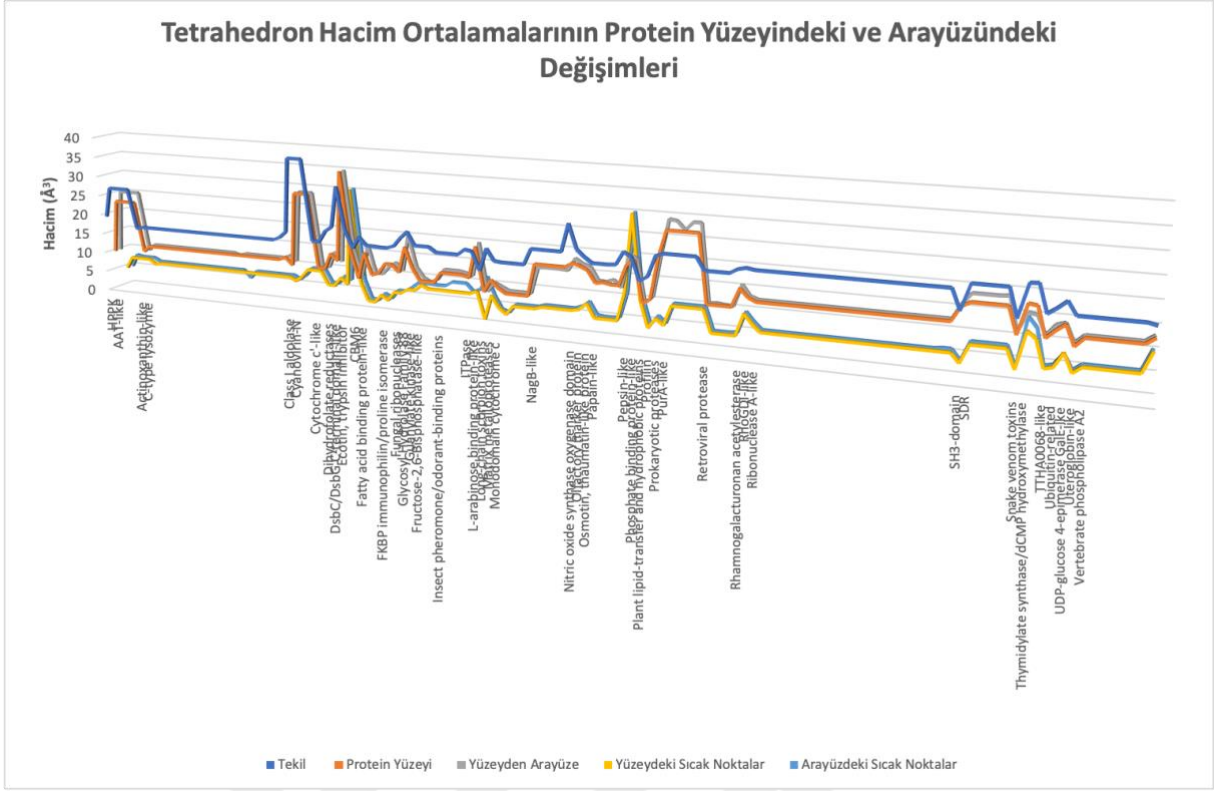
Yüzey tetrahedronlarının ortalama hacmi tüm tetrahedronların ortalama hacmi ile oranlanmış ve elde edilen oranlar tüm proteinler için ortalaması alınarak 0,71 olarak hesaplanmıştır. Arayüze dönüşen yüzey tetrahedronlarında bu değer 0,72 olmuştur. Her bir protein için yüzeydeki keskin tetrahedron hacim ortalamaları tüm protein tetrahedronlarının hacmine olan oranı hesaplanmıştır. Bu oranın bütün proteinlerdeki ortalaması 0,32'dir. Bu oran proteindeki tüm tetrahedronlar yerine yüzey tetrahedronları kullanıldığında 0,45'tir. Keskin tetrahedronların diğer yüzey tetrahedronların yarısı kadar büyüklüğe sahip oldukları ortaya çıkmıştır. Toplam proteine orana 0,32, yüzey tetrahedronlarına oranı 0,45 olarak bulunmuştur.

Çizelge 6.4. Tekil, yüzey ve arayüzdeki tetrahedronlara ait ortalama hacim değerleri

Tekil PDB	Komp PDB	Tekil Protein Hacim (Å ³)	Protein Yüzeyi Hacim (Å ³)	Protein Yüzeyinden Protein Arayüzüne Hacim (Å ³)	Yüzeydeki Keskin Tetrahedronlar Hacim (Å ³)	Arayüzdeki Keskin Tetrahedronlar Hacim (Å ³)
1f9h	1g4c	19,49	9,87	9,87	4,48	4,48
1aaw	1arg	26,96	23,16	25,37	7,28	7,61
	1asl	26,96	23,16	25,37	7,28	7,28
	1asm	26,96	23,16	25,37	7,28	7,28
	1asn	26,96	23,16	25,37	7,28	7,28
1noa	1nco	16,96	10,53	10,58	6,12	6,42
193l	1b2k	17,37	11,94	11,94	6,46	6,46
	1hf4	17,37	11,94	11,94	6,46	6,46
	1lcn	17,37	11,94	11,94	6,46	6,46
	1lj3	17,37	11,94	11,94	6,46	6,46
	1lj4	17,37	11,94	11,94	6,46	6,46
	1lje	17,37	11,94	11,94	6,46	6,46
	1ljf	17,37	11,94	11,94	6,46	6,46
	1ljg	17,37	11,94	11,94	6,46	6,46
	1ljh	17,37	11,94	11,94	6,46	6,46
	1lji	17,37	11,94	11,94	6,46	6,46
	1ljj	17,37	11,94	11,94	6,46	6,46
	1ljk	17,37	11,94	11,94	6,46	6,46
	1lkr	17,37	11,94	11,94	6,46	6,46
	1lys	17,37	11,94	11,94	6,46	6,46
	1rcm	17,37	11,94	11,46	6,46	4,64
	1uco	17,37	11,94	11,94	6,46	6,46
	1v7t	17,37	11,94	11,94	6,46	6,46
	2d4i	17,37	11,94	11,94	6,46	6,46
	2d4k	17,37	11,94	11,94	6,46	6,46
	3lyt	17,37	11,94	11,94	6,46	6,46
4lyt	17,37	11,94	11,94	6,46	6,46	
5lym	17,37	11,94	11,94	6,46	6,46	
1dkj	1dkk	18,15	12,63	12,63	5,62	5,62
1gqn	1qfe	19,66	10,84	11,49	6,94	6,94
3ezm	1l5b	37,98	29,01	29,01	8,86	8,86
	1m5j	37,98	29,01	29,01	8,86	8,86
	1m5m	37,98	29,01	29,01	8,86	8,86
1cpq	1nbb	17,97	10,82	10,48	5,48	5,63
	1rcp	17,97	10,82	10,78	5,48	5,63
1drf	2c2t	20,69	14,66	14,64	7,62	7,62
1tjd	1eej	21,93	13,12	12,46	6,18	6,89
1ecy	1ecz	31,92	35,26	35,26	30,16	30,16
1ejc	1ejd	21,28	15,99	15,36	7	7,19
1uz0	1uxz	17,18	9,14	9,49	2,5	2,5
1fe3	1fdq	20,05	15,66	17,17	2,32	2,40
1tvq	1tw4	18,1	10,32	10,32	4,21	4,21

2cbr	1cbi	18,05	10,93	10,37	3,19	3,19
1fkb	1a7x	17,96	13,43	12,26	5,33	5,33
	1fki	17,96	13,43	13,43	5,33	5,33
2aae	2aad	18,6	11,70	11,70	6,41	6,41
1vd5	2d5j	20,7	18,05	20,59	6,38	7,80
1ex7	1ex6	22,36	12,90	12,90	7,85	7,85
1c81	1c7z	19,15	10,04	10,37	7,06	7,80
	1c80	19,15	10,04	9,47	7,06	7,71
	1tip	19,15	10,04	10,37	7,06	7,80
1ooi	1oof	17,95	12,60	12,91	7,02	8,89
	1oog	17,95	12,60	12,91	7,02	8,89
	1t14	17,95	12,60	12,91	7,02	8,89
	2gte	17,95	12,60	12,60	7,02	7,02
1v7r	2dvn	19,35	12,13	12,13	7,77	7,77
1rpj	1gud	19,03	19,70	20,31	1,03	11,12
1sn1	1djt	14,64	9,03	8,59	7,08	6,59
1utt	1utz	20,05	12,02	10,97	4,2	3,36
1hrc	1crc	17,24	10,25	10,10	2,8	2,8
5cyt	1i54	17,05	99,08	9,26	4,96	4,96
	1i55	17,05	99,08	9,26	4,96	4,96
	1lfm	17,05	99,08	9,26	4,96	4,96
	3cyt	17,05	99,08	9,26	4,96	4,96
1cd5	1dea	20,9	16,84	15,61	5,53	5,53
	1fgo	20,9	16,84	15,76	5,53	5,53
	1frz	20,9	16,84	15,76	5,53	5,53
	1fs5	20,9	16,84	15,75	5,53	5,53
	1hor	20,9	16,84	15,62	5,53	5,53
1qw6	1zvl	27,56	17,96	18,59	5,93	6,20
1job	1f35	21,79	17,46	17,46	7,55	7,55
1kwn	1pp3	20,14	16,58	16,98	4,19	4,45
1atk	1nlj	18,88	13,75	13,88	4,22	4,22
	1tu6	18,88	14,10	13,75	4,22	4,22
	2ftd	18,88	13,75	14,39	4,22	4,49
1nqc	1npz	19,06	13,75	12,79	11,15	10,65
1yx9	1psa	22,16	18,66	20,46	29,76	29,76
1hpb	1hsl	21,11	20,00	20,00	9,79	9,79
1rzl	1uvc	15,76	10,26	9,62	3,25	3,51
2acg	1f2k	16,98	11,22	11,34	5,65	5,82
1dua	1agj	21,86	21,08	21,08	4,26	4,26
1gim	1ade	22,51	27,37	29,49	9,05	9,13
	1adi	22,51	27,37	29,24	9,05	9,13
	1hon	22,51	27,37	27,36	9,05	9,13
	1hoo	22,51	27,37	29,24	9,05	9,13
	1hop	22,51	27,37	29,24	9,05	9,13
1zp8	2o4k	19,49	11,38	11,11	3,63	3,63
	2o4p	19,49	11,38	11,33	3,63	3,63
	2o4s	19,49	11,38	11,09	3,63	3,63
	2p3b	19,49	11,38	11,01	3,63	3,84

1deo	1pp4	20,54	15,78	16,15	8,93	8,93
2jhy	2jhx	20,97	13,86	13,70	6,77	6,77
1aqp	1a2w	20,66	13,02	13,02	5,01	5,01
	1afk	20,66	13,02	13,02	5,01	5,01
	1jn4	20,66	13,02	13,02	5,01	5,01
	1o0f	20,66	13,02	13,02	5,01	5,01
	1o0h	20,66	13,02	13,02	5,01	5,01
	1o0m	20,66	13,02	13,02	5,01	5,01
	1o0n	20,66	13,02	13,02	5,01	5,01
	1o0o	20,66	13,02	13,02	5,01	5,01
	1qhc	20,66	13,02	13,02	5,01	5,01
	1rbb	20,66	13,02	13,02	5,01	5,01
	1ulb	20,66	13,02	13,02	5,01	5,01
	1w4o	20,66	13,02	13,02	5,01	5,01
	1w4q	20,66	13,02	13,02	5,01	5,01
	1wbu	20,66	13,02	13,02	5,01	5,01
	1xps	20,66	13,02	13,02	5,01	5,01
	1xpt	20,66	13,02	13,02	5,01	5,01
	1z6d	20,66	13,02	13,02	5,01	5,01
	1z6s	20,66	13,02	13,02	5,01	5,01
	2g4w	20,66	13,02	13,02	5,01	5,01
	2g8q	20,66	13,02	13,02	5,01	5,01
2g8r	20,66	13,02	13,02	5,01	5,01	
8rsa	20,66	13,02	13,02	5,01	5,26	
9rsa	20,66	13,02	13,02	5,01	5,26	
1yp5	1zuy	16,02	16,40	16,50	2,88	2,88
1b2l	1a4u	21,97	17,50	18,93	7,22	7,39
	1b14	21,97	17,50	18,93	7,22	7,39
	1b15	21,97	17,50	18,93	7,22	7,39
	1b16	21,97	17,50	18,93	7,22	7,39
	1sby	21,97	17,50	19,11	7,22	7,39
1qke	1qkd	15,58	11,51	11,51	2,82	2,82
1ajm	1aiq	23,14	18,37	15,94	11,43	14,02
1ffl	1fwm	23,29	18,08	15,6	9,93	11,72
2cwy	2cxd	17,09	11,43	11,43	3,71	3,71
1ubi	1aar	18,47	13,25	13,90	4,19	4,19
1udc	2udp	20,11	14,73	14,52	7,54	6,80
1utg	2utg	17,23	10,26	10,40	3,39	3,39
1q7a	1cl5	17,24	12,31	12,31	4,89	4,89
	1fb2	17,24	12,31	12,31	4,89	4,89
	1fv0	17,24	12,31	12,31	4,89	4,89
	1kpm	17,24	12,31	12,31	4,89	4,89
	1oxl	17,24	12,31	12,31	4,89	4,89
	1tp2	17,24	12,31	12,31	4,89	4,89
	1y38	17,24	12,31	12,31	4,89	4,89



Şekil 6.3. Tetrahedron hacim ortalamaları

6.4. Tekil Proteine Ait Tetrahedron Ortalama Küresellik Değerlerinin Protein Yüzeyindeki ve Arayüzündeki Değişimleri

Proteinlerin tetrahedronlarının küresellik değerlerinin ortalamaları hesaplanmıştır (Çizelge 6.5 ve Şekil 6.4) ve tüm proteinler için elde edilen bu değerlerin ortalamaları hesaplanmıştır: tekil proteinlerinde 0,47, protein yüzeyinde ve arayüze dönüşen protein yüzeylerinde 0,39, yüzeydeki keskin tetrahedronlarda ve arayüze dönüşmüş keskin tetrahedronlarda 0,22. En büyük değer regüler tetrahedronda 0,67 olur. Bu değerden küçük olan değerler tetrahedronun irregüler olduğunu gösterir. Keskin tetrahedronların önemli derecede diğer tetrahedronlardan daha irregüler oldukları görülmüştür (0,47 ve 0,39'a karşılık 0,22 küresellik değeri)



Şekil 6.4. En düşük ortalama küresellik değerine sahip (0,41) 3ezm PDB kodlu protein (solda), en büyük ortalama küresellik değerine sahip (0,50) 1vd5 PDB kodlu protein (sağda) (Sehnal ve ark. 2021)

Çizelge 6.5. Tekil, yüzey ve arayüzdeki tetrahedronlarına ait sayı ortalama küresellik değerleri

Tekil PDB	Kompleks PDB	Tekil Protein	Protein Yüzeyi	Protein Yüzeyinden Protein Arayüzüne	Yüzeydeki Sıcak Noktalar	Arayüzdeki Sıcak Noktalar
1f9h	1g4c	0,47	0,34	0,34	0,18	0,18
1aaw	1arg	0,48	0,49	0,54	0,22	0,22
	1asl	0,48	0,49	0,54	0,22	0,22
	1asm	0,48	0,49	0,54	0,22	0,22
	1asn	0,48	0,49	0,54	0,22	0,22
1noa	1nco	0,45	0,21	0,21	0,11	0,12
193l	1b2k	0,48	0,40	0,40	0,19	0,19
	1hf4	0,48	0,40	0,40	0,19	0,19
	1lcn	0,48	0,40	0,40	0,19	0,19
	1lj3	0,48	0,40	0,40	0,19	0,19
	1lj4	0,48	0,40	0,40	0,19	0,19
	1lje	0,48	0,40	0,40	0,19	0,19
	1ljf	0,48	0,40	0,40	0,19	0,19
	1ljg	0,48	0,40	0,40	0,19	0,19
	1ljh	0,48	0,40	0,40	0,19	0,19
	1lji	0,48	0,40	0,40	0,19	0,19
	1ljj	0,48	0,40	0,40	0,19	0,19
	1ljk	0,48	0,40	0,40	0,19	0,19
	1lkr	0,48	0,40	0,40	0,19	0,19
	1lys	0,48	0,40	0,40	0,19	0,19
	1rcm	0,48	0,40	0,39	0,19	0,18
	1uco	0,48	0,40	0,40	0,19	0,19
	1v7t	0,48	0,40	0,40	0,19	0,19
	2d4i	0,48	0,40	0,40	0,19	0,19
	2d4k	0,48	0,40	0,40	0,19	0,19
	3lyt	0,48	0,40	0,40	0,19	0,19
4lyt	0,48	0,40	0,40	0,19	0,19	
5lym	0,48	0,40	0,40	0,19	0,19	
1dkj	1dkk	0,48	0,39	0,39	0,23	0,23

1gqn	1qfe	0,48	0,39	0,34	0,21	0,21
3ezm	1l5b	0,41	0,53	0,53	0,36	0,36
	1m5j	0,41	0,53	0,53	0,36	0,36
	1m5m	0,41	0,53	0,53	0,36	0,36
1cpq	1nbb	0,47	0,36	0,31	0,23	0,21
	1rcp	0,47	0,36	0,33	0,23	0,21
1drf	2c2t	0,47	0,39	0,38	0,14	0,14
1tjd	1eej	0,47	0,40	0,36	0,21	0,27
1ecy	1ecz	0,43	0,35	0,35	0,17	0,17
1ejc	1ejd	0,49	0,31	0,36	0,18	0,20
1uz0	1uxz	0,47	0,37	0,35	0,13	0,13
1fe3	1fdq	0,47	0,37	0,41	0,07	0,04
1tvq	1tw4	0,45	0,29	0,29	0,1	0,1
2cbr	1cbi	0,45	0,28	0,29	0,09	0,09
1fkb	1a7x	0,46	0,40	0,37	0,24	0,24
	1fki	0,46	0,40	0,40	0,24	0,24
2aae	2aad	0,47	0,32	0,32	0,12	0,12
1vd5	2d5j	0,5	0,41	0,38	0,24	0,33
1ex7	1ex6	0,48	0,38	0,38	0,19	0,19
1c81	1c7z	0,48	0,35	0,33	0,22	0,16
	1c80	0,48	0,35	0,33	0,22	0,21
	1tip	0,48	0,35	0,33	0,22	0,16
1ooi	1oof	0,49	0,45	0,44	0,33	0,36
	1oog	0,49	0,45	0,44	0,33	0,36
	1t14	0,49	0,45	0,44	0,33	0,36
	2gte	0,49	0,45	0,45	0,33	0,33
1v7r	2dvn	0,48	0,35	0,35	0,25	0,25
1rpj	1gud	0,49	0,47	0,46	0,25	0,18
1sn1	1djt	0,46	0,30	0,30	0,34	0,31
1utt	1utz	0,48	0,33	0,31	0,16	0,13
1hrc	1crc	0,47	0,37	0,38	0,14	0,14
5cyt	1i54	0,47	0,32	0,36	0,24	0,24
	1i55	0,47	0,32	0,36	0,24	0,24
	1lfm	0,47	0,32	0,36	0,24	0,24
	3cyt	0,47	0,32	0,36	0,24	0,24
1cd5	1dea	0,49	0,45	0,44	0,2	0,2
	1fgo	0,49	0,45	0,43	0,2	0,2
	1frz	0,49	0,45	0,43	0,2	0,2
	1fs5	0,49	0,45	0,39	0,2	0,2
	1hor	0,49	0,45	0,39	0,2	0,2
1qw6	1zvl	0,48	0,35	0,28	0,06	0,09
1job	1f35	0,46	0,41	0,41	0,19	0,19
1kwn	1pp3	0,47	0,40	0,38	0,15	0,13
1atk	1nlj	0,49	0,36	0,38	0,18	0,18
	1tu6	0,49	0,31	0,36	0,18	0,18
	2ftd	0,49	0,36	0,36	0,18	0,16
1nqc	1npz	0,49	0,36	0,31	0,27	0,32
1yx9	1psa	0,48	0,48	0,50	0,45	0,45

1hpb	1hsl	0,49	0,64	0,64	0,39	0,39
1rzl	1uvc	0,49	0,37	0,40	0,12	0,13
2acg	1f2k	0,48	0,35	0,34	0,24	0,22
1dua	1agj	0,49	0,35	0,35	0,11	0,11
1gim	1ade	0,49	0,48	0,63	0,3	0,28
	1adi	0,49	0,48	0,53	0,3	0,28
	1hon	0,49	0,48	0,53	0,3	0,28
	1hoo	0,49	0,48	0,53	0,3	0,28
	1hop	0,49	0,48	0,53	0,3	0,28
1zpz8	2o4k	0,45	0,32	0,32	0,12	0,12
	2o4p	0,45	0,32	0,32	0,12	0,12
	2o4s	0,45	0,32	0,31	0,12	0,12
	2p3b	0,45	0,32	0,32	0,12	0,18
1deo	1pp4	0,49	0,39	0,42	0,26	0,26
2jhy	2jhx	0,46	0,42	0,42	0,25	0,25
1aqp	1a2w	0,48	0,34	0,34	0,25	0,25
	1afk	0,48	0,34	0,34	0,25	0,25
	1jn4	0,48	0,34	0,34	0,25	0,25
	1o0f	0,48	0,34	0,34	0,25	0,25
	1o0h	0,48	0,34	0,34	0,25	0,25
	1o0m	0,48	0,34	0,34	0,25	0,25
	1o0n	0,48	0,34	0,34	0,25	0,25
	1o0o	0,48	0,34	0,34	0,25	0,25
	1qhc	0,48	0,34	0,34	0,25	0,25
	1rbb	0,48	0,34	0,34	0,25	0,25
	1u1b	0,48	0,34	0,34	0,25	0,25
	1w4o	0,48	0,34	0,34	0,25	0,25
	1w4q	0,48	0,34	0,34	0,25	0,25
	1wbu	0,48	0,34	0,34	0,25	0,25
	1xps	0,48	0,34	0,34	0,25	0,25
	1xpt	0,48	0,34	0,34	0,25	0,25
	1z6d	0,48	0,34	0,34	0,25	0,25
	1z6s	0,48	0,34	0,34	0,25	0,25
	2g4w	0,48	0,34	0,34	0,25	0,25
	2g8q	0,48	0,34	0,34	0,25	0,25
	2g8r	0,48	0,34	0,34	0,25	0,25
8rsa	0,48	0,34	0,33	0,25	0,25	
9rsa	0,48	0,34	0,33	0,25	0,25	
1yp5	1zuy	0,47	0,51	0,51	0,07	0,07
1b2l	1a4u	0,48	0,38	0,42	0,23	0,23
	1b14	0,48	0,38	0,42	0,23	0,23
	1b15	0,48	0,38	0,42	0,23	0,23
	1b16	0,48	0,38	0,42	0,23	0,23
	1sby	0,48	0,38	0,38	0,23	0,23
1qke	1qkd	0,43	0,34	0,34	0,1	0,1
1ajm	1aiq	0,48	0,57	0,51	0,18	0,18
1ffl	1fwm	0,48	0,41	0,43	0,23	0,22
2cwy	2cxd	0,49	0,37	0,37	0,22	0,22

1ubi	1aar	0,46	0,43	0,42	0,21	0,21
1udc	2udp	0,49	0,37	0,34	0,15	0,16
1utg	2utg	0,44	0,40	0,44	0,25	0,25
1q7a	1cl5	0,46	0,39	0,39	0,22	0,22
	1fb2	0,46	0,39	0,39	0,22	0,22
	1fv0	0,46	0,39	0,39	0,22	0,22
	1kpm	0,46	0,39	0,39	0,22	0,22
	1oxl	0,46	0,39	0,39	0,22	0,22
	1tp2	0,46	0,39	0,39	0,22	0,22
	1y38	0,46	0,39	0,39	0,22	0,22

6.5. Rezidü Sayıları

Tekil proteini oluşturan rezidülerin ortalama %51'i proteinin yüzey rezidülerini oluştururken, ortalama %22'si "kilit noktalardan" oluşmaktadır. Tekil proteinleri oluşturan rezidülerin ortalama %50'si protein arayüzüne yerleşmektedir. Hem bu yüzey rezidülerinin hem de yüzeydeki "kilit noktaların" %97'si protein arayüzüne yerleşmektedir. Arayüz rezidülerinin ortalama %43'ü "kilit noktalardan" oluşmaktadır.

Çizelge 6.6. Tekil protein ve komplekslerinin arayüz bölgelerindeki rezidü sayıları

Tekil PDB	Kompleks PDB	Tekil Protein Protein	Protein Yüzey-Arayüz		Keskin Tetrahedronlar	
			Yüzey	Arayüz	Yüzey	Arayüz
1f9h	1g4c	158	83	83	45	45
1aaw	1arg	396	95	85	33	31
1aaw	1asl	396	95	85	33	33
1aaw	1asm	396	95	85	33	33
1aaw	1asn	396	95	85	33	33
1noa	1nco	113	64	60	33	33
193l	1b2k	129	78	78	32	28
193l	1hf4	129	78	78	32	32
193l	1lcn	129	78	78	32	32
193l	1lj3	129	78	78	32	32
193l	1lj4	129	78	78	32	32
193l	1lje	129	78	78	32	32
193l	1ljf	129	78	78	32	32
193l	1ljg	129	78	78	32	32
193l	1ljh	129	78	78	32	32
193l	1lji	129	78	78	32	32
193l	1ljj	129	78	78	32	32
193l	1ljk	129	78	78	32	32
193l	1lkr	129	78	78	32	32
193l	1lys	129	78	78	32	32
193l	1rcm	129	78	77	32	32
193l	1uco	129	78	78	32	32
193l	1v7t	129	78	78	32	32

193l	2d4i	129	78	78	32	32
193l	2d4k	129	78	78	32	32
193l	3lyt	129	78	78	32	32
193l	4lyt	129	78	78	32	32
193l	5lym	129	78	78	32	32
1dkj	1dkk	129	78	77	39	39
1gqn	1qfe	252	97	86	38	38
3ezm	1l5b	101	54	54	16	16
3ezm	1m5j	101	54	54	16	16
3ezm	1m5m	101	54	54	16	16
1cpq	1nbb	129	77	69	37	30
1cpq	1rep	129	77	69	37	30
1drf	2c2t	186	81	80	41	41
1tjd	1eej	216	92	84	49	43
1ecy	1ecz	142	53	53	26	26
1ejc	1ejd	419	138	132	62	62
1uz0	1uxz	131	84	83	43	43
2cbr	1cbi	125	71	69	27	27
1fe3	1fdq	136	72	65	20	12
1tvq	1tw4	131	65	65	26	26
1fkb	1a7x	107	53	52	24	24
1fkb	1fki	107	53	53	24	24
2aae	2aad	104	59	59	29	29
1vd5	2d5j	377	112	87	19	15
1ex7	1ex6	186	89	89	32	32
1c81	1c7z	191	84	76	46	38
1c81	1c80	191	84	73	46	41
1c81	1tip	191	84	76	46	38
1ooi	1oof	124	83	70	31	26
1ooi	1oog	124	83	70	31	26
1ooi	1t14	124	83	70	31	26
1ooi	2gte	124	83	83	31	31
1v7r	2dvn	186	82	82	35	35
1rpj	1gud	288	100	99	38	37
1sn1	1djt	64	47	43	29	27
1utt	1utz	159	77	71	35	31
1hrc	1crc	104	70	68	22	22
5cyt	1i54	103	68	68	23	23
5cyt	1i55	103	68	68	23	23
5cyt	1lfm	103	68	68	23	23
5cyt	3cyt	103	68	68	23	23
1cd5	1dea	266	107	101	34	34
1cd5	1fqo	266	107	101	34	34
1cd5	1frz	266	107	101	34	34
1cd5	1fs5	266	107	99	34	34
1cd5	1hor	266	107	100	34	34
1qw6	1zvl	420	96	84	46	39
1job	1f35	162	67	67	30	30

1kwn	1pp3	207	83	81	38	36
1atk	1nlj	215	84	84	31	31
1nqc	1npz	215	85	84	40	39
1atk	1tu6	217	84	84	31	31
1atk	2ftd	215	84	80	31	31
1yx9	1psa	326	93	76	31	31
1hpb	1hsl	238	97	97	39	39
1rzl	1uvc	91	64	61	32	30
2acg	1f2k	125	67	67	29	26
1dua	1agj	242	82	82	33	33
1gim	1ade	431	101	98	36	33
1gim	1adi	431	101	98	36	33
1gim	1hon	431	101	95	36	33
1gim	1hoo	431	101	98	36	33
1gim	1hop	431	101	98	36	33
1zp8	2o4k	99	64	63	25	25
1zp8	2o4p	99	64	64	25	25
1zp8	2o4s	99	64	64	25	25
1zp8	2p3b	99	64	62	25	22
1deo	1pp4	233	100	98	41	41
2jhy	2jhx	138	79	79	30	30
1aqp	1a2w	124	67	67	35	35
1aqp	1afk	124	67	67	35	35
1aqp	1jn4	124	67	67	35	35
1aqp	1o0f	124	67	67	35	35
1aqp	1o0h	124	67	67	35	35
1aqp	1o0m	124	67	67	35	35
1aqp	1o0n	124	67	67	35	35
1aqp	1o0o	124	67	67	35	35
1aqp	1qhc	124	67	67	35	35
1aqp	1rbb	124	67	67	35	35
1aqp	1u1b	124	67	67	35	35
1aqp	1w4o	124	67	67	35	35
1aqp	1w4q	124	67	67	35	35
1aqp	1wbu	124	67	67	35	35
1aqp	1xps	124	67	67	35	35
1aqp	1xpt	124	67	67	35	35
1aqp	1z6d	124	67	67	35	35
1aqp	1z6s	124	67	67	35	35
1aqp	2g4w	124	67	67	35	35
1aqp	2g8q	124	67	67	35	35
1aqp	2g8r	124	67	67	35	35
1aqp	8rsa	124	67	64	35	34
1aqp	9rsa	124	67	64	35	34
1yp5	1zuy	58	47	46	15	15
1b2l	1a4u	254	104	90	37	36
1b2l	1b14	254	104	90	37	36
1b2l	1b15	254	104	90	37	36

1b2l	1b16	254	104	90	37	36
1b2l	1sby	254	104	90	37	36
1qke	1qkd	62	44	44	28	28
1ajm	1aiq	264	98	88	40	33
1ffl	1fwm	264	95	85	38	32
2cwy	2cxd	94	59	59	22	22
1ubi	1aar	76	55	51	11	11
1udc	2udp	338	115	98	43	37
1utg	2utg	70	51	49	24	24
1q7a	1cl5	121	68	68	31	31
1q7a	1fb2	121	68	68	31	31
1q7a	1fv0	121	68	68	31	31
1q7a	1kpm	121	68	68	31	31
1q7a	1oxl	121	68	68	31	31
1q7a	1tp2	121	68	68	31	31
1q7a	1y38	121	68	68	31	31
2bch	2bd1	123	66	66	27	27

6.6. Arayüzdeki “Kilit Noktaların” Rezidü Kompozisyonu

Her bir protein kompleksindeki arayüzde bulunan “kilit noktaları” oluşturan rezidülerin bulunma sıklıkları Çizelge 6.7’de gösterilmiştir.

Çizelge 6.7. Arayüzdeki “kilit noktaların” rezidü içeriği

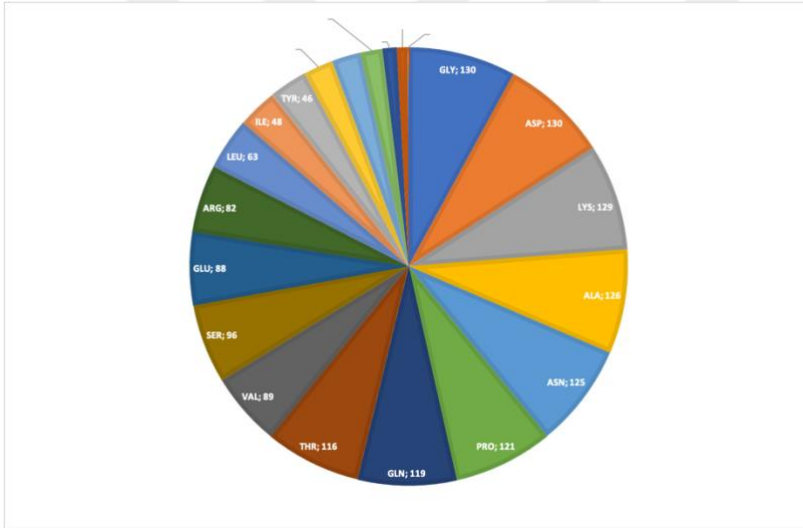
PDB Tekil	PDB Komp	GLY	ASP	LYS	ALA	ASN	PRO	GLN	THR	VAL	SER	GLU	ARG	LEU	ILE	TYR	PHE	MET	HIS	CYS	TRP	IAS
1f9h	1g4c	0,07	0,07	0,09	0,11	0,09	0,09	0,05	0,07	0,05	0,05	0,14	0,07	0,02					0,05			
1aaw	1arg	0,03	0,13	0,03	0,17	0,10	0,03	0,07	0,07			0,27	0,07					0,03				
1aaw	1asl	0,06	0,13	0,03	0,16	0,09	0,03	0,06	0,06			0,25	0,09					0,03				
1aaw	1asm	0,06	0,13	0,03	0,16	0,09	0,03	0,06	0,06			0,25	0,09					0,03				
1aaw	1asn	0,06	0,13	0,03	0,16	0,09	0,03	0,06	0,06			0,25	0,09					0,03				
1noa	1nco	0,14	0,11	0,04	0,25	0,04	0,04	0,04	0,11	0,07	0,11		0,04								0,04	
193l	1b2k	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
1dkj	1dkk	0,23	0,08	0,05	0,05	0,13	0,05	0,05	0,05	0,03	0,10		0,10	0,05	0,03							
193l	1hf4	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
193l	1lcn	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
193l	1lj3	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
193l	1lj4	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
193l	1lje	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
193l	1ljf	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
193l	1ljg	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
193l	1ljh	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
193l	1lji	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
193l	1ljj	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
193l	1ljk	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
193l	1lkr	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
193l	1lys	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
193l	1rcm	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
193l	1uco	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
193l	1v7t	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
193l	2d4i	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
193l	2d4k	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
193l	3lyt	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
193l	4lyt	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							

193l	5lym	0,30	0,07	0,03	0,03	0,20	0,03	0,03	0,10	0,03			0,10	0,03	0,03							
1gqn	1qfe		0,03	0,05	0,24	0,08	0,05	0,08	0,05		0,13	0,08	0,03	0,03	0,03			0,05	0,08			
3ezm	1i5b	0,19	0,19		0,06	0,19		0,13	0,06		0,06	0,06		0,06								
3ezm	1m5j	0,19	0,19		0,06	0,19		0,13	0,06		0,06	0,06		0,06								
3ezm	1m5m	0,19	0,19		0,06	0,19		0,13	0,06		0,06	0,06		0,06								
1cpq	1nbb	0,07	0,28	0,07	0,24	0,03	0,07		0,03	0,03		0,14		0,03								
1cpq	1rcp	0,07	0,28	0,07	0,24	0,03	0,07		0,03	0,03		0,14		0,03								
1drf	2c2t	0,10	0,10	0,12		0,12	0,12	0,05	0,02	0,05	0,02	0,24	0,05									
1tjd	1eej	0,09	0,12	0,14	0,19		0,05	0,09	0,02	0,02	0,07	0,05		0,02	0,02	0,02		0,05	0,02	0,02		
1ecy	1ecz	0,08	0,08	0,15	0,04	0,04	0,12	0,04	0,04	0,08	0,04	0,19	0,04					0,08				
1ejc	1ejd	0,21	0,02	0,08	0,03	0,06	0,03	0,02	0,06	0,05	0,05	0,26	0,05						0,06			0,02
1uz0	1uxz	0,26	0,02		0,12	0,09	0,07	0,09	0,16	0,02	0,12	0,02							0,02			
2cbr	1cbi	0,15	0,11	0,04	0,07	0,15	0,04	0,04			0,04	0,22	0,04	0,04				0,04	0,04			
1fe3	1fdq	0,17	0,08	0,08		0,08			0,17			0,33						0,08				
1tvq	1tw4	0,12	0,15	0,19	0,04	0,08	0,08	0,04	0,08		0,04	0,12	0,04						0,04			
1fkb	1a7x	0,21	0,08	0,21	0,04	0,04	0,13		0,04			0,08	0,13						0,04			
1fkb	1fki	0,21	0,08	0,21	0,04	0,04	0,13		0,04			0,08	0,13						0,04			
2aae	2aad	0,17	0,10	0,03	0,07	0,07	0,03		0,07		0,34	0,07				0,03						
1vd5	2d5j	0,13	0,27		0,07	0,07	0,07	0,07			0,07	0,13	0,07						0,07			
1ex7	1ex6	0,16	0,19	0,06	0,16	0,06	0,06		0,06	0,03	0,09	0,13										
1c81	1c7z	0,05	0,11	0,11	0,08	0,05	0,05	0,05		0,03	0,11	0,16	0,05	0,03		0,03		0,05	0,05			
1c81	1c80	0,07	0,10	0,10	0,07	0,05	0,05	0,05		0,05	0,12	0,15	0,05	0,02		0,02		0,05	0,05			
1c81	1tip	0,05	0,11	0,11	0,08	0,05	0,05	0,05		0,03	0,11	0,16	0,05	0,03		0,03		0,05	0,05			
1ooi	1oof	0,08	0,04	0,19	0,08	0,08	0,12	0,08				0,19						0,04	0,08	0,04		
1ooi	1oog	0,08	0,04	0,19	0,08	0,08	0,12	0,08				0,19						0,04	0,08	0,04		
1ooi	1t14	0,08	0,04	0,19	0,08	0,08	0,12	0,08				0,19						0,04	0,08	0,04		
1ooi	2gte	0,10	0,03	0,23	0,06	0,10	0,10	0,10				0,16						0,03	0,06	0,03		
1v7r	2dvn	0,14	0,06	0,17	0,03	0,03	0,03	0,03	0,09	0,03		0,23	0,03	0,03	0,03	0,03	0,03	0,03	0,03			
1rpi	1gud	0,11	0,08	0,24	0,11	0,05	0,08	0,11			0,11	0,05	0,03		0,03							
1sn1	1djt	0,15	0,11	0,11	0,04	0,07	0,07	0,04		0,04	0,04	0,07	0,04		0,04	0,07			0,07			0,04
1utt	1utz	0,16	0,13	0,13	0,06	0,13	0,06		0,06	0,03	0,06	0,03	0,10						0,03			
1hrc	1erc	0,23	0,05	0,32			0,05	0,05	0,14			0,14			0,05							
5cyt	1i54	0,26	0,09	0,35	0,04	0,04		0,04		0,04	0,09	0,04										
5cyt	1i55	0,26	0,09	0,35	0,04	0,04		0,04		0,04	0,09	0,04										
5cyt	1lfm	0,26	0,09	0,35	0,04	0,04		0,04		0,04	0,09	0,04										
5cyt	3cyt	0,26	0,09	0,35	0,04	0,04		0,04		0,04	0,09	0,04										
1cd5	1dea	0,06	0,13	0,13	0,09	0,03	0,03	0,03	0,06	0,03	0,06	0,06	0,09	0,03		0,03	0,03	0,06		0,03		

1cd5	1fgo	0,06	0,13	0,13	0,09	0,03	0,03	0,03	0,06	0,03	0,06	0,06	0,09	0,03		0,03	0,03	0,06		0,03		
1cd5	1frz	0,06	0,13	0,13	0,09	0,03	0,03	0,03	0,06	0,03	0,06	0,06	0,09	0,03		0,03	0,03	0,06		0,03		
1cd5	1fs5	0,06	0,13	0,13	0,09	0,03	0,03	0,03	0,06	0,03	0,06	0,06	0,09	0,03		0,03	0,03	0,06		0,03		
1cd5	1hor	0,06	0,13	0,13	0,09	0,03	0,03	0,03	0,06	0,03	0,06	0,06	0,09	0,03		0,03	0,03	0,06		0,03		
1qw6	1zvl	0,10	0,15	0,21	0,03	0,08	0,08		0,05	0,03	0,08	0,08	0,03		0,03				0,03		0,05	
1job	1f35	0,03	0,20	0,17	0,10	0,03	0,03	0,17				0,07	0,07	0,07	0,07							
1kwn	1pp3	0,17	0,06	0,06	0,06	0,06	0,06	0,06	0,14		0,11	0,11	0,03			0,08	0,03					
1atk	1nlj	0,10	0,10	0,23	0,06	0,10	0,03		0,06	0,03	0,06	0,06	0,06				0,03	0,06				
1nqc	1npz	0,10	0,10	0,15	0,03	0,10	0,10	0,03	0,05		0,13	0,05	0,05	0,03		0,03		0,03	0,03			
1atk	1tu6	0,10	0,10	0,23	0,06	0,10	0,03		0,06	0,03	0,06	0,06	0,06				0,03	0,06				
1atk	2ftd	0,10	0,10	0,21	0,07	0,07	0,03		0,07	0,03	0,07	0,07	0,07				0,03	0,07				
1yx9	1psa	0,13	0,16		0,03	0,03	0,03	0,03	0,06		0,29	0,03	0,03	0,03	0,06	0,03				0,03		
1hpb	1hsl	0,13	0,10	0,21	0,15	0,08	0,05	0,10		0,03		0,10	0,03				0,03					
1rzl	1uvc	0,20	0,03	0,07	0,17	0,07	0,03	0,03	0,10		0,23		0,03		0,03							
2acg	1f2k	0,27	0,04	0,12	0,04	0,04		0,08	0,15		0,12	0,04	0,04		0,04	0,04						
1dua	1agj	0,09	0,09	0,18	0,03	0,15	0,03	0,03		0,03	0,12	0,15	0,06			0,03						
1gim	1ade	0,06	0,12	0,03	0,06	0,06		0,12	0,03		0,09	0,21	0,12			0,06	0,03					
1gim	1adi	0,06	0,12	0,03	0,06	0,06		0,12	0,03		0,09	0,21	0,12			0,06	0,03					
1gim	1hon	0,06	0,12	0,03	0,06	0,06		0,12	0,03		0,09	0,21	0,12			0,06	0,03					
1gim	1hoo	0,06	0,12	0,03	0,06	0,06		0,12	0,03		0,09	0,21	0,12			0,06	0,03					
1gim	1hop	0,06	0,12	0,03	0,06	0,06		0,12	0,03		0,09	0,21	0,12			0,06	0,03					
1zp8	2o4k	0,20		0,04		0,08	0,16	0,12	0,08		0,04	0,04	0,04		0,04		0,08	0,04			0,04	
1zp8	2o4p	0,20		0,04		0,08	0,16	0,12	0,08		0,04	0,04	0,04		0,04		0,08	0,04			0,04	
1zp8	2o4s	0,20		0,04		0,08	0,16	0,12	0,08		0,04	0,04	0,04		0,04		0,08	0,04			0,04	
1zp8	2p3b	0,18		0,05		0,09	0,14	0,14	0,09		0,05	0,05			0,05		0,09	0,05			0,05	
1deo	1pp4	0,20			0,05	0,05	0,02		0,27	0,12	0,15	0,05		0,05	0,02					0,02		
2jhy	2jhx	0,17	0,20	0,13	0,03		0,07		0,03	0,03	0,10	0,10			0,03			0,07			0,03	
1aqp	1a2w	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03						
1aqp	1afk	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03						
1aqp	1jn4	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03						
1aqp	1o0f	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03						
1aqp	1o0h	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03						
1aqp	1o0m	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03						
1aqp	1o0n	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03						
1aqp	1o0o	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03						
1aqp	1qhc	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03						
1aqp	1rbb	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03						

1aqp	1u1b	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03					
1aqp	1w4o	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03					
1aqp	1w4q	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03					
1aqp	1wbu	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03					
1aqp	1xps	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03					
1aqp	1xpt	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03					
1aqp	1z6d	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03					
1aqp	1z6s	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03					
1aqp	2g4w	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03					
1aqp	2g8q	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03					
1aqp	2g8r	0,07	0,07	0,10	0,10	0,14	0,07	0,07	0,03	0,07	0,24					0,03					
1aqp	8rsa	0,07	0,07	0,11	0,11	0,14	0,04	0,07	0,04	0,07	0,25					0,04					
1aqp	9rsa	0,07	0,07	0,11	0,11	0,14	0,04	0,07	0,04	0,07	0,25					0,04					
1yp5	1zuy	0,13	0,07	0,07	0,07		0,13		0,07		0,20	0,07	0,07			0,07			0,07		
1b2l	1a4u		0,03	0,03	0,08		0,17	0,03	0,14	0,08	0,06	0,14	0,06	0,03	0,06				0,08		0,03
1b2l	1b14		0,03	0,03	0,08		0,17	0,03	0,14	0,08	0,06	0,14	0,06	0,03	0,06				0,08		0,03
1b2l	1b15		0,03	0,03	0,08		0,17	0,03	0,14	0,08	0,06	0,14	0,06	0,03	0,06				0,08		0,03
1b2l	1b16		0,03	0,03	0,08		0,17	0,03	0,14	0,08	0,06	0,14	0,06	0,03	0,06				0,08		0,03
1b2l	1sby		0,03	0,03	0,08		0,17	0,03	0,14	0,08	0,06	0,14	0,06	0,03	0,06				0,08		0,03
1qke	1qkd	0,11		0,07		0,04	0,11	0,07	0,14	0,04	0,18	0,07	0,07		0,04					0,07	
1ajm	1aiq	0,09	0,19	0,06	0,06	0,09	0,09		0,03	0,03	0,03	0,13	0,06	0,03	0,03		0,03		0,03		
1ffl	1fwm	0,06	0,19	0,06	0,06	0,09	0,09	0,03	0,03	0,03	0,03	0,09	0,09	0,03	0,03		0,03		0,03		
2cwy	2cxd	0,18	0,05	0,05	0,09		0,05		0,05		0,05	0,09	0,27	0,09						0,05	
1ubi	1aar	0,27	0,09	0,18				0,09				0,18	0,09					0,09			
1udc	2udp	0,05	0,16	0,16	0,11	0,08	0,11	0,05			0,03	0,05	0,05	0,03			0,03	0,03	0,05		
1utg	2utg	0,04	0,08	0,13	0,04	0,04	0,13	0,08	0,21	0,04	0,04	0,13		0,04							
1q7a	1cl5	0,10	0,10	0,23	0,03	0,03	0,06	0,03	0,06		0,06	0,06		0,16			0,03			0,03	
1q7a	1fb2	0,10	0,10	0,23	0,03	0,03	0,06	0,03	0,06		0,06	0,06		0,16			0,03			0,03	
1q7a	1fv0	0,10	0,10	0,23	0,03	0,03	0,06	0,03	0,06		0,06	0,06		0,16			0,03			0,03	
1q7a	1kpm	0,10	0,10	0,23	0,03	0,03	0,06	0,03	0,06		0,06	0,06		0,16			0,03			0,03	
1q7a	1oxl	0,10	0,10	0,23	0,03	0,03	0,06	0,03	0,06		0,06	0,06		0,16			0,03			0,03	
1q7a	1y38	0,10	0,10	0,23	0,03	0,03	0,06	0,03	0,06		0,06	0,06		0,16			0,03			0,03	
1q7a	1tp2	0,10	0,10	0,23	0,03	0,03	0,06	0,03	0,06		0,06	0,06		0,16			0,03			0,03	
2bch	2bd1	0,04	0,11	0,11		0,19	0,04	0,04	0,07	0,07	0,15	0,07		0,07				0,04			

136 protein kompleksi incelendiğinde glisin130 protein kompleksinde, aspartat 130 protein kompleksinde, lizin 129 protein kompleksinde, alanin 126 protein kompleksinde, asparjin 125 protein kompleksinde, prolin 121 protein kompleksinde, glutamin 119 protein kompleksinde ve treonin 116 protein kompleksinde, serin 96 protein kompleksinde, valin 89 protein kompleksinde, glutamat 88 protein kompleksinde, arjinin 82 protein kompleksinde, izolösün 48 protein kompleksinde, lösün 46 protein kompleksinde, tirozin 46 protein kompleksinde, fenilalanin 36 protein kompleksinde, metiyonin 35 protein kompleksinde, histidin 27 protein kompleksinde, sistein 17 protein kompleksinde ve triptofan 13 protein kompleksinde kilit nokta olarak bulunmaktadır. Enteresan bir şekilde sadece bir tane komplekste Beta-L-Aspartik asit sıcak arayüz rezidüsü olarak bulunmuştur. Geriye kalan serin, valin, glutamat, arjinin ve lösünün protein komplekslerinde sırasıyla görülme sıklıkları; 96, 89, 88, 82 ve 46 olarak bulunmuştur (Şekil 6.5).



Şekil 6.5. Protein kompleksinde bulunan sıcak noktaların rezidülerinin dağılımı

Her bir protein kompleksinin arayüzünde bulunan “kilit noktalara” ait rezidüler incelendiğinde glisinin 40, lizinin 36, serinin 30, glutamatın 22, aspartatın 17, prolinin 5, asparajinin 4, alaninin 3, tirozinin 2 ve arjininin 1 defa arayüzde en sık rastlanan rezidü tipi olduğu görülmüştür. Glisinin loop bölgelerinde sık bulunduğu bilinmektedir, dolayısıyla keskin tetrahedronlarda çok rastlanması glisinin durumunu teyit etmektedir.

7. COMSIN VERİ TABANINDA BULUNAN RETROVİRAL PROTEAZ PROTEİN AİLESİNE AİT PROTEİN KOMPLEKSLERİNİN ANALİZİ

AIDS pandemisi dünya çapında yayılmaya devam ederken, AIDS hastalığına neden olan (insan immün yetmezlik virüsü) HIV'in temel enzimlerinin yeni ve güçlü inhibitörlerinin geliştirilmesi, bilim insanlarının odak noktası haline gelmiştir.

Virüs T hücrelerini enfekte ettiğinde, viryonun başarılı bir şekilde yayılması ve olgunlaşması için üç anahtar viral enzim (integraz, ters transkriptaz ve proteaz) gereklidir. HIV aspartik proteaz (HIV PR) dahil olmak üzere retroviral proteazlar, 99 ila 125 amino asit boyutunda, simetrik homodimerlerdir. HIV olgunlaşabilmek için PR'ye ihtiyaç duyar. Ayrıca HIV PR, dimerizasyon üzerine enzimatik olarak aktif olan bir aspartik proteazdır (McQuade ve ark. 1990). HIV PR daha fazla enfeksiyon yapabilmek için de gereklidir. HIV PR'nin etkisizleştirilmesi, olgunlaşmamış, bulaşıcı olmayan viral partiküllerin oluşmasına sebep olur (Wlodawer ve Erickson 1993).

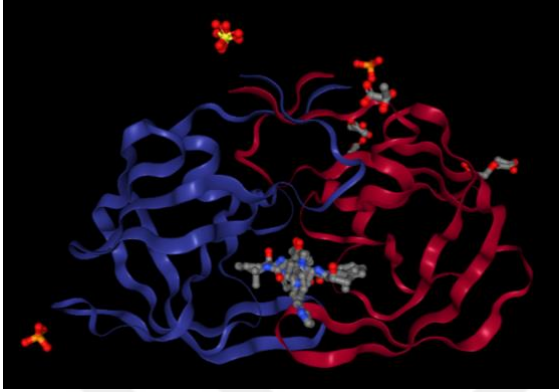
Olgun proteaz dimerini bir arada tutan etkileşimlerin kabaca yarısı, dimer arayüzü olarak bilinen bir bölgede meydana gelir. Bu arayüzü, iki monomeri (1 ila 4 arası rezidüel ve 96 ila 99 arası rezidüel) birbirine bağlayan iki N- ve iki C-terminal rezidüelinden oluşan dört iplikli beta tabakaları oluşturur (Şekil 7.1) (Pettit ve ark. 2003).



Şekil 7.1. 2o4k PDB kodlu HIV proteaz proteininde 4 iplikli beta-tabakalar (Sehnal ve ark. 2021)

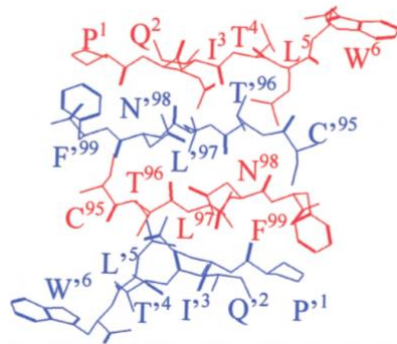
HIV PR dimerizasyonunun inhibitörleri, HIV PR'nin arayüzey bölgelerine karşılık gelen bu peptitlere dayanmaktadır (Şekil 7.2). Örneğin, HIV PR'nin 96-99 rezidüelinden oluşan bir C-terminal pentapeptidine karşılık gelen bir inhibitör peptidin, HIV PR'nin zayıf bir inhibitörü

olmasına rağmen, N-terminalinin açılasyonunun ilave olarak dört kat aktivite kaybına neden olduğu bulunmuştur (Şekil 7.3) (Bowman ve Chmielewski 2002). Benzer şekilde olgun proteazın N terminalindeki ilk 4 rezidüsünün serbest proteazın yapı bütünlüğünün sürdürülebilmesi için önemli olduğu bulunmuştur (Louis ve ark. 1999).



Şekil 7.2. 2avs PDB kodlu HIV Proteaz ve ligandları

Retroviral proteaz ailesine ait 1zp8 PDB monomerinin, homo-dimer oluşturduğu 4 kompleksin (her biri ayrı ligandlar içerir) PDB kodları 2o4k, 2o4p, 2o4s ve 2p3b'dir. 2o4s ve 2p3b homo-dimer kompleksleri 351 tetrahedron oluştururken, 2o4k 356 tetrahedron, 2o4p 360 tetrahedron oluşturmaktadır. Bu 4 kompleks incelendiğinde yüzeyi oluşturan tetrahedronlarının sayısının hepsinde 63 olduğu bulunmuştur. Bu yüzey tetrahedronlarından arayüze geçen tetrahedronların sayılarında farklılıklar görülmektedir. 2p3b'de 56 tetrahedron, 2o4k'da 58 tetrahedron, 2o4p'de 60 tetrahedrondan yüzeyden protein arayüzüne geçmiştir. Keskin tetrahedronlar açısından incelendiğinde tüm 4 kompleksin yüzeyinde 9 tane keskin tetrahedron bulunmuş ve 2o4k, 2o4p ve 2o4s'de bunların tamamı protein arayüzüne geçerken 2p3b'de sadece 1 tanesi geçmemiş, 8 tanesi protein yüzeyinden arayüzüne geçmiştir (Çizelge 7.1)



Şekil 7.3. HIV proteazın arayüzündeki 4 iplikli beta-tabaka dimerleşmesi (Bowman ve Chmielewski 2002).

Çizelge 7.1. Retroviral Proteaz protein ailesine ait bazı protein komplekslerinin tetrahedron, rezidü sayısı, hacim ve küresellik bilgileri

Tekil PDB	Kompleks PDB	Tetrahedron Sayısı	Tekil Protein Rezidü Sayısı	Hacim (Å ³)	Küresellik	Yüze Tetrahedron Sayısı	Yüze Rezidü Sayısı	Hacim (Å ³)	Küresellik	Arayüze Geçen Tetrahedron Sayısı	Arayüz Rezidü Sayısı	Hacim (Å ³)	Küresellik	Yüze Bulunan Keskin Tetrahedronlar	Yüze Bulunan Keskin Tetrahedronların Rezidü Sayısı	Hacim (Å ³)	Küresellik	Arayüze Geçen Keskin Tetrahedronlar	Arayüze Bulunan Keskin Tetrahedronların Rezidü Sayısı	Hacim (Å ³)	Küresellik
1zp8	2o4k	356	99	11,38	0,45	63	64	11,38	0,32	60	64	11,33	0,32	9	25	3,63	0,12	9	25	3,63	0,12
1zp8	2o4p	360	99	11,38	0,45	63	64	11,38	0,32	60	64	11,33	0,32	9	25	3,63	0,12	9	25	3,63	0,12
1zp8	2o4s	351	99	11,38	0,45	63	64	11,38	0,32	59	64	11,09	0,31	9	25	3,63	0,12	9	25	3,63	0,12
1zp8	2p3b	351	99	11,38	0,45	63	64	11,38	0,32	56	62	11,01	0,32	9	25	3,63	0,12	8	22	3,84	0,18

Keskin tetrahedronların arayüze geçen rezidülerinin (1, 2, 4, 6, 35, 37, 39, 40, 41, 44, 46, 48, 50, 51, 52, 53, 55, 61, 79, 88, 91, 92, 94, 98, 99), 22 tanesi (1, 2, 4, 6, 35, 37, 44, 46, 48, 50, 51, 52, 53, 55, 61, 79, 88, 91, 92, 94, 98, 99) bütün komplekslerde ortak iken 39,40 ve 41 rezidüler 2p3b’de arayüze geçmeyen rezidülerdir. “Kilit noktaları” oluşturan rezidülerin isimleri de sırasıyla Pro, Gln, Thr, Trp, Glu, Ser, Pro, Gly, Arg, Pro, Met, Gly, Ile, Gly, Gly, Phe, Lys, Gln, Pro, Asn, Thr, Gln, Gly, Asn ve Phe’dir. 2p3b’de arayüze geçmeyen rezidülerin isimleri Gly, Arg ve Pro’dur.

Daha detaylı bir araştırma olarak 2o4p HIV proteaz protein kompleksinin arayüzde etkileşen rezidülerine bakılmıştır. Bu protein kompleksinin seçilmesinin nedenleri; bu kompleksteki ligandın ağır atomlar bakımından diğerlerinden daha hafif olması ve hem tekil hem de kompleks proteinde bulunan ligandların atom tiplerinin aynılıklarının yüksek olmasıdır. Tekil proteinin sahip olduğu rezidüler ile arayüzde etkileşen rezidülerin sıklığı ve isimleri Çizelge 7.2’de gösterilmiştir. Arayüzde etkileşen 64 adet rezidüleme ait sıra numaraları ve isimleri Çizelge 7.3’de gösterilmiştir.

Çizelge 7.2. Tekil protein ve arayüzdeki rezidülerin sayısı ve isimleri

Rezidü Tipi	Tekil Protein Rezidüleri	Arayüzdeki Rezidüler
ALA	3	2
ARG	4	1
ASN	3	3
ASP	4	3
CYS	2	2
GLN	5	3
GLU	4	1
GLY	13	7
HIS	1	1
ILE	13	7
LEU	12	7
LYS	7	5
MET	2	2
PHE	2	2
PRO	6	4
SER	1	0
THR	8	7
TRP	2	2
TYR	1	0
VAL	6	3
TOPLAM	99	62

Çizelge 7.3. 2o4p protein kompleksindeki arayüzündeki rezidüler

Rezidü Sırası	Rezidü Adı	Rezidü Sırası	Rezidü Adı
1	PRO	50	ILE
2	GLN	51	GLY
3	ILE	52	GLY
4	THR	53	PHE
5	LEU	54	ILE
6	TRP	55	LYS
7	LYS	56	VAL
8	ARG	61	GLN
9	PRO	62	ILE
10	LEU	67	CYS
12	THR	68	GLY
19	LEU	69	HIS
20	LYS	70	LYS
21	GLU	74	THR
22	ALA	75	VAL
24	LEU	80	THR
25	ASP	81	PRO
26	THR	82	VAL
27	GLY	83	ASN
28	ALA	85	ILE
29	ASP	88	ASN
30	ASP	89	LEU
33	LEU	91	THR
36	MET	92	GLN
42	TRP	93	ILE
44	PRO	94	GLY
45	LYS	95	CYS
46	MET	96	THR
47	ILE	97	LEU
48	GLY	98	ASN
49	GLY	99	PHE

Tekil proteinde bulunan Asn, Cys, His, Met, Phe ve Trp'nin tamamı etkileşen rezidüler arasında bulunmuşken, Ser ve Tyr hiç görülmemiştir. Thr %88, Asp %75, Lys %71, Ala ve Pro %67, Gln %60 oranında bulunmuştur. Arg ve Glu %25, Val %50, Gly ve Ile %54, Leu %58 oranında az sıklıkla bulunan rezidüler arasında yer almışlardır.

HIV PR ile ilgili yapılan çalışmalarda ilk 4 rezidünün önemli olduğu vurgulanmış, bunlardan üçü (1,2 ve 4. rezidüler) “kilit noktalar” olarak bu çalışmada bulunmuştur. Yine proteinin C terminalindeki pentapeptitin ikisi (98 ve 99) kilit nokta olarak bulunmuştur.

8. TARTIŞMA VE SONUÇ

Yukarıdaki bulgular yüzey rezidülerinin bilinmesinin arayüz rezidülerinin tahmini için önemli bir kriter olduğunu göstermektedir çünkü yüzey tetrahedronlarının %94'ü homo-dimerlerde arayüze dönüşmektedir. Yine benzer şekilde yüzeyde bulunan keskin tetrahedronların da %96'sı arayüzde görülmektedir. Benzer şekilde yapılan bir çalışmada homo-dimer, trimer, tetramer ve hegzamerlerin protein arayüzündeki rezidü kompozisyonunun proteinin iç kısımlarından ziyade yüzeyinde bulunan rezidülerle daha ilişkili olduğunu ortaya koymuştur (Kabir 2004). Homolog olmayan 28 homo-dimer protein ile yapılan başka bir çalışmada da protein yüzeyinin %54-%87 oranında arayüze geçmesi ile protein yüzeyinin bilinmesinin gözlemlenen arayüzün kaba bir şekilde örneklenmesi anlamına geldiğini vurgulamışlardır (Jones ve Thornton 1997).

Yüzey tetrahedronlarının ortalama hacminin oranının tüm proteinin tetrahedronlarının ortalama hacmine oranı 0,71'dir ve bu proteinin iç kısımlarının daha sıkı bir şekilde paklendiğinin bir göstergesidir. Yüzey rezidüleri arayüze dönüşürken hacimlerinin değişimi %1 oranında artmış, bu da hacimlerini bir nevi koruduklarını göstermektedir. Yüzeydeki rezidülerin arayüz tahmini için önemli bir kriter olduğunu hacim değerleri de desteklemektedir. Yüzeydeki keskin tetrahedronlar arayüze geçerken hacimlerinde %9'luk bir artış görülmüştür. Bu da bu keskin tetrahedronların arayüzde önemli bir etkileşim yapmalarına olanak sağlayabilir. Protein yüzeyindeki atomların hacimlerinin araştırıldığı çalışmada; Voronoi hacim hesaplamalarının, protein yüzeyinde birçok atomun proteinin kor bölgesindekilere kıyasla ~%6 daha büyük olduğu, protein yüzeyinin proteinin koru kadar sıkı paketlenmediğini, protein arayüzündeki protein atomlarının Voronoi hacimlerinin daha büyük olduğu bulunmuştur (Gerstein ve ark. 1995). Protein yüzeyindeki atomlardan suya maruz kalanların daha büyük Voronoi hacimlerine sahip olduğu keşfedilmiştir (Gerstein ve Chothia 1996).

Arayüzdeki tetrahedronların ve keskin tetrahedronların küçük küresellik değerlerine sahip olduğu bulunmuştur. Arayüzdeki tetrahedronlarının ve keskin tetrahedronların ortalama küresellik değerlerinin daha az olmaları bu tetrahedronların irregüler (büyük olasılıkla daha düzlemsel) olduğunu ifade etmektedir. Tetrahedronun daha düzlemsel olması arayüzdeki rezidülerin daha fazla etkileşim yüzeyi oluşturmalarına fırsat sağlamış olabilir. Bu sonuç protein-protein etkileşimlerinin arayüzlerinin düzlemsel olma eğilimiyle uyumludur (Lu ve ark. 2020). Başka bir çalışmada, geometrileri yüksek derecede dejenere olan arayüzlerin protein

paketlenmesi sırasında düzlemsel olduğu Gao ve Skolnick tarafından ifade edilmiştir (Gao ve Skolnick 2010).

Protein yüzeyinde iki üçgene sahip olan tetrahedronların oluşturduğu keskin tetrahedronların arayüze geçme oranlarının tek üçgene sahip yüzeyde bulunan tetrahedronlara göre daha fazla olması bu tetrahedronların ve bu tetrahedronları oluşturan rezidüleri daha önemli olabileceğine işaret etmiştir. Protein arayüzündeki bu “kilit noktaların” bilinmesi, protein etkileşimlerinin anlaşılması için önemli bir ipucu vermektedir. Yapılan hesaplamalar ile anlamlı bir biçimde önemli olduğu düşünülen protein çiftlerinde bu “kilit noktaların” bulunma oranı %83,3 olarak bulunmuştur.

Her bir protein kompleksinin arayüzündeki “kilit noktalar” incelendiğinde en sık rezidülerin genellikle hidrofobik (glisin, alanin ve prolin), yüklü (lizin, arjinin, aspartat ve glutamat) ve polar (serin ve asparajin) rezidüler oldukları bulunmuştur. 142 zorunlu homolog kompleksinin arayüz rezidülerinin kompozisyonunun %45’i hidrofobik, %25’i yüklü ve %31’i polar olarak bulunması, bu çalışmada elde edilen verilerle benzerlik göstermektedir (Kabir 2004). Glisinin keskin tetrahedronları oluşturan rezidüler içerisinde en yaygın olması nedenleri; glisinin yan zinciri olmaması dolayısıyla diğer rezidülere kıyasla kolay paketlenmesi ve korunmuş glisinin aromatikler ve küçük hidrofobikler gibi diğer rezidülerle etkileşime girmesinin tercih edilmesinden dolayı olabilir (Keskin ve ark. 2005). Glisin ayrıca loop’larda çokça bulunmaktadır.

Serbest proteazın yapı bütünlüğünün sürdürülebilmesi için önemli olduğu ortaya konmuş olan önemli sekiz rezidüsünün (1-4, 96-99) beş tanesi (1, 2, 4, 98 ve 99) yapılan hesaplamalar sonucu kilit rezidüler arasında bulunmuştur. Kilit rezidü sayısı olan 22’nin, toplam rezidü sayısı olan 99’dan önemli ölçüde küçük olduğu göz önüne alındığında, tezde sunulan geometrik yaklaşımın deneysel çalışmalara (mutasyonla önemli rezidülerin ortaya çıkarılmasında, örneğin alan hedefli mutajenez) yardımcı olması beklenmektedir.

Hücre içerisindeki on binlerce proteinin ikili kombinasyonlarının sayısının ne kadar büyük olduğu düşünüldüğünde, protein-protein komplekslerinin tahmin edilmesinin ne kadar zorlu olduğu ortaya çıkar. Bu nedenle, bu çalışmada hesaplamalı yöntemler ile bulunan “kilit noktalar” protein etkileşimlerinin daha iyi anlaşılmasında yeni bir yaklaşım sunmaktadır. Bu çalışmada elde edilen “kilit noktalar” toplam protein rezidülerinin ortalama %21’ini oluşturmaktadır. “Kilit noktaların” önemli rezidüleri de içerdiği HIV Proteaz’da gösterilmiştir.

Dolayısıyla HIV Proteaz gibi hastalık yapıcı proteinlere karşı inhibitörler geliştirilmek istendiğinde başlangıç noktası olarak “kilit noktalar” seçildiğinde zaman, maliyet gibi kriterler açısından büyük avantajlar sağlayabilir.

Sonuç olarak yüzeyi oluşturan tetrahedronların bulunması, bunlara ait hacim hesapları ve bunların küreselliklerinin bilinmesi protein arayüzünü tahmin edebilmek için önemli kriterlerdir. Delaunay teselasyonu sonucunda elde edilen tetrahedronlar ile bulunan keskin tetrahedronların varlığı ve bunlara ait hacim ve küresellik değerlerinin bilinmesi protein arayüzlerinin tahmini için önemli görünmektedir.



KAYNAKLAR

Altundag, Ö. (2012). Bazı polimer yapıların küçük açı x-ışını saçılma yöntemi (SAXS) ile incelenmesi/Characterization by small angle x-ray scattering (SAXS) method of some polymer structures.

Andreeva, A., Kulesha, E., Gough, J., ve Murzin, A. G. (2019). The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures. *Nucleic Acids Research*, 48(D1), D376-D382. doi:10.1093/nar/gkz1064

Anikeenko, A. V., Gavrilova, M. L., ve Medvedev, N. N. (2008). Shapes of Delaunay Simplexes and Structural Analysis of Hard Sphere Packings. In M. L. Gavrilova (Ed.), *Generalized Voronoi Diagram: A Geometry-Based Approach to Computational Intelligence* (pp. 13-45). Berlin, Heidelberg: Springer Berlin Heidelberg.

Ansari, S., ve Helms, V. (2005). Statistical analysis of predominantly transient protein-protein interfaces. *Proteins: Structure, Function, and Bioinformatics*, 61(2), 344-355. doi:https://doi.org/10.1002/prot.20593

Atanasov, A. G., Waltenberger, B., Pferschy-Wenzig, E.-M., Linder, T., Wawrosch, C., Uhrin, P., ve Stuppner, H. (2015). Discovery and resupply of pharmacologically active plant-derived natural products: A review. *Biotechnology Advances*, 33(8), 1582-1614. doi:https://doi.org/10.1016/j.biotechadv.2015.08.001

Aydinkal, R. M., ve Bagci, E. Z. (2018). Residue packing in globular and intrinsically disordered proteins. *Proteins: Structure, Function, and Bioinformatics*, 86(4), 434-438. doi:https://doi.org/10.1002/prot.25459

Bagci, E. Z., Senguler-Ciftci, F., Ciftci, U., ve Demir, A. (2021). A novel measure to analyze protein structures: Aspect ratio in protein alpha shapes. *Proteins: Structure, Function, and Bioinformatics*, n/a(n/a). doi:https://doi.org/10.1002/prot.26148

Bahar, I., Jernigan, R. L., ve Dill, K. A. (2017). *Protein actions: Principles and modeling*: Garland Science.

Baker, T. J. (1989). Element quality in tetrahedral meshes. Paper presented at the Proc. 7th Int. Conf. on Finite Element Methods in Flow Problems.

Baspinar, A., Cukuroglu, E., Nussinov, R., Keskin, O., ve Gursoy, A. (2014). PRISM: a web server and repository for prediction of protein–protein interactions and modeling their 3D complexes. *Nucleic Acids Research*, 42(W1), W285-W289. doi:10.1093/nar/gku397

Bogan, A. A., ve Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces¹¹Edited by J. Wells. *Journal of Molecular Biology*, 280(1), 1-9. doi:https://doi.org/10.1006/jmbi.1998.1843

Bowman, M. J., & Chmielewski, J. (2002). Novel strategies for targeting the dimerization interface of HIV protease with cross-linked interfacial peptides. *Peptide Science*, 66(2), 126-133. doi:https://doi.org/10.1002/bip.10232

Bradley, M. P., ve Crippen, G. M. (1993). Voronoi modeling: the binding of triazines and pyrimidines to L. casei dihydrofolate reductase. *Journal of Medicinal Chemistry*, 36(21), 3171-3177. doi:10.1021/jm00073a018

Branden, C. I., ve Tooze, J. (2012). *Introduction to protein structure*: Garland Science.

Budowski-Tal, I., Kolodny, R., ve Mandel-Gutfreund, Y. (2018). A Novel Geometry-Based Approach to Infer Protein Interface Similarity. *Scientific Reports*, 8(1), 8192. doi:10.1038/s41598-018-26497-z

Burkowski, F. J. (2008). *Structural bioinformatics: an algorithmic approach*: CRC Press.

Caendish, J. C., Field, D. A., ve Frey, W. H. (1985). An approach to automatic three-dimensional finite element mesh generation. *International Journal for Numerical Methods in Engineering*, 21(2), 329-347. doi:https://doi.org/10.1002/nme.1620210210

Crippen, G. M. (1987). Voronoi binding site models. *Journal of Computational Chemistry*, 8(7), 943-955. doi:https://doi.org/10.1002/jcc.540080703

Çukuroğlu, E. (2015). *Clustering Protein Interfaces and Their Analysis*. (Doctoral dissertation). Koç University,

Dannelongue, H. H., ve Tanguy, P. A. (1991). Three-dimensional adaptive finite element computations and applications to non-Newtonian fluids. *International Journal for Numerical Methods in Fluids*, 13(2), 145-165. doi:https://doi.org/10.1002/flid.1650130203

de Cougny, H., Shephard, M., ve Georges, M. (1990). Explicit node point smoothing within the octree mesh generator. SCOREC Report, 10-1990.

De, S., Krishnadev, O., Srinivasan, N., ve Rekha, N. (2005). Interaction preferences across protein-protein interfaces of obligatory and non-obligatory components are different. *BMC Structural Biology*, 5(1), 15. doi:10.1186/1472-6807-5-15

Dune Core Modules. (2021). Erişim tarihi 24 Temmuz 2021, Erişim adresi https://www.dune-project.org/doxygen/2.6.0/classDune_1_1GridFactoryInterface.html

Edelsbrunner, H., ve Mücke, E. P. (1994). Three-dimensional alpha shapes. *ACM Trans. Graph.*, 13(1), 43–72. doi:10.1145/174462.156635

Erten, S., Li, X., Bebek, G., Li, J., ve Koyutürk, M. (2009). Phylogenetic analysis of modularity in protein interaction networks. *BMC Bioinformatics*, 10(1), 333. doi:10.1186/1471-2105-10-333

Esmailbeiki, R., Krawczyk, K., Knapp, B., Nebel, J.-C., ve Deane, C. M. (2016). Progress and challenges in predicting protein interfaces. *Briefings in bioinformatics*, 17(1), 117-131. doi:10.1093/bib/bbv027

Galperin, M. Y., Wolf, Y. I., Makarova, K. S., Vera Alvarez, R., Landsman, D., ve Koonin, E. V. (2020). COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Research*, 49(D1), D274-D281. doi:10.1093/nar/gkaa1018

Gao, M., ve Skolnick, J. (2010). Structural space of protein–protein interfaces is degenerate, close to complete, and highly connected. *Proceedings of the National Academy of Sciences*, 107(52), 22517-22522. doi:10.1073/pnas.1012820107

Gerstein, M., Tsai, J., ve Levitt, M. (1995). The Volume of Atoms on the Protein Surface: Calculated from Simulation, using Voronoi Polyhedra. *Journal of Molecular Biology*, 249(5), 955-966. doi:https://doi.org/10.1006/jmbi.1995.0351

Gerstein, M., ve Chothia, C. (1996). Packing at the protein-water interface. *Proceedings of the National Academy of Sciences*, 93(19), 10167-10172. doi:10.1073/pnas.93.19.10167

Glaser, F., Steinberg, D. M., Vakser, I. A., ve Ben-Tal, N. (2001). Residue frequencies and pairing preferences at protein–protein interfaces. *Proteins: Structure, Function, and Bioinformatics*, 43(2), 89-102. doi:[https://doi.org/10.1002/1097-0134\(20010501\)43:2<89::aid-prot1021>3.0.co;2-h](https://doi.org/10.1002/1097-0134(20010501)43:2<89::aid-prot1021>3.0.co;2-h)

Grigorieff, N., ve Harrison, S. C. (2011). Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy. *Curr Opin Struct Biol*, 21(2), 265-273. doi:10.1016/j.sbi.2011.01.008

Grigoriev, A. (2001). A relationship between gene expression and protein interactions on the proteome scale: analysis of the bacteriophage T7 and the yeast *Saccharomyces cerevisiae*. *Nucleic Acids Research*, 29(17), 3513-3519. doi:10.1093/nar/29.17.3513

Güney, E. (2007). Large scale characterization of protein interactions: Identification of hot spots and spatial motifs in protein-protein interfaces, Yüksek Lisans, Emre Güney, Koç Üniversitesi / Fen Bilimleri Enstitüsü / Elektrik ve Bilgisayar Mühendisliği Ana Bilim Dalı, 145 s., Bilgisayar Mühendisliği Bilimleri

Jones, S., ve Thornton, J. M. (1997). Prediction of protein-protein interaction sites using patch analysis11Edited by G. von Heijne. *Journal of Molecular Biology*, 272(1), 133-143. doi:<https://doi.org/10.1006/jmbi.1997.1233>

Kabir, T. (2004). The characterisation and prediction of protein-protein interfaces, (Doctoral Dissertation), ProQuest Dissertations Publishing, University of London, University College London (United Kingdom).

Keskin, O., Ma, B., & Nussinov, R. (2005). Hot Regions in Protein–Protein Interactions: The Organization and Contribution of Structurally Conserved Hot Spot Residues. *Journal of Molecular Biology*, 345(5), 1281-1294. doi:<https://doi.org/10.1016/j.jmb.2004.10.077>

Kim, M.-S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., ve Pandey, A. (2014). A draft map of the human proteome. *Nature*, 509(7502), 575-581. doi:10.1038/nature13302

Kobe, B., Guncar, G., Buchholz, R., Huber, T., Maco, B., Cowieson, N., Forwood, ve Jade K. (2008). Crystallography and protein–protein interactions: biological interfaces and crystal contacts. *Biochemical Society Transactions*, 36(6), 1438-1441. doi:10.1042/bst0361438

Kundrotas, P. J., Vakser, I. A., ve Janin, J. (2013). Structural templates for modeling homodimers. *Protein Science*, 22(11), 1655-1663. doi:<https://doi.org/10.1002/pro.2361>

Lander, G. C., Estrin, E., Matyskiela, M. E., Bashore, C., Nogales, E., ve Martin, A. (2012). Complete subunit architecture of the proteasome regulatory particle. *Nature*, 482(7384), 186-191. doi:[10.1038/nature10774](https://doi.org/10.1038/nature10774)

Lesk, A. M. (2001). *Introduction to protein architecture: the structural biology of proteins*: Oxford University Press Oxford.

Letunic, I., Khedkar, S., ve Bork, P. (2020). SMART: recent updates, new developments and status in 2020. *Nucleic Acids Research*, 49(D1), D458-D460. doi:[10.1093/nar/gkaa937](https://doi.org/10.1093/nar/gkaa937)

Levy, E. D., De, S., ve Teichmann, S. A. (2012). Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *Proceedings of the National Academy of Sciences*, 109(50), 20461-20466. doi:[10.1073/pnas.1209312109](https://doi.org/10.1073/pnas.1209312109)

Liang, J., Edelsbrunner, H., Fu, P., Sudhakar, P. V., ve Subramaniam, S. (1998a). Analytical shape computation of macromolecules: I. Molecular area and volume through alpha shape. *Proteins*, 33(1), 1-17.

Liang, J., Edelsbrunner, H., Fu, P., Sudhakar, P. V., ve Subramaniam, S. (1998b). Analytical shape computation of macromolecules: II. Inaccessible cavities in proteins. *Proteins*, 33(1), 18-29.

Lipinski, C., ve Hopkins, A. (2004). Navigating chemical space for biology and medicine. *Nature*, 432(7019), 855-861. doi:[10.1038/nature03193](https://doi.org/10.1038/nature03193)

Lobanov, M. Y., Shoemaker, B. A., Garbuzynskiy, S. O., Fong, J. H., Panchenko, A. R., ve Galzitskaya, O. V. (2010). ComSin: database of protein structures in bound (complex) and unbound (single) states in relation to their intrinsic disorder. *Nucleic Acids Research*, 38(Database issue), D283-D287. doi:[10.1093/nar/gkp963](https://doi.org/10.1093/nar/gkp963)

Louis, J. M., Clore, G. M., ve Gronenborn, A. M. (1999). Autoprocessing of HIV-1 protease is tightly coupled to protein folding. *Nature Structural Biology*, 6(9), 868-875. doi:[10.1038/12327](https://doi.org/10.1038/12327)

Lu, H., Zhou, Q., He, J., Jiang, Z., Peng, C., Tong, R., ve Shi, J. (2020). Recent advances in the development of protein–protein interactions modulators: mechanisms and clinical trials. *Signal Transduction and Targeted Therapy*, 5(1), 213. doi:10.1038/s41392-020-00315-3

Ma, B., Elkayam, T., Wolfson, H., ve Nussinov, R. (2003). Protein–protein interactions: Structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proceedings of the National Academy of Sciences*, 100(10), 5772-5777. doi:10.1073/pnas.1030237100

Madaoui, H., ve Guerois, R. (2008). Coevolution at protein complex interfaces can be detected by the complementarity trace with important impact for predictive docking. *Proceedings of the National Academy of Sciences*, 105(22), 7708-7713. doi:10.1073/pnas.0707032105

McQuade, T., Tomasselli, A., Liu, L., Karacostas, V., Moss, B., Sawyer, T., ve Tarpley, W. (1990). A synthetic HIV-1 protease inhibitor with antiviral activity arrests HIV-like particle maturation. *Science*, 247(4941), 454-456. doi:10.1126/science.2405486

Medvedev, N. N., ve Naberukhin, Y. I. (1987). Shape of the delaunay simplices in dense random packings of hard and soft spheres. *Journal of Non-Crystalline Solids*, 94(3), 402-406. doi:https://doi.org/10.1016/S0022-3093(87)80074-1

Medvedev, N. N., Voloshin, V. P., ve Naberukhin, Y. I. (1988). Structure of simple liquids as a percolation problem on the Voronoi network. *Journal of Physics A: Mathematical and General*, 21(4), L247-L252. doi:10.1088/0305-4470/21/4/011

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, Gustavo A., Sonnhammer, E. L. L., ve Bateman, A. (2020). Pfam: The protein families database in 2021. *Nucleic Acids Research*, 49(D1), D412-D419. doi:10.1093/nar/gkaa913

Mitra, P., ve Pal, D. (2010). New measures for estimating surface complementarity and packing at protein–protein interfaces. *FEBS Letters*, 584(6), 1163-1168. doi:https://doi.org/10.1016/j.febslet.2010.02.021

Moreira, I. S., Fernandes, P. A., ve Ramos, M. J. (2007). Hot spots—A review of the protein–protein interface determinant amino-acid residues. *Proteins: Structure, Function, and Bioinformatics*, 68(4), 803-812. doi:https://doi.org/10.1002/prot.21396

Murakami, Y., ve Mizuguchi, K. (2010). Applying the Naïve Bayes classifier with kernel density estimation to the prediction of protein–protein interaction sites. *Bioinformatics*, 26(15), 1841-1848. doi:10.1093/bioinformatics/btq302

Nadassy, K., Tomás-Oliveira, I., Alberts, I., Janin, J., ve Wodak, S. J. (2001). Standard atomic volumes in double-stranded DNA and packing in protein–DNA interfaces. *Nucleic Acids Research*, 29(16), 3362-3376. doi:10.1093/nar/29.16.3362

Nadassy, K., Wodak, S. J., ve Janin, J. (1999). Structural Features of Protein–Nucleic Acid Recognition Sites. *Biochemistry*, 38(7), 1999-2017. doi:10.1021/bi982362d

Ofran, Y., ve Rost, B. (2007). ISIS: interaction sites identified from sequence. *Bioinformatics*, 23(2), e13-e16. doi:10.1093/bioinformatics/btl303

Ogmen, U. (2006). PRISM: A System for the Query, Visualization and Analysis of Protein Interfaces and Their Networks. (Doctoral dissertation). Koç University

Ozbek, P., Soner, S., & Haliloglu, T. (2013). Hot Spots in a Network of Functional Sites. *PLOS ONE*, 8(9), e74320. doi:10.1371/journal.pone.0074320

Parthasarathy, V. N., Graichen, C. M., ve Hathaway, A. F. (1994). A comparison of tetrahedron quality measures. *Finite Elements in Analysis and Design*, 15(3), 255-261. doi:https://doi.org/10.1016/0168-874X(94)90033-7

Pazos, F., Juan, D., Izarzugaza, J. M. G., Leon, E., ve Valencia, A. (2008). Prediction of Protein Interaction Based on Similarity of Phylogenetic Trees. In J. D. Thompson, M. Ueffing, ve C. Schaeffer-Reiss (Eds.), *Functional Proteomics: Methods and Protocols* (pp. 523-535). Totowa, NJ: Humana Press.

Pettit, S. C., Gulnik, S., Everitt, L., ve Kaplan, A. H. (2003). The Dimer Interfaces of Protease and Extra-Protease Domains Influence the Activation of Protease and the Specificity of GagPol Cleavage. *Journal of Virology*, 77(1), 366-374. doi:doi:10.1128/JVI.77.1.366-374.2003

Pontius, J., Richelle, J., ve Wodak, S. J. (1996). Deviations from Standard Atomic Volumes as a Quality Measure for Protein Crystal Structures. *Journal of Molecular Biology*, 264(1), 121-136. doi:https://doi.org/10.1006/jmbi.1996.0628

Poupon, A. (2004). Voronoi and Voronoi-related tessellations in studies of protein structure and interaction. *Current Opinion in Structural Biology*, 14(2), 233-241. doi:<https://doi.org/10.1016/j.sbi.2004.03.010>

R Core Team. (2020). R: a language and environment for statistical computing [Internet]. R Foundation for Statistical Computing.

Rambo, R. P., ve Tainer, J. A. (2013). Accurate assessment of mass, models and resolution by small-angle scattering. *Nature*, 496(7446), 477-481. doi:10.1038/nature12070

Relman, D. A. (1998). Detection and identification of previously unrecognized microbial pathogens. *Emerging infectious diseases*, 4(3), 382-389. doi:10.3201/eid0403.980310

Reš, I., Mihalek, I., ve Lichtarge, O. (2005). An evolution based classifier for prediction of protein interfaces without using protein structures. *Bioinformatics*, 21(10), 2496-2501. doi:10.1093/bioinformatics/bti340

Jones, S., ve Thornton, J. M. (1997). Analysis of protein-protein interaction sites using surface patches. Edited by G. Von Heijne. *Journal of Molecular Biology*, 272(1), 121-132. doi:<https://doi.org/10.1006/jmbi.1997.1234>

Schreiber, G., ve Keating, A. E. (2011). Protein binding specificity versus promiscuity. *Current Opinion in Structural Biology*, 21(1), 50-61. doi:<https://doi.org/10.1016/j.sbi.2010.10.002>

Schrodinger, L. L. C. (2010). The AxPyMOL Molecular Graphics Plugin for Microsoft PowerPoint.

Sehna, D., Bittrich, S., Deshpande, M., Svobodová, R., Berka, K., Bazgier, V., ve Rose, A. S. (2021). Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Research*, 49(W1), W431-W437. doi:10.1093/nar/gkab314

Soner, S., (2009). Study of protein-protein interfaces using Gaussian network model, Yüksek Lisans, 78 s., Boğaziçi Üniversitesi / Fen Bilimleri Enstitüsü / Kimya Mühendisliği Bölümü / Kimya Mühendisliği Ana Bilim Dalı

Shoemaker, B. A., Panchenko, A. R., ve Bryant, S. H. (2006). Finding biologically relevant protein domain interactions: conserved binding mode analysis. *Protein Sci*, 15(2), 352-361. doi:10.1110/ps.051760806

Singh, R. K., Tropsha, A., ve Vaisman, I. I. (1996). Delaunay tessellation of proteins: four body nearest-neighbor propensities of amino acid residues. *Journal of Computational Biology*, 3(2), 213-221.

Smith, H. (2001). Discovery of the anthrax toxin: the beginning of studies of virulence determinants regulated in vivo. *International Journal of Medical Microbiology*, 291(6), 411-417. doi:<https://doi.org/10.1078/1438-4221-00147>

Srivastava, A., Nagai, T., Srivastava, A., Miyashita, O., & Tama, F. (2018). Role of Computational Methods in Going beyond X-ray Crystallography to Explore Protein Structure and Dynamics. *International Journal of Molecular Sciences*, 19(11), 3401. Retrieved from <https://www.mdpi.com/1422-0067/19/11/3401>

Stank, A., Kokh, D. B., Fuller, J. C., & Wade, R. C. (2016). Protein Binding Pocket Dynamics. *Accounts of Chemical Research*, 49(5), 809-815. doi:10.1021/acs.accounts.5b00516

STRING. (2021). Erişim tarihi 26 Temmuz 2021, Erişim adresi https://string-db.org/cgi/about?footer_active_subpage=statistics

Stumpf, M. P. H., Thorne, T., de Silva, E., Stewart, R., An, H. J., Lappe, M., ve Wiuf, C. (2008). Estimating the size of the human interactome. *Proceedings of the National Academy of Sciences*, 105(19), 6959-6964. doi:10.1073/pnas.0708078105

Tong, A. H. Y., Evangelista, M., Parsons, A. B., Xu, H., Bader, G. D., Pagé, N., ve Boone, C. (2001). Systematic Genetic Analysis with Ordered Arrays of Yeast Deletion Mutants. *Science*, 294(5550), 2364-2368. doi:10.1126/science.1065810

Tsai, J., ve Gerstein, M. (2002). Calculations of protein volumes: sensitivity analysis and parameter database. *Bioinformatics*, 18(7), 985-995. doi:10.1093/bioinformatics/18.7.985

Tsai, J., Taylor, R., Chothia, C., ve Gerstein, M. (1999). The packing density in proteins: standard radii and volumes11Edited by J. M. Thornton. *Journal of Molecular Biology*, 290(1), 253-266. doi:<https://doi.org/10.1006/jmbi.1999.2829>

Tsai, J., Voss, N., ve Gerstein, M. (2001). Determining the minimum number of types necessary to represent the sizes of protein atoms. *Bioinformatics*, 17(10), 949-956. doi:10.1093/bioinformatics/17.10.949

Tunçbağ, N. (2007). *Characterization and Analysis of Protein-protein Interfaces*. (Doctoral dissertation). Koç University,

Tunçbağ, N., Güney, E., Ulubaş, M. C., Keskin Özkaya, Z. Ö., ve Gürsoy, A. (2007). PRISM A Web Server for Prediction and Visualization of Protein Protein Interactions.

TÜBA (Türkiye Bilimler Akademisi). (2013). *Türkçe Bilim Terimleri Sözlüğü*. Erişim tarihi 24 Eylül 2020, Erişim adresi <http://www.tubaterim.gov.tr/>

Urakubo, Y., Ikura, T., ve Ito, N. (2008). Crystal structural analysis of protein–protein interactions drastically destabilized by a single mutation. *Protein Science*, 17(6), 1055-1065. doi:<https://doi.org/10.1110/ps.073322508>

Vaisman, I. I., Tropsha, A., ve Weifan, Z. (1998, 23-23 May 1998). Compositional preferences in quadruplets of nearest neighbor residues in protein structures: statistical geometry analysis. Paper presented at the Proceedings. IEEE International Joint Symposia on Intelligence and Systems (Cat. No.98EX174).

Wadell, H. (1935). Volume, Shape, and Roundness of Quartz Particles. *The Journal of Geology*, 43(3), 250-280. doi:10.1086/624298

Wlodawer, A., ve Erickson, J. W. (1993). STRUCTURE-BASED INHIBITORS OF HIV-1 PROTEASE. *Annual Review of Biochemistry*, 62(1), 543-585. doi:10.1146/annurev.bi.62.070193.002551

Zhang, J., ve Kurgan, L. (2017). Review and comparative assessment of sequence-based predictors of protein-binding residues. *Briefings in bioinformatics*, 19(5), 821-837. doi:10.1093/bib/bbx022

Zhang, Q. C., Deng, L., Fisher, M., Guan, J., Honig, B., ve Petrey, D. (2011). PredUs: a web server for predicting protein interfaces using structural neighbors. *Nucleic Acids Research*, 39(suppl_2), W283-W287. doi:10.1093/nar/gkr311

Zhang, Q. C., Petrey, D., Norel, R., ve Honig, B. H. (2010). Protein interface conservation across structure space. *Proceedings of the National Academy of Sciences*, 107(24), 10896-10901. doi:10.1073/pnas.1005894107

