

T.C.
HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

FENOTİP ÇEŞİTLİLİĞİNİ AÇIKLAMADA ALX4
VARYANTLARININ PROTEİN MODELLERİNDE
İNCELENMESİ

Bengü Yaren BEYAZ

Biyoinformatik Programı
YÜKSEK LİSANS TEZİ

ANKARA

2024

T.C.
HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

FENOTİP ÇEŞİTLİLİĞİNİ AÇIKLAMADA ALX4
VARYANTLARININ PROTEİN MODELLERİNDE
İNCELENMESİ

Bengü Yaren BEYAZ

Biyoinformatik Programı
YÜKSEK LİSANS TEZİ

TEZ DANIŞMANI

Prof. Dr. Ayşe Nurten AKARSU

İKİNCİ DANIŞMAN

Prof. Dr. Turgut BAŞTUĞ

ANKARA

2024

**FENOTİP ÇEŞİTLİLİĞİNİ AÇIKLAMADA ALX4 VARYANTLARININ PROTEİN
MODELLERİNDE İNCELENMESİ**

Öğrenci: Bengü Yaren BEYAZ

Danışman: Prof. Dr. Ayşe Nurten AKARSU

İkinci Danışman: Prof. Dr. Turgut BAŞTUĞ

Bu tez çalışması 25/07/2024 tarihinde jürimiz tarafından "Biyoinformatik Programı"
nda yüksek lisans tezi olarak kabul edilmiştir.

Jüri Başkanı: Doç. Dr. Yeşim Aydın SON
(Orta Doğu Teknik Üniversitesi)

Tez Danışmanı: Prof. Dr. Ayşe Nurten AKARSU
(Hacettepe Üniversitesi)

Üye: Doç. Dr. Ceren SUCULARLI
(Hacettepe Üniversitesi)

Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca yukarıdaki jüri tarafından uygun bulunmuştur.

23 Ağustos 2024

Prof. Dr. Müge YEMİŞÇİ ÖZKAN
Enstitü Müdürü

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan “**Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge**” kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir. ⁽¹⁾
- Enstitü / Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ... ay ertelenmiştir. ⁽²⁾
- Tezimle ilgili gizlilik kararı verilmiştir.

23/08/2024

Bengü Yaren BEYAZ

¹“*Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge*”

- (1) *Madde 6. 1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.*
- (2) *Madde 6. 2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internette paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç imkanı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.*
- (3) *Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir *. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir. Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir*

** Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.*

ETİK BEYAN

Bu alıřmadaki bütn bilgi ve belgeleri akademik kurallar erevesinde elde ettiđimi, grsel, iřitsel ve yazılı tm bilgi ve sonuları bilimsel ahlak kurallarına uygun olarak sunduđumu, kullandıđım verilerde herhangi bir tahrifat yapmadıđımı, yararlandıđım kaynaklara bilimsel normlara uygun olarak atıfta bulunduđumu, tezimin kaynak gsterilen durumlar dıřında zgn olduđunu, tez danıřmanı Prof. Dr. Ayře Nurten AKARSU ve eř danıřman Prof. Dr. Turgut BAŐTUĐ'un danıřmanlıđında tarafımdan retildiđini ve Hacettepe niversitesi Sađlık Bilimleri Enstits Tez Yazım Ynergesine gre yazıldıđını beyan ederim.

Beng Yaren BEYAZ

TEŞEKKÜR

Bu tezi hazırlamamda ve Ankara’da yaşam kurmamda bana çok yardımcı olan danışmanım Prof. Dr. Ayşe Nurten Akarsu’ya ve eş danışmanım Prof. Dr. Turgut Baştuğ’a, teze çok büyük katkısı olan, benimle birebir ilgilenen ve sorularıma sabırla cevap veren Dr. Öğr. Üyesi Arda Çetinkaya’ya bilgilerini paylaştığı ve değerli vaktini ayırdığı için minnetimi sunarım. Bu süreçte beni yalnız bırakmayan Doç. Dr. Ceren Sucularlı’ya, TÜSEB Aziz Sancar Araştırma Merkezindeki çalışma arkadaşlarıma çok teşekkür ederim. Tüm eğitim hayatım boyunca beni yalnız bırakmayan aileme, henüz lisede olmasına rağmen otomatik işlerimde bana çok yardımcı olan kardeşim Dildar Naz Beyaz’a ve odaklanmamda ve tezimde bana çok yardımcı olan Akif Koç’a ne kadar şanslı olduğumu ve şükranlarımı sunarım.

ÖZET

Beyaz, B.Y., Fenotip Çeşitliliğini Açıklamada ALX4 Varyantlarının Protein Modellerinde İncelenmesi, Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü Biyoinformatik Programı Yüksek Lisans Tezi, Ankara, 2024. *ALX* gen ailesi, embriyonik gelişimde önemli rol oynayan ve mutasyonlarında insanlarda kafatasında ve yüzde malformasyonlar gösteren homeodomain içerikli transkripsiyon faktörlerini kodlar. Ailenin üyelerinden *ALX1* ve *ALX3* mutasyonları proteinde işlev kaybı mutasyonlarına yol açarken, *ALX4*'te işlev kaybı ve kazanımı sonucu ortaya çıkan farklı fenotipler gözlenmektedir. *ALX4*'ün tamamının üç boyutlu yapısı ve bu davranışının nedeni henüz bilinmemektedir. Bu tez çalışmasında önce *ALX4* proteini modellenmeye çalışılmıştır. Güvenilir modellenebilen homeodomain bölgesindeki değişikliklerin amino asit etkileşimlerine bakılarak fenotipe etkilerinin işlev kaybı veya kazanımı olması hakkında tahmin yürütülebileceği saptanmıştır.

Anahtar Kelimeler: *ALX4*, protein modelleme, şablon tabanlı modelleme, şablon tabansız modelleme, işlev kaybı, işlev kazanımı

ABSTRACT

Beyaz, B.Y., Investigation of ALX4 Variants in Protein Models to Explain Phenotype Diversity, Hacettepe University Graduate School of Health Sciences Bioinformatics Program Master Thesis, Ankara, 2024. The *ALX* gene family encodes homeodomain-containing transcription factors that play an important role in embryonic development and whose mutations cause craniofacial malformations in human. While *ALX1* and *ALX3* mutations lead to loss-of-function mutations in the protein, *ALX4* shows different phenotypes resulting in loss and gain of function. The full three-dimensional structure of ALX4 and the reason for this behavior is not yet known. In this thesis, we first tried to model the ALX4 protein. It was determined that the effects of changes in the homeodomain region, which can be reliably modeled, can be predicted to be loss or gain of function by looking at amino acid interactions.

Key Words: ALX4, protein modeling, template-based modeling, template-free modeling, loss of function, gain of function

İÇİNDEKİLER

ONAY SAYFASI	iii
YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI	iv
ETİK BEYAN SAYFASI	v
TEŞEKKÜR	vi
ÖZET	vii
ABSTRACT	viii
İÇİNDEKİLER	ix
SİMGELER ve KISALTMALAR	xii
ŞEKİLLER	xiii
TABLolar	xiv
1. GİRİŞ	1
2. GENEL BİLGİLER	3
2.1. Protein ve Yapısı	3
2.1.1. Protein	3
2.1.2. Protein Yapısı	5
2.1.3. Protein Katlanması	12
2.1.4. Genetik Varyantlar ve Proteinlerde Fonksiyon Kaybettiren ve Kazandıran Değişiklikler	13
2.1.5. Proteinler Arası Yapı	16
2.2. ALX4 Geni ve Proteini	16
2.2.1. Homeobox Genleri ve Proteinleri	16
2.2.2. ALX Gen Ailesi	17
2.2.3. ALX4 Proteini	19
2.3. Protein Yapısının Eldesi	20
2.3.1. Protein Yapısının Deneysel Eldesi	21
2.3.2. Protein Yapısının Hesaplamalı (<i>in silico</i>) Yöntemlerle Eldesi – Protein Modelleme	21
2.3.3. CAPS Yarışması ve Günümüzde Kullanılan Modelleyiciler	25
2.3.4. Homolog Protein Araştırması	27

2.3.5. Protein Yapısının Görselleştirilmesi	28
2.3.6. Proteinlerde Yapı Kontrolü ve Yapıların Karşılaştırılması	28
2.3.7. Minimizasyon	28
3. GEREÇ VE YÖNTEM	30
3.1. Yabani Tip ALX4 Proteininin Üç Boyutlu Yapısının Araştırılması	30
3.2. Kullanılacak Modelleyicilerin Modellenmesi	30
3.3. Yabani Tip ALX4 Proteininin Üç Boyutlu Yapısının Tahmin Edilmesi	30
3.3.1. Dizinin Belirlenmesi	30
3.3.2. Homolog Protein Araştırması	30
3.3.3. Belirlenen Modelleyicilerde Modelleme	31
3.3.4. Eksik Atomların Tamamlanması	33
3.3.5. Modelleyicilerin En İyi Modelinin Belirlenmesi	34
3.3.6. NAMD ile Minimizasyon Yapılması	36
3.3.7. ALX4 için En İyi Modelleyicinin ve Modelin Belirlenmesi	38
3.4. 2M0C'nin Yabani Homeodomain Olarak Hazırlanması	39
3.5. Varyant Proteinlerin Seçilmesi	39
3.6. Varyant Proteinlerin Modellenmesi	40
3.6.1. Dizilerin Üretilmesi	40
3.6.2. Modelleme ve PSF Oluşturma	40
3.6.3. En İyi Modellerin Belirlenmesi	40
3.7. Yabani Homeodomain ile Varyant Proteinlerin Karşılaştırılması	41
4. BULGULAR	42
4.1. Yabani Tip ALX4 Üç Boyutlu Yapısı ve Dizisi	42
4.2. Yabani Tip ALX4 Üç Boyutlu Yapısının Modellenmesi	45
4.2.1. Homolog Protein	45
4.2.2. Yabani ALX4 Modellerinin Karşılaştırılması ve Minimizasyon	46
4.3. ALX4 Homeodomain Varyant Modelleri	55
4.3.1. Modelleme için Kullanılacak Varyant Proteinler ve Dizileri	56
4.3.2. Varyant Protein Modelleri	57
4.4. Yabani Homeodomain ile Varyantların Karşılaştırılması	60
5. TARTIŞMA	64

6. SONUÇ ve ÖNERİLER	69
7. KAYNAKLAR	70
8. EKLER	74
EK-1: Tez Çalışması ile İlgili Etik Kurul İzinleri	
EK-2: Tez Çalışması Orijinallik Raporu	
9. ÖZGEÇMİŞ	



SİMGELER ve KISALTMALAR

ALX	<i>Aristaless-like homeobox</i>
BLASTP	<i>protein-protein Basic Local Alignment Tool</i>
CASP	<i>Critical Assesment of Protein Structure Prediction</i>
CHARMM	<i>Chemistry at HARvard Macromolecular Mechanics</i>
D-I-TASSER	<i>Deep Learning-based Iterative Threading ASSEmbly Refinement</i>
GDT_TS	<i>The Global Distance Test_Total Score</i>
gnomAD	<i>The Genome Aggregation Databse</i>
i-TASSER	<i>Iterative Threading ASSEmbly Refinement</i>
IDR	<i>Intrinsically Disordered Region</i>
IUPAC	<i>International Union of Pure and Applied Chemistry</i>
MSA	<i>Multiple Sequence Alignment</i>
NAMD	<i>Nanoscale Molecular Dynamics</i>
NCBI	<i>National Center for Biotechnology Information</i>
NMR	<i>Nuclear Magnetic Resonance</i>
PAE	<i>Predicted Aligned Error</i>
PDB	<i>Protein Data Bank</i>
pLDDT	<i>predicted Local Distance Difference Test</i>
pLM-BLAST	<i>Protein Local Match-Basic Local Alignment Tool</i>
PSF	<i>Protein Structure File</i>
RMSD	<i>Root-Mean-Square Deviation</i>
UniProt	<i>The Universal Protein Resource</i>
VMD	<i>Visual Molecular Dynamics</i>

ŞEKİLLER

Şekil		Sayfa
2.1.	<i>ALX4</i> mutasyon dağılımı ve fenotip çeşitliliği.	19
4.1.	<i>ALX4</i> proteininin tahmini yapısı.	42
4.2.	<i>ALX4</i> proteini homeodomain bölgesinin NMR yapısı.	43
4.3.	<i>ALX4</i> proteininin tahmini yapısı (Q9H161) ile homeodomain bölgesinin NMR yapısının (2M0C) hizalanmış görüntüsü.	44
4.4.	<i>ALX4</i> için homolog arama sonuçları.	46
4.5.	Modelleyicilerin çalışılan algoritmalarının ortalama MolProbit ve RMSD sonuçları.	48
4.6.	Modelleyicilerde üretilen en iyi modellerin doğruluk grafikleri.	49
4.7.	Modelleme sonuçları.	52
4.8.	Minimizasyon sonuçları.	54
4.9.	Varyant modelleri ve PAE grafikleri.	58
4.10.	<i>ALX4</i> protein varyant modellerinin homeodomain bölgesi için yabancı <i>ALX4</i> ile üst üste karşılaştırılması.	60
4.11.	p.Ala243Val varyantının etkileşimleri.	62
4.12.	p.Lys211Glu varyantının etkileşimleri.	62
4.13.	p.Gln225Glu ve p.Arg272Pro varyantlarının etkileşimleri.	63
4.14.	p.Arg218Gln varyantının etkileşimleri.	63
5.1.	<i>ALX4</i> için UniProt veri tabanından elde edilen protein domainleri.	65
5.2.	<i>ALX4</i> homeodomaininde arjininin DNA'ya bağlanması.	67
5.3.	Arjinin ve glutaminin kimyasal yapıları.	68

TABLÖLAR

Tablo		Sayfa
2.1.	Protenlerde yer alan amino asitler ve kimyasal özellikleri.	9
2.2.	CASP13-15 yarışmalarında en yüksek değerleri alan modelleyiciler ve Z değerleri.	26
4.1.	En iyi modeller ve bu modellerin minimize yapılarının RMSD ve ayrıntılı MolProbity değerleri.	51
4.2.	Varyantların pLDDT ve detaylı MolProbity değerleri.	59
4.3.	Modellenen ALX4 varyantlarının özellikleri.	61

1. GİRİŞ

Proteinlerin biyolojik olarak işlevsel üç boyutlu yapılarını bilmek proteinin işlevini daha iyi anlayabilmek, proteinlerin diğer biyomoleküllerle ilişkilerini anlamak, hastalıklarla mücadelede genetik mutasyonların fonksiyon üzerindeki etkilerini araştırmak, ilaç tasarımlarına rehberlik etmek gibi farklı alanlarda yardımcıdır. Başlangıçta deneysel yöntemlerle proteinlerin üç boyutlu yapısı anlaşılmaya çalışılmıştır. Bu amaçla farklı yöntemler bulunmakla birlikte halen en yaygın kullanılan yöntemler X-ışını kristalografisi ve nükleer manyetik rezonanstır. Yapı eldesinde deneysel modeller en güvenilir veriyi sunmakla birlikte bu yöntemler çok pahalı olup zaman, özelleşmiş alt yapı ve insan gücü gerektiren işlemlerdir. İnsan genom projesi sonrası DNA ve protein dizi bilgilerinin hızla çoğalması, bu bilgilerin veri tabanlarına eklenerek erişime açılması, yapay zeka modellerinin giderek gelişmesi sonucunda dizi bilgilerinden hesaplamalı yöntemlerle yapı tahmin metodlarının gelişmesine doğru yöntemler evrilmiştir. Bu kapsamda çok sayıda yazılım üretilmiş, daha iyi tahmin araçlarının tespit edilmesine yönelik skorlama sistemleri geliştirilmiştir.

Genetik hastalıklar alanında da protein yapısal tahmin araçları DNA varyantlarının fonksiyona etkilerinin tahmin edilmesinde etkin olarak kullanılmaktadır. Tez danışmanı tarafından ilk kez 2009 yılında insanlarda yüz gelişim anomalileri ile ilişkilendirilmiş olan *ALX4* geni varyantlarının bu tez çalışmasında üç boyutlu protein yapısında nasıl değişikliklere yol açtığı araştırılması ve bunun fenotip çeşitliliğine olası katkısının hesaplamalı protein modelleme yöntemleri ile test edilmesi planlanmıştır.

Aristaless - Like Homeobox (*ALX*) gen ailesi omurgalılarda *ALX1*, *ALX3* ve *ALX4* olarak isimlendirilen üç üyeye sahip paired tipi homeodomain içeren bir transkripsiyon faktörüdür. Her üç gendeki mutasyonlar orta yüz gelişimini etkilemekte ve farklı şiddette frontozal displazi fenotiplerine yol açmaktadır. *ALX1* ve *ALX3* gen mutasyonları biallelik fonksiyon kayıpları ile seyirli otozomal resesif

yüz malformasyonlarına yol açarken *ALX4* geninde monoallelizm (heterozigot) ve biallelik patojenik varyantlar hastalığa sebep olmaktadır. Bu niteliği ile *ALX4* semidominant (biallelik formlarının daha ağır fenotip gösterdiği) kalıtım kalıbı özelliklerine uyar şekilde malformasyonlara neden olmaktadır. *ALX4* geni fonksiyon kaybettiren ve fonksiyon kazandıran mutasyonları ile ortaya çıkan fenotip çeşitliliği de dikkat çekicidir. *ALX4* geni fonksiyon kaybettiren heterozigot mutasyonları kafa kemiklerinden parietal kemik içinde kemikleşme kusuru ile giden “parietal foramina” hastalığına yol açarken, genin fonksiyon kazandıran heterozigot mutasyonları ise kafa kemikleri arasındaki bağ dokunun bebeklik döneminde erken kemikleşmesi sonucu ortaya çıkan ve farklı bir hastalık olan kraniyosinostoz fenotipi ile ilişkilendirilmiştir. *ALX4* geni biallelik (homozigot ve bileşik heterozigot) mutasyonları ise yüz orta hattı ve burunda malformasyon ve kafa kemiklerinde kemikleşme kusurları ile giden daha ağır bir hastalığa neden olmaktadır.

ALX4 geni 11. kromozomun kısa kolu 11 bandında yerleşik 4 ekzonlu bir gen olup 411 amino asit uzunluğunda ALX4 proteinini kodlamaktadır. ALX4 proteinin yapısı proteinin 214-273 amino asitleri arasında kalan ve üç alfa heliksin bir hidrofobik merkez çevresinde organize olduğu DNA bağlanma bölgesi (homeodomain) dışında bilinmemektedir. Oysa ki *ALX4* geninin kafatası ve yüz malformasyonları ile ilişkilendirilmesinin ardından bu genin farklı bölgelerinde çok sayıda varyant tanımlanmış, bunların bir kısmının fonksiyon kazandırıcı özellikleri deneysel olarak da gösterilmiştir. Bu nedenle tez kapsamında ilk olarak yabancı tip ALX4 proteini için başarılı bir model oluşturulması hedeflenmiştir. Sonrasında da fonksiyon kazandıran, fonksiyon kaybettiren ve nötral varyantları içeren grupların ortak ve ayırtıcı özelliklerinin belirlenmesi hedeflenmiştir.

2. GENEL BİLGİLER

2.1. Protein ve Yapısı

2.1.1. Protein

Proteinler amino asit adı verilen yapıtaşlarının bir kovalent bağ olan peptid bağı ile bağlanması sonucu oluşan makromoleküllerdir (1). Genelde uzunlukları 50 ile 2000 amino asidin birleşimi kadardır. Bundan daha kısa ve uzun proteinler de bulunmaktadır, ama daha nadirdir. Hücre birçok faaliyetini proteinler üzerinden yürütür. Başka hücrelerle etkileşimini ve çevre ile iletişimini de proteinler üzerinden gerçekleştirir. Dolayısıyla proteinler organizmanın yaşama süreci ve halinde etken rol oynar. Proteinler bu hayati fonksiyonları üç boyutlu yapılarına göre farklı fonksiyonlar kazanıp, bu fonksiyonların ilişığı görevleri yerine getirerek gerçekleştirirler (2).

Protein DNA'dan başlayıp RNA'ya devam eden bir bilgi akışı (santral dogma) ile hücre içerisinde elde edilir. Santral dogma, hücrede bilgi akışının en temel açıklamasıdır. RNA (ribonükleik asit) polimeraz DNA (deoksiribonükleik asit)'nın gen bölgesinin *promotor* kısmına bağlanarak kopması gereken bölgeye kadar ilerler. İleride anlatılacak tRNA biyomolekülünün kimyasal yapısından dolayı her üç nükleotit birlikteliğine *kodon* denir ve kodon bir amino asite karşılık gelir. RNA polimeraz proteini dur (*stop*) kodonuna varınca kopar. Promotor bölgeden hemen sonra başlayan başlama (*start*) kodonu ile ise gen boyunca ilerlemesi gerektiğini bilir ve ilerlemeye başlar. Bu sırada ortamdaki serbest nükleotitleri karşısına tamamlayıcı şekilde bağlayarak tamamlayıcı zincir yani pre-mRNA [haberci (*messenger*) ribonükleik asit öncesi] üretir. Diğer bir deyişle transkripsiyon gerçekleşir. mRNA'da genin birebir yapısı yani hem kodlanan (*exon*), hem de kodlanmayan (*intron*) bölgeler bulunur. Bunlarla birlikte baş (5') ve son (3') uçlarında kodlanmayan, mRNA [haberci (*messenger*) ribonükleik asit]'nın ortamda kırılmadan daha uzun süre bulunmasını sağlayan proteine dönüşmeyen bölgeleri [UTR (*untranslated*

region)] bulunmaktadır. UTR dışındaki kodlanmayan bölgeye bağlanan proteinler bu bölgeyi mrRNA'dan uzaklaştırır ve kodlanan bölgeler sırasıyla birbirine bağlanır. Bu olaya uç birleştirme (*splicing*) denir. Uç birleştirme ile olgun mRNA oluşur. Çekirdekte oluşan bu mRNA'lar çekirdek dışına çıkıp dışarıda bulunan ribozoma bağlanır. Buna ek olarak üç nükleotit ve buna karşılık gelen 1 amino asit (kodon) bulunduran tRNA [taşıyıcı (*transfer*) ribonükleik asit]'lar da tamamlayıcı olacak şekilde ribozoma bağlanır ve her bağlandığında bir önceki bağlanan amino asitine bir amino asit daha ekler. Böylece protein lineer zinciri oluşarak translasyon işlemi gerçekleşir (2,3).

Bu aşamadan sonra protein, katlanma geçirerek biyolojik olarak kullanılabilir hale gelir. Böylece işlevsel üç boyutlu yapılarını katlanma ile veya katlanmadan sonra başka bir protein ya da molekül (*ligant*) ile bağ yaparak (*dock*) elde ederler. Her yaptıkları bağ sonucu üç boyutlu şekilleri (*konformasyon*) değişir. Kazanılan her farklı konformasyon ile farklı görev yapabilir (*fonksiyon*) hale gelirler. Proteinler, katlanmadan sonraki konformasyon ve fonksiyonlarını devamlı sürdürmek zorunda değillerdir. Çevresel koşullara göre, bağ yaparak yeni konformasyona kavuşabilir, bu bağı kırarak da eski konformasyonuna dönebilir veya yeni bir bağ yaparak başka konformasyona sahip olabilirler. Böylece yeni fonksiyon kazanabilir, kısmi zamanlı fonksiyon edinebilir ya da eski fonksiyonlarına geri dönebilirler (4). Fosforilasyon ve defosforilasyon bu duruma örnek olarak verilebilir.

Protein üç boyutlu yapısının bu denli kolayca değişebilmesinin sebebi, protein katlanmasının tamamen atomların birbirine uyguladığı kuvvetler ve bu kuvvetler sonucunda her atomun konumunun sistem yani protein en düşük enerjiye sahip olacak şekilde sabitlenmesidir. Diğer bir deyişle sistem dengededir, protein bir şekle sahiptir. Her yeni bağda bağlanan tarafların atomları birbirine kuvvet uygulayarak konumlarını değiştirir ve dengedeki sistemler bozulur, dolayısıyla proteinin ve bağlanan protein veya ligantın şekli değişir. Birlikte sistemi dengede tutacak şekilde yeni bir şekle sahip olurlar.

2.1.2. Protein Yapısı

Latince *-mer* son eki yapı taşı anlamına gelmekte olup, *mono-* baş eki tek, *poli-* baş eki birçok anlamına gelmektedir. Proteinler, amino asit monomerlerinin biraraya gelerek oluşturduğu polimer bir yapıdır. Amino asit moleküllerinin her biri, diğer bir amino asit ile bir *kovalent bağ* türü olan *peptit bağı* yapar ve uzun amino asit zinciri oluşur. Bu nedenle proteinlere aynı zamanda *polipeptit*, amino asit zincirine de *polipeptit zinciri* de denmektedir. Proteinin bu uzun polipeptit zincir yapısı (aynı zamanda tRNA ile ribozomda oluşan uzun zincir yapısı) birincil (*primer*) yapı olarak bilinir. Amino asitlerin polipeptit zinciri oluştururken ortaya çıkardığı sıra, o proteinin *amino asit dizisini* oluşturur.

Amino asitler 20 farklı çeşittir. C (karbon), H (hidrojen), O (oksijen) ve N (azot) atomlarının farklı şekillerde biraraya gelmesiyle oluşur. Her amino asit NH₂ (amin) molekülünü içeren amino grubu (amino), COOH (karboksilik asit) molekülünü içeren karboksil grubu (asit) ve ortada bu iki grubun bağlandığı karbon atomu bulundurur. Bu karbon atomu aynı zamanda hidrojen atomuna ve amino asite karakterini veren radikal gruba (yan zincir) bağlıdır. Yan zincir dışındaki yapı tüm amino asitlerde tamamen aynıdır. Bu değişmeyen kısma amino asidin *belkemiği* denir. Tek değişken kısım yan zincirdir.

Protein, amino asitlerin yan zinciri dışarıda çıkıntı olarak kalacak şekilde sadece belkemiği kısmının birbiri ile bağ yapmasıyla oluşur. Protein birincil yapısında amino asitler yeni gelen amino asidin belkemiğinde bulunan amino grubunun önceki amino asidin belkemiğinde bulunan karboksil grubuna peptit bağıyla bağlanması ile art arda dizilir. Bu peptit bağı amino grubu NH₂ molekülünden bir H atomu ile karboksil grup COOH molekülünden OH (hidroksil) molekülünün birleşip koparak H₂O (su) molekülü ortaya çıkarmasıyla oluşur. Amino asitlerin belkemiğinin birleşimi protein belkemiğini (*backbone*) oluşturur. Belkemiği dışında dışarıya doğru çıkıntı yapan yan zincirler de proteinin yan zincirleridir. Tüm

proteinlerin belkemiği birbirinin aynısıdır. Proteinleri birbirinden farklı yapan yan zincirlerin sıralanışı ve amino asit sayısıdır.

Protein birincil yapısının bir başı bir de sonu vardır. Başı, ilk amino asidin amino grubu, sonu ise son amino asidin karboksil grubudur. Bu başlangıç amino grubuna *N terminal (amino terminali)*, bitiş karboksil grubuna *C terminal (karboksil terminali)* denir. Protein N terminalden C terminale doğru okunur ve birincil yapı oluşurken her yeni gelen amino asit zincire bu yönde katılır (1, 2, 4).

Organik kimyada bir molekül zincirinde birden çok C atomu bulunduğunda hangi C atomuna ne bağlandığını kolayca ifade edebilmek için kimya terminolojisi üzerine çalışan kurum IUPAC (*International Union of Pure and Applied Chemistry – Uluslararası Temel (Teorik) ve Uygulamalı Kimya Birliği*) tarafından her C atomunun numaralandırılması önerilmiştir. Böylelikle, Yunanca numaralandırmada fonksiyonel gruba bağlanan ilk C atomuna Yunancada bir anlamına gelen *alfa* (α), ikinciye iki anlamına gelen *beta* (β), üçüncüye üç anlamına gelen *gamma* (γ) C atomu denmiştir. Organik kimyada fonksiyonel grup, molekülde diğer atomlardan farklı davranan ve belirli bağlar yapan özel atom grubunu ifade eder. Amino asit belkemiğinde amino ve karboksil fonksiyonel grupları bulunduğundan ortadaki C atomu C_α atomu olarak adlandırılır. Protein belkemiğinde ise N terminalden sonra gelen C atomu C_α atomu olup, ikinci gelen C atomu bazen C bazen C_β atomu olarak adlandırılır (5).

Protein belkemiğinde her amino asidin 3 bağı bulunur. Bu bağlardan $C_\beta - N$ peptit bağı dönme hareketi yapamazken $N - C_\alpha$ bağı ve $C_\alpha - C_\beta$ bağı dönme hareketi yapar. N ve C_α atomları arasındaki bağı dönüş açısına *phi* (Φ), C_α ve C_β atomları arasındaki bağı dönüş açısına *psi* (Ψ) denir. Bu iki açının farklı kombinasyonları ile protein üç boyutlu yapısı şekil alır. Ancak özellikle yan zincirin birbirine uygulayacağı kuvvet nedeniyle her değeri alamazlar. Bu açılara *dönme (torsion) açıları* denir. Dönme hareketi yapmayan bağ merkez alınarak dönmeyen kısımlar bir

düzlem içerisinde gösterilir. Bu düzlemler dönerek proteine şekil aldırarak kısımlar ile aldırılmayan kısımlar birbirinden ayrıldığından hareket daha kolay anlaşılır (6).

Molekül içi bağların yanında moleküller arası etkileşimlerden doğan bağlar da bulunmakta ve yapı oluşturmaktadır. Moleküller arası bağlara *van der Waals bağları* denir. Fizikçi Johannes Diderik van der Waals (1837 – 1923)'in moleküller arasında etkileşim olduğunu bulması sonucu bu bağlara onun adı atfedilmiştir. Molekül içi bağlara kıyasla çok zayıf bağlardır, çok daha kolay kırılabilir. İyon – iyon etkileşimi, dipol – dipol etkileşimi ve London dağılım kuvvetleri sonucu oluşan bağ ile hidrojen bağı olmak üzere 4 çeşit van der Waals bağı vardır. Atomlar elektron çekme isteklerine (*elektronegativite*) göre *polar* (kutuplu) ya da *apolar* (kutupsuz) kovalent bağ yapar.

Kovalent bağda bağ yapan iki atom arası elektronlar ortaklaşa kullanılır, ancak elektronegativitelerine göre ortak kullanılan elektronlar bir atoma biraz yakın diğer atomdan biraz uzakta bulunabilir. Bu durum da bir atomun toplam yükünün negatif diğerinin pozitif olmasına ve dolayısıyla molekülün polar olmasına sebep olur. Eğer iki atomun elektronegativitesi aynı ya da birbirine çok yakınsa ortak paylaşılacak elektronlar iki atoma da diğerinden daha fazla yaklaşmayıp ortada bulunacaktır. Bu durumda iki atomun da yükü nötr olacağından molekül apolar olur. Elektronlar elektronegativiteye göre paylaşıldıktan sonra molekülün toplamı pozitif yüklü ise *pozitif polar*, negatif yüklü ise *negatif polar*, eşit yükte ise *nötr polar* olur. Polar moleküllerin iki ucunda zıt yükler bulunduğundan bu moleküller iki kutupludur ve bunlara iki kutuplu anlamı *dipol* denir. Moleküllerin bu farklı elektriksel karakterine göre birbirleri arasında uzaklıklarına bağlı elektrostatik kuvvetler oluşur. Çekme kuvveti sonucunda ise zayıf bir bağ oluşur. Zıt yüklü moleküller arasında çekme kuvvetiyle oluşan etkileşime *iyon – iyon*, iki dipol molekül arasında çekme kuvvetiyle oluşan etkileşime *dipol – dipol etkileşimi* denir. Özellikle bir molekülün H atomu ile başka bir molekülün ondan daha elektronegatif atomu arasındaki dipol – dipol etkileşimi sonucu oluşan bağa *H bağı* denir. Bir molekül atomu geçici yük

kazandığı zaman başka moleküle olan çekim kuvvetine de *London dağılım kuvveti* denir ve atom eski haline döndüğünde bağ kopar (7).

Bu moleküller arası bağlardan özellikle H bağı ile protein birincil zincirinde bazı yerel bölgelerde bükülmeler meydana gelir. Bu bükülmeler sonucu oluşan şekillere ve bu bükülmelerle proteinin aldığı şekle proteinin *ikincil (sekonder) yapısı* denir. İkincil yapıdaki yerel bükülmeler çoğunlukla *alfa sarmal (alpha helix)* ve *beta yaprak (beta sheet)* olarak adlandırılan iki farklı yapıdadır. Sarmal ve yaprak şeklinde tanımlamalar 1900'lerin ortalarında ipek ve saç ile yapılan çalışmalardan gelmektedir. Bu çalışmalarda, 1900'lerin ortalarında ipek ve saç ile yapılan çalışmalarda proteinlerde ilk olarak keratin proteini ile sarmal yapı, bir yıl içinde ise ikinci olarak fibroin proteini ile yaprak şeklinde yapılar olduğu keşfedilmiştir. Bu yapılar protein belkemiğinin C = O bağındaki O atomu ile N – H bağındaki H atomunun birbiri ile H bağı yapması sonucu oluşur. H bağı birbirine birkaç amino asit uzakta iki amino asit arasında oluşuyorsa sarmal yapı, bükülmüş zincirde karşılıklı amino asitler arasında oluşuyorsa yaprak yapı meydana gelir. Bu yapıların oluşmasında yan zincirlerin bir rolü yoktur. Bu nedenle proteinlerde alfa sarmal ve beta yaprak yapılar sıkça rastlanır. İkincil yapısı bulunmayan azınlıkta küçük proteinler de bulunmaktadır. Bunlara *kendinden düzensiz proteinler (intrinsically disordered proteins)* denilmektedir (8).

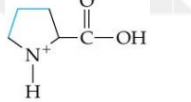
Yan zincirlerin Tablo 2.1'de gösterilen özelliklerine göre birbirine moleküller arası kuvvet uygulamasıyla proteinin ikincil yapı dışındaki kısımlarında çeşitli bükülmeler, dönüşler (*turns*), döngü (*loops*) ve bobin (*coils*) yapılar oluşur. Bu yapıların görseli ikincil yapılar kadar anlamlı değildir. Dönüslü yapılar genelde iki ikincil yapı arasında bulunur. Protein bu aşamada *üçüncül (tertiary) yapısına* ulaşmıştır ve daha fazla şekil değiştirmeye ihtiyacı yoktur. Biyolojik işlevini tek başına yerine getirecek proteinlerin son hali üçüncül yapılarıdır. Ancak bir ya da birden fazla aynı ya da farklı proteinin biraraya gelmesi ile biyolojik işlevini yerine getirebilecek proteinler üçüncül yapıdaki bu protein ya da proteinlerle biraraya gelerek *dördüncül (quaternary) yapıyı* oluşturur. Üçüncül yapıdaki her bir protein

monomer olarak adlandırılırken birçok proteinin biraraya gelmesiyle oluşmuş dördüncül yapıdaki proteine polimer ya da *multimer* denilmektedir (8).

Tablo 2.1. Protenerlerde yer alan amino asitler ve kimyasal özellikleri.

Amino Asit	Kodonları	3 Harfli Kısaltması	Tek Harfli Kısaltması	Kimyasal Yapı	Yan Zincir Karakteri
Aspartik asit	GAU, GAC	Asp	D		Negatif polar
Glutamik asit	GAA, GAG	Glu	E		Negatif polar
Arjinin	CGG, CGA, CGC, CGU, AGA, AGG	Arg	R		Pozitif polar
Histidin	CAU, CAC	His	H		Pozitif polar
Lisin	AAA, AAG	Lys	K		Pozitif polar
Asparajin	AAC, AAU	Asn	N		Nötr polar
Glutamin	CAG, CAA	Gln	Q		Nötr polar
Serin	UCU, UCC, UCA, UCG, AGU, AGC	Ser	S		Nötr polar
Treyonin	ACU, ACC, ACA, ACG	Thr	T		Nötr polar
Tirozin	UAU, UAC	Tyr	Y		Nötr polar

Tablo 2.1 (devamı). Protenlerde yer alan amino asitler ve kimyasal özellikleri.

Amino Asit	Kodonları	3 Harfli Kısaltması	Tek Harfli Kısaltması	Kimyasal Yapı	Yan Zincir Karakteri
Alanin	GCU, GCC, GCA, GCG	Ala	A	$\begin{array}{c} \text{H} \quad \text{O} \\ \quad \\ \text{H}_2\text{N}-\text{C}-\text{C}-\text{OH} \\ \\ \text{CH}_3 \end{array}$	Apolar
Glisin	GGU, GGC, GGA, GGG	Gly	G	$\begin{array}{c} \text{H} \quad \text{O} \\ \quad \\ \text{H}_2\text{N}-\text{C}-\text{C}-\text{OH} \\ \\ \text{H} \end{array}$	Apolar
Valin	GUU, GUC, GUA, GUG	Val	V	$\begin{array}{c} \text{H} \quad \text{O} \\ \quad \\ \text{H}_2\text{N}-\text{C}-\text{C}-\text{OH} \\ \\ \text{H}_3\text{C}-\text{C}-\text{H} \\ \\ \text{CH}_3 \end{array}$	Apolar
Lösin	UUA, UUG, CUU, CUC, CUA, CUG	Leu	L	$\begin{array}{c} \text{H} \quad \text{O} \\ \quad \\ \text{H}_2\text{N}-\text{C}-\text{C}-\text{OH} \\ \\ \text{CH}_2 \\ \\ \text{H}_3\text{C}-\text{C}-\text{H} \\ \\ \text{CH}_3 \end{array}$	Apolar
İzolösin	AUA, AUC, AUU	Ile	I	$\begin{array}{c} \text{H} \quad \text{O} \\ \quad \\ \text{H}_2\text{N}-\text{C}-\text{C}-\text{OH} \\ \\ \text{CH}-\text{CH}_3 \\ \\ \text{CH}_2 \\ \\ \text{CH}_3 \end{array}$	Apolar
Prolin	CCU, CCC, CCA, CCG	Pro	P		Apolar
Fenilalanin	UUU, UUC	Phe	F	$\begin{array}{c} \text{H} \quad \text{O} \\ \quad \\ \text{H}_2\text{N}-\text{C}-\text{C}-\text{OH} \\ \\ \text{CH}_2 \\ \\ \text{C}_6\text{H}_5 \end{array}$	Apolar
Metiyonin	AUG	Met	M	$\begin{array}{c} \text{H}_2\text{N} \quad \text{C} \quad \text{C} \quad \text{OH} \\ \quad \quad \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{S} \\ \\ \text{CH}_3 \end{array}$	Apolar
Triptofan	UGG	Trp	W	$\begin{array}{c} \text{H} \quad \text{O} \\ \quad \\ \text{H}_2\text{N}-\text{C}-\text{C}-\text{OH} \\ \\ \text{CH}_2 \\ \\ \text{C}_8\text{H}_6\text{N}_2 \end{array}$	Apolar
Sistein	UGU, UGC	Cys	C	$\begin{array}{c} \text{H} \quad \text{O} \\ \quad \\ \text{H}_2\text{N}-\text{C}-\text{C}-\text{OH} \\ \\ \text{CH}_2 \\ \\ \text{SH} \end{array}$	Apolar

Monomer proteinler biraraya gelerek dördüncül yapı oluştururken bazen multimer yapıda bir simetri meydana gelir. Protein simetrisi, proteinin bir eksen etrafında döndürüldüğünde yine aynı gözükmelerini anlatan bir terimdir. Simetrisinin mertebesi ise proteinin hangi açı ile döndürüldüğünde aynı gözükeceğini anlatan bir ölçüdür. Simetri mertebesi n ise proteinin döndürüldüğünde aynı gözükeceği en küçük açı derecesi $\frac{360}{n}$ 'dir. Protein bir eksen etrafında n derecesi kadar döndürüldüğünde aynı gözükyorsa bu simetriye *döngüsel (cyclic)*, iki birbirine dik eksen etrafında döndürüldüğünde aynı gözükyorsa bu simetriye *dihedral*, birbirine dik olmayan üç eksen etrafında döndürüldüğünde aynı gözükyorsa bu simetriye *kübik (cubic) simetri* denir (9).

Protein yapısında şimdiye kadar anlatılanlardan bağımsız olarak bir de *domain* yapısı bulunur. Domain, proteinin işlevsel kısmıdır. Örneğin protein bir molekülü fosforilleyecekse moleküle bağlanır ve fosforilleme işlevinden sorumlu domainden fosforiller. Proteinlerin DNA'ya bağlanma bölgelerini oluşturan alanlar da bağlanma bölgesi domaini (BBD) olarak adlandırılır. Domain genelde 40 – 350 amino asit arasında bir uzunluğa sahiptir. Bir proteinde birden fazla da bulunabilir. Proteinde birden çok domain bulunduğu bu domainleri birbirine bağlayan kısa ve şekli olmayan polipeptit zinciri dizisine ise *kendinden düzensiz diziler (intrinsically disordered sequences)* denir. Domaini olmayan proteinler de bulunmaktadır. Bunlar genelde iki biyomolekülün birarada durmasını sağlayan bağlayıcı ara proteinlerdir (8).

Protein son şeklini aldığı anda görünüşü genel olarak globular ve fibröz olmak üzere ikiye ayrılır. Küresel yapıdaki proteinler tam yuvarlak olmasa da yuvarlak yapıya benzeyen yapılardır. Fibröz yapıdaki proteinler daha çok uzun ince silindirik yapıdadır. Protein bu şekline göre fonksiyonlarını gerçekleştirir.

Protein fonksiyonuna göre birkaç temel çeşide ayrılır. Yapısal proteinler hücrenin yapısına katılır. İskele (*scaffold*) proteinleri bir fonksiyonu yerine getirmek

için birden fazla proteini biraraya toplar. Enzimler kimyasal reaksiyonun hızlandırılmasında görev alır. Zar ulaşım (membrane transport) proteinleri zardan iyon ve moleküllerin geçişini sağlar. Düzenleyici proteinler hücre aktivitesini kontrol eder. Sinyal proteinleri de düzenleyici proteinlerdir ve hücredeki sinyal alış verişini kontrol eder. Motor proteinler hücredeki diğer proteinleri hareket ettirmekten sorumludur. Transkripsiyon faktörleri, ilgili genin ifadesini kontrol eder (4).

2.1.3. Protein Katlanması

Proteinin lineer birincil yapısında kalmayıp üçüncül yapısına geçme sürecine *protein katlanması* denir. Tüm sistemler varolmak için efor sarfetmeyeceği durgun bir halde bulunmak ister. Bunun için ayrıca enerji harcamak istemez. Bu nedenle bileşenlerinin birbirine uyguladığı kuvvetler toplamının yani net kuvvetin sıfır olacağı şekilde bir konformasyona bürünür. Yan zincirler bu özelliklerine göre itme-çekme karakteri kazanarak hem birbirlerine elektriksel kuvvet uygular hem de protein su ortamında bulunduğundan apolar bağlı yan zincirler sudan uzaklaşmak isterken polar bağlı zincirler yaklaşmak ister. Su polar bağlı bir moleküldür ve polar bağlı moleküllerle elektriksel etkileşime girebilirken apolar bağlı moleküllerle bir etkileşime giremez. Katlanmadaki bu etkiye *hidrofobik etki* denir. Polipeptit zincirleri çok esnek olduğundan uygulanan bu kuvvetler proteinin birincil yapıda kalmasını zorlaştırır ve sonucunda protein bükülüp *protein katlanması* denilen olay gerçekleşir. Protein katlanmasında *şaperon* adlı proteinler proteine katlanırken eşlik etme, vakti gelmeden katlanmayı engelleme, yanlış katlanmış proteinleri onarma ve onarılmıyorsa parçalama gibi proteine katlanırken kısmi olarak yardımcı olsa da asıl katlanmayı gerçekleştiren şey çoğunlukla yan zincir karakteridir (2).

Proteinde amino asit yan zincirlerinin ve diğer kalan atomların birbirine kuvvet uygulamayacak duruma gelmesiyle proteindeki net kuvvet sıfır olur. Aşağıdaki denklemde de görüldüğü gibi, net kuvvetin sıfır olması, enerjinin minimum değere sahip olduğunu gösterir. Bu minimum enerjiye *Gibbs serbest enerjisi* denir ve Formül 2.1 ile hesaplanır (10).

$$F_{(\bar{x})} = -\nabla U_{(\bar{x})} \quad \text{Formül 2.1}$$

(Bu denklemdeki gradyeni operatörü $\nabla U = \frac{\partial U}{\partial x_1}, \dots, \frac{\partial U}{\partial x_n}$ şeklinde hesaplanır.)

2.1.4. Genetik Varyantlar ve Proteinde Fonksiyon Kaybettiren ve Kazandıran Değişiklikler

İnsanda anne ve babadan kalıtılan 22 adet eş kromozom ile biri anneden ve diğeri babadan gelen birer adet cinsiyet kromozomu (X ve Y) olmak üzere toplamda 46 kromozom bulunmaktadır (diploid yapı). Bu kromozomları oluşturan DNA'da bir değişiklik olursa bu durum genetik varyasyon olarak adlandırılır. DNA varyasyonları tüm genomda oluşabilse de protein fonksiyonlarında değişiklik yapan genetik varyasyonlar genler ya da o genin fonksiyonunu düzenleyen bölgelerde yer alan varyantlardır. Genetik varyasyonda iki kromozomdaki karşılıklı genlerden sadece biri değişmişse buna *heterozigot varyant*, ikisi de değişmişse buna *homozigot varyant* adı verilir. Genetik varyantın ifadesi sonucunda ortaya çıkan protein de *varyant proteindir*. Varyant proteinler canlılarda göz rengi, yüz şekli gibi çeşitliliğe sebep olabildikleri gibi genetik hastalıkların ortaya çıkmasına da neden olabilirler. İstenmeyen ve hastalığa sebep olan varyantlara özel olarak *mutant*, değişme sürecine de *mutasyon* denir.

Genetik varyantlar hücre tipine, büyüklükleri ve protein etkilerine göre gruplandırılabilir. Hücre tipine göre germ-hattı (*germline*) varyantları ve somatik varyantlar olarak ikiye ayrılır. Germ-hattı, üreme hücrelerinin (yumurta ve sperm) oluşumu boyunca izlediği yoldur. Germ-hattı varyantları yumurta ve sperm hücrelerinde meydana gelen varyantlardır. Üreme hücrelerinde meydana geldiğinden genetik olarak yavruya aktarılır ve etkileri yavrunun yaşamı boyunca sürer. Somatik varyantlar ise vücut hücrelerinde meydana gelen varyantlardır. Bireyde ortaya çıkıp bireyi etkiler ve gelecek nesillere aktarılmaz. Genelde çevresel etkilerden kaynaklı olup kişinin hayatının bir anında ortaya çıkabilir.

Büyükliklerine göre genetik varyantları sınıflayabilmek için önce DNA'nın yapısından bahsetmek gerekir. DNA iki zincirli sarmal bir yapıdır. Bu zincirler adenin, timin, guanin ve sitozin bazlarının biraraya gelmesiyle oluşur. Bu bazların riboz şekeri ve fosfat grupları ile oluşturduğu yapıya nükleotit denir. Sarmal yapı, iki zincirdeki karşılıklı bazların birbirine Hidrojen bağı ile bağlanmasıyla oluşur. DNA sarmalındaki hidrojen bağı ile karşılıklı bağı yapı DNA dizisinin en küçük brimi oluşturur ve *baz çifti* olarak adlandırılır. İnsanlarda genom varyantları tek bir baz çiftinin değişime uğradığı varyantlardan kromozom değişikliklerine kadar başlıca 3 grup altında incelenebilir (11).

- 1) Kromozom değişiklikleri: 3-10 milyon baz çiftinden büyük değişikliklerdir. 46 kromozom sayısında artış ve azalmalar ile karşımıza çıkabildiği gibi yapısal yeniden düzenlenmeler ile de oluşabilirler. Çok sayıda gen içerdikleri için sendrom adı verilen ve pekçok organı ilgilendiren hastalıklarla karşımıza çıkarlar.
- 2) Yapısal varyantlar: 1000 baz çifti ile 10 milyon baz çifti arasında değişen varyantlardır. Diploid yapıyı değiştirmelerine göre; a) Kopya sayısını değiştiren silinmeler ve ikiye katlanmalar (duplikasyon) kopya sayısı değişiklikleri (Copy number variation, CNV) olarak adlandırılırlar; b) Kopya sayısını değiştirmeyen ters çevrilme (inversiyon) ve translokasyon anomalileridir. Ters çevrilmede kromozomda nükleotit dizisinin duruş yönü değişir, translokasyonda ise bir kromozomun kopan bir parçası başka bir kromozoma yapışır, böylelikle nükleotit dizisi bir kromozomdan başka bir kromozoma aktarılmış olur.
- 3) Kısa varyantlar: 1000 baz çiftinden daha kısa varyantlar kısa varyant olarak kabul edilmektedir. Bunlar tek nükleotit değişiklikleri (SNV – *Single Nucleotide Variant*) olabildiği gibi 1000 baz çiftinden küçük ekleme (insersiyon), silinmeler (delesyon) bulunabilir. Eklenme ve silinmeler beraberce indel kısaltması ile de tanımlanabilir.

Genetik varyantlar yarattıkları etkiye göre Latince iyi huylu anlamına gelen *benign*, muhtemelen benign (*likely benign*), bilinmeyen etkili (*unknown significance*), muhtemelen hastalık yapıcı (*likely pathogenic*) ve hastalık yapıcı (*pathogenic*) olmak üzere 5 sınıfa ayrılır. Hastalık yapıcı anlamına gelen patojen kelimesi Yunanca hastalık anlamına gelen “*pathos*” kelimesi ile Fransızca üretmek anlamına gelen “-*génique*” ekinin birleşiminden gelmektedir. Varyant etkisi söylenmek istendiğinde bu 5 sınıfa göre ifade edilir.

Protein etkilerinde ise tek nükleotit değişiminde sessiz (*silent* veya *synonymous*), yanlış anlamlı (*missense* veya *nonsynonymous*) ve dur kazanımı (*stop gain* veya *nonsense*), birden fazla nükleotit değişiminde indel (*insertion – deletion*), çerçeve kayması (*frameshift*) ve uç birleştirme meydana gelir. Sessiz varyantlarda, bir kodon birden fazla amino asite karşılık gelebileceğinden, protein amino asit sekansında bir değişiklik olmaz. Yanlış anlamlı varyantlarda tek nükleotit değişimi amino asit dizisinde bir amino asit değişikliğine karşılık gelir. Bunun patojenik etkisi olabilir ya da olmayabilir. Dur kazanımında varyasyonla amino asit dur kodonuna değişir ve amino asit dizisi orada kesilir. Bu proteinler genelde iş yapabilir halde olmaz. İndel varyantlarda nükleotit dizisinde birden fazla nükleotide silinme ya da birden fazla nükleotidin eklenmesi söz konusudur. Çerçeve kaymasında birden fazla nükleotide eklenme ya da silinme gibi bir değişiklik meydana gelmesinin ardından amino asidi kodlayan üçlü nükleotit yapısı kayar ve farklı üçlü gruplar haline gelip karşılık geldiği amino asidi değiştirir. Uç birleştirmede ise birleşen kısımlarda yanlış birleşme sonucu proteine dönüşen bölgede azalma veya proteine dönüşmeyen bölgeden kalıntılar meydana gelir. Bunun sonucunda mRNA’dan olması gerekenden farklı bir protein oluşur.

Genetik varyantların proteinlerin dizisi üzerindeki etkileri proteinlerin işlevlerini de farklı şekillerde etkileyebilir. Genlerin kodladığı proteinlerin işlevini azaltan ya da yok eden mutasyonlar fonksiyon kaybedici mutasyonlar (*loss-of function mutations*) olarak tanımlanmaktadır. Fonksiyon kaybettiren mutasyonlarda gen işlevsel bir protein kodlayamaz hale gelmiş olabilir. Bu şekilde, gen ürünü işlev

ve fonksiyonunu tamamen kaybetmişse buna “null” alel veya “amorf” alel denir. Orijinal fonksiyonların bir bölümü muhafaza ediliyorsa bu da “hipomorf” alel olarak adlandırılır. Hipomorfik mutasyonlar gen ürününü azaltan ancak tamamen yok etmeyen mutasyonlardır. Fonksiyon kaybettiren mutasyonlar her iki alelde mutasyonla ortaya çıkan çeknik kalıtım gösteren hastalıklarda daha çok görülürler.

Herhangi bir genin transkripsiyonunu artıran, susması gereken bir genin işlevini devam ettirmesine neden olan ya da gen ürününe orijinal işlevinden farklı yeni ve anormal işlevler kazandıran mutasyonlar ise işlev kazandıran mutasyonlar (*gain-of-function mutations*) olarak adlandırılırlar. Fonksiyon kazandıran mutasyonlar genellikle baskın (dominant) ya da eksik dominans (semidominant) kalıtım gösteren hastalıklarda ortaya çıkmaktadır. Gen ürünü üzerinde herhangi bir etkiye sahip olmayan varyasyonlar ise nötral mutasyonlar olarak adlandırılırlar.

2.1.5. Proteinler Arası Yapı

Proteinlerin omurgası (backbone) benzer proteinler arasında yan zincirlere göre daha fazla korunma eğilimindedir. Bir protein kendi ile dizi açısından homoloji gösteren proteinlerle 3 boyutlu yapısı açısından da benzerdir. Ancak dizi ile yapı arasındaki ilişki her zaman keskin sınırlarla tarif edilmemektedir. Dizileri benzer olduğu halde 3 boyutlu yapısı farklı olan proteinler olduğu gibi, farklı dizilere sahip olduğu halde 3 boyutlu yapıları açısından benzer olan proteinler de mevcuttur.

2.2. ALX4 Geni ve Proteini

2.2.1. Homeobox Genleri ve Proteinleri

Homeobox gen ailesi, homeodomain isimli DNA bağlanmasını sağlayan protein motifini içeren genleri içermektedir. Bu ailenin pek çok üyesi çok hücreli canlılarda embriyonik gelişimi kontrol eden transkripsiyon faktörlerini (TF) kodlayan genlerden oluşmaktadır. İlk kez meyve sineği *Drosophila Melanogaster*'in

embriyolojik gelişiminin incelenmesi sırasında fark edilmiş, ardından mantarlardan omurgalı canlılara kadar tüm çok hücreli ökaryotik canlılarda embriyonik gelişimde önemli roller oynadıkları anlaşılmıştır. Homeobox genleri farklı özellikteki proteinleri kodlamasına karşın, yaklaşık 60 amino asitten oluşan ve yüksek oranda korunmuş homeodomain motifleri içermeleri ile karakterize edilirler (12). Homeodomain proteinlerinin çoğunun erken embriyonik dönemde morfogenezde ve hücre farklılaşmasında önemli rolleri bulunmakta, birçoğu ise insan hastalıkları ve konjenital anormalliklerle ilişkilendirilmektedir.

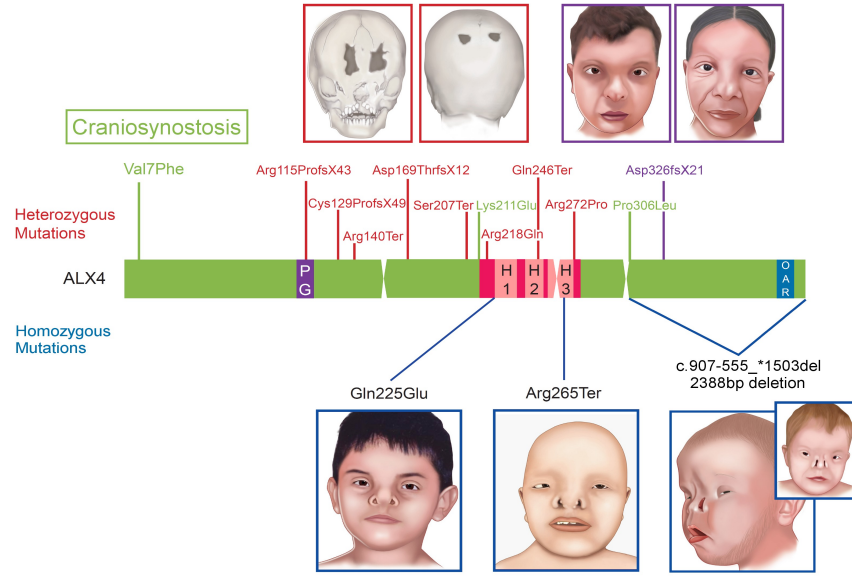
Günümüzde ökaryotik genomda filogenetik analizler kullanılarak üst sınıflara, sınıflara, alt sınıflara veya gruplara ayrılmış olan ve 235 tanesi fonksiyonel, 65 tanesi psödogen olmak üzere yaklaşık 300 tane homeobox geni bulunduğu tahmin edilmektedir (12). Bu genler bünyesinde farklı sayıda gen içeren toplamda 11 üst sınıf (ANTP, PRD, LIM, POU, HNF, SINE, TALE, CUT, PROS, ZF ve CERS) altında toplanmıştır. Tez konusu olan *ALX4* geninin bağlı olduğu ALX gen ailesi ise Paired (PRD) üst sınıfı altında yer almaktadır.

2.2.2. *ALX* Gen Ailesi

Bir homeobox gen sınıfı olan *ALX* ailesinde *ALX1*, *ALX3* ve *ALX4* olmak üzere 3 gen bulunmaktadır. Omurgalıdaki ALX gen ailesinin üç üyesi omurgalı evriminin erken dönemlerinde meydana gelen iki tüm genom duplikasyonunun sonucunda ortaya çıkmıştır. ALX gen ailesinin dördüncü üyesi olması gereken ALX2 ise canlılarda kaybolmuş ve hiçbir omurgalıda gözlenmemiştir. (13)

ALX gen ailesine mensup 3 üyenin mutasyonları, insanlarda kafatası ve yüz malformasyonlarına (frontonasal displazi, FND) yol açmaktadır. Bunlardan *ALX4* ve *ALX1* genindeki mutasyonların yüz malformasyonları ile ilişkisi tez danışmanı Dr. Akarsu ve ekibi tarafından gösterilmiştir (14,15). Bu çalışmalar sırasında *ALX4* geninin ailenin diğer üyelerinden farklı özellikleri olduğu gözlenmiştir. *ALX1* (FND3; OMIM 613456) ve *ALX3* (FND1; OMIM 136760) gen mutasyonlarına bağlı

olarak ortaya çıkan yüz malformasyonları otozomal resesif kalıtımla uyumlu iken *ALX4* geninin fonksiyon kaybettiren mutasyonları semidominant bir kalıtım kalıbı göstermektedir. Bu tip kalıtımla uyumlu olarak homozigot fonksiyon kaybı mutasyonlarına sahip bireyler heterozigot fonksiyon kaybı mutasyonlarına sahip bireylerden daha ağır bir fenotip göstermektedirler Fenotip çeşitliliği bununla da sınırlı kalmayıp gerek heterozigot gerekse homozigot mutasyona sahip bireylerde beklenenin dışında yeni fenotipler ortaya çıkabilmektedir (Şekil 2.1). *ALX4* geninde heterozigot fonksiyon kaybı mutasyonlarına sahip bireylerde görülen klasik fenotipte kafatasının parietal kemik bölgelerindeki kemik oluşumunun eksik gerçekleşmesine bağlı kemik defektleri (parietal foramina) ortaya çıkmaktadır. Ancak bazı nadir heterozigot mutasyonlarla bu fenotipe yüz malformasyonları eklenebilmekte, ya da tamamen farklı bir hastalık olan kraniyosinostoz (kafa kemiği eklem çizgilerinin erken kapanması) görülebilmektedir. Fonksiyon çalışmaları kraniyosinostoz fenotipinin fonksiyon kazanımı mekanizmaları ile açıklanabileceğini desteklemektedir. *ALX4* genindeki çeşitli mutasyonların farklı işlev değişimlerine neden olması, farklı kalıtım kalıpları göstermesi ve farklı fenotipler ortaya çıkarması bir gen için nadiren karşılaşılan bir durumdur. Bu nedenle, tez kapsamında *ALX4* proteininin *in-silico* protein modellerinde detaylı incelenmesi ve farklı varyantların etkilerinin modeller kapsamında ortaya konabilirliğinin araştırılması hedeflenmiştir.



Şekil 2.1. *ALX4* mutasyon dağılımı ve fenotip çeşitliliği. *ALX4* protein yapısı ve ilgili domanleri ortadaki yeşil kutu üzerinde göstermiştir. Mor kutu proteindeki poliglutamin (PG) uzantısını, mavi kutu OAR alanını temsil ederken kırmızı ile gösterilen kısım üç heliks yapısındaki (H1, H2, H3) homeodomaini göstermektedir. Heterozigot ve homozigot mutasyonlara karşılık gelen protein dizisi değişiklikleri protein yapısının üstünde ve altında yer almaktadır. Yeşil ile gösterilen mutasyonlar kraniyosinostoz; kırmızı ile gösterilenler parietal foraminaya; mavi ile gösterilenler ise *ALX4* ile ilişkili FND'nin değişken fenotiplerine yol açmaktadır. *ALX4* geni homozigot mutasyonları farklı derecelerde frontonasal displaziye yol açarken saç gelişiminde normalden (solda) alopesi (ortada) ve yüz hipertrikozuna (sağda) kadar değişen geniş bir fenotipik çeşitlilik görünmektedir. Heterozigot değişikliklerde FND fenotipi beklenmezken p.Asp326LeufsX81 mutasyonunun (mor renkte) heterozigot durumda FND fenotipine yol açtığı görünmektedir (16 numaralı yayından alınmıştır).

2.2.3. *ALX4* Proteini

ALX4 proteini, *ALX4* geninin kodladığı ve UniProt'a göre 411 amino asit uzunluğunda olan bir protein olup bu proteinin 2 önemli domaini bulunmaktadır. Homeodomain, paired tipi bir homeodomain olup 214.-276. amino asitler arasında yer almaktadır. Bu domain DNA bağlanmasından sorumlu olup hedef genlerin

civarındaki DNA dizilerine bağlanmayı sağlar. Böylece hedef gen civarına yerleşen ALX4 proteini transkripsiyon aktivasyonunda rol alır. OAR domaini ise 391.-404. amino asitler arasında yer almakta olup bu domainin işlevi kesin olarak bilinmemektedir. Bazı çalışmalar bu domainin proteinin otokontrolünde görev alabileceğini düşündürmektedir. ALX protein ailesinin üyeleri birbirleri ile N terminal bölgelerinden kafa-kafaya dimer yapılar oluşturarak hedef DNA dizilerine bağlanmaktadır. Tüm bu bilgilere karşın ALX Proteininin yapısı ile işlevlerini ilişkilendiren yayınlar kısıtlıdır. Deneysel bir yöntem ile ALX4 proteininin tamamı için üç boyutlu yapı mevcut olmamakla birlikte sadece homeodomain kısmının 3 boyutlu yapısı nükleer manyetik rezonans (NMR) ile ortaya konmuştur (PDB Kodu: 2M0C).

2.3. Protein Yapısının Eldesi

Bir molekül olan proteinlerin biyolojik işlevleri, kimyasına göre aldıkları üç boyutlu şekle bağlıdır. Proteinlerin üç boyutlu yapısını bilmek protein biyolojisinin daha iyi anlayabilmeye, proteinlerin diğer biyomoleküller ile ilişkilerini anlamaya, protein dizayn edebilmeye ve ilaç geliştirmeye yardımcı olmaktadır (17,18).

Bir proteinin üç boyutlu yapısı elde edildiğinde bu yapıyı farklı şekillerde ifade etmek mümkündür. Bunun en basit şekli proteini oluşturan her bir atomun XYZ koordinatlarındaki pozisyonları ve atomlar arasındaki bağları göstermektir. Ancak bu yöntem proteinin üç boyutlu yapısını sadece bir bakış açısından ortaya koyması nedeniyle sınırlıdır. Bir diğer yöntemde ise protein ait atomların birbirine göreli pozisyonlarını iç (internal) koordinatlar şeklinde ifade etmektir. İç koordinatlar, bir peptit bağının rotasyonunu ve atomlar arasındaki açığı belirtmekte, ayrıca yan zincirin omurgaya göre konumunu kaydetmektedir. İç koordinatlarla proteinin üç boyutlu bilgisini ifade etmek protein farklı bakış açılarından bakabilecek şekilde döndürmeyi sağlaması nedeniyle kullanışlı bir yöntemdir.

Proteinlerin üç boyutlu yapılarının incelenmesi için öncelikli üç boyutlu yapıyı elde etmek gerekmektedir. Yapı eldesi için hem deneysel hem de hesaplamalı çeşitli yöntemlere aşağıda yer verilmektedir (19).

2.3.1. Protein Yapısının Deneysel Eldesi

Protein yapısını deneysel metotlarda elde etmek için kullanılan yöntemler en çok kullanılanlardan en az kullanılanlara göre sıralandığında x-ray kristalografisi, çözelti (solution) NMR, elektron mikroskopi, katı hal NMR, hibrit, nötron difraksiyon, fiber difraksiyon, elektron kristalografisi ve solution scattering metodlarını karşımıza çıkmaktadır. Bu yöntemler dışında başka yöntemler de bulunsa da daha az tercih edilmektedir (20). Protein Data Bank (PDB) bahsedilen yöntemler kullanılarak elde edilmiş protein yapılarının bulunabileceği bir veri tabanıdır (21). Bu veri tabanı kullanılarak proteinlerin üç boyutlu yapıları PDB uzantılı dosya olarak indirilebilir. Bu dosyalarda proteinde bulunan her bir atom ve bu atomların x,y,z koordinat sistemindeki lokasyon bilgileri bulunur.

Bu yöntemler içinde en çok kullanılan ilk 2 yöntemden X-ray kristalografide X ışını protein kristalinden geçerek saçılır ve dedektöre çarpar. Dedektör saçılan ışınların lokasyonlarını okur ve bu bilgiyi bilgisayara iletir. Bilgisayar bu lokasyonların açı bilgisinden yoğunluk haritası çıkararak atom konumlarını elde eder. NMR’da ise protonlardan atom uzaklıkları elde edilerek protein yapıları ortaya konulmaktadır. Bu deneysel yöntemlerde atomlar doğrudan gözlenmediği ve dolaylı verilerden yapılan analizlerle sonuçlara ulaşıldığı için “protein gözlenmez çözülür” denilmektedir. Deneysel veriler üzerinden yapılan analizlerde protein dizisi ile en çok uyuşan konformasyon protein yapısını oluşturmakta kullanılır (22).

2.3.2. Protein Yapısının Hesaplamalı (*in silico*) Yöntemlerle Eldesi – Protein Modelleme

Protein modellemesi, protein amino asit sekansından üç boyutlu protein yapısının elde edilmesidir. Bu fikir ilk kez 1973'te Anfinsen tarafından ortaya atılmış ve Anfinsen'in dogması olarak kalmıştır. Anfinsen, amino asitlerin birbiri arasında oluşturduğu kuvvetlere göre proteinin bir enerjiye sahip olması ve bu enerjiyi minimum halde tutacak şekilde katlanması için gerekli tüm bilginin amino asit sekansında olduğunu söylemiştir. İnsan genom projesi sonrasında DNA dizi ve beraberinde protein sekans verileri artmıştır. Böylelikle protein yapısını sekans tahmin edebilecek birçok biyoinformatik yöntem geliştirilmiştir (23).

Bir proteinin deneysel olarak yapısını elde etmek hem çok maliyetli hem de çok zahmetlidir. Bu noktada hem zamandan hem bütçeden tasarruf edilebilirliği hem de süreci daha yakından inceleyebilmek, molekülün şekil ve hareketini yakından görebilmek açısından hesaplamalı yöntemler kullanışlı bir protein yapı eldesi yöntemi olarak karşımıza çıkmaktadır. Hesaplamalı yöntemler bilgisayar ortamında gerçekleştirildiğinden ve bilgisayardaki işlemci silikondan yapıldığından bu çalışma tarzına Latince'de "silikonda" anlamına gelen *in silico* adı verilmiştir.

Bu yöntemlerin esası protein'in backbone yapısının korunması ve daha çok yan zincirlerinin değişmesini göz önüne alır. Yan zincire göre konformasyonlar da bilinen proteinlerin konformasyonları kullanılarak hesaplanabilmektedir. Böylelikle bir proteinin sekansından backbone ve yan zinciri çözüldüğü zaman üç boyutlu yapısı da çözülebilmektedir. Proteinin backbone'u çözüldükten sonra N terminus ve C terminus kısımları eklenip, backbone'da peptit bağı yapan bölgeler ve torsiyon açıları belirlendiğinde yan zincir sadece bu torsiyon açılarının izin verdiği bölgelere doğru dönebildiğinden bu bilgi ile bir protein modeli çıkarılır. Çıkarılan proteinin serbest enerjisi hesaplanıp uygunluğu değerlendirilir.

Bugün proteinlerin hesaplamalı yöntemle modellemeleri pek çok farklı amaçla kullanılmaktadır. CASP (*Critical Assessment of Protein Structure Prediction*) yarışması protein modelleyicilerden en iyisini seçebilmek için her 2 senede bir yarışma düzenlemektedir. Bu yarışmada deneysel olarak çözülmüş ama

yayınlanmamış protein dizileri modelleyicilere yüklenerek üç boyutlu yapısı elde edilmeye çalışılmaktadır. Modelleyicilerin ürettiği modeller deneysel modellerle karşılaştırılıp bir Z skor üretilmekte ve en yüksek Z skoruna sahip modelleyici en iyi modelleyici olarak kabul edilmektedir (24, 25).

3 çeşit in silico protein modeli elde etme yöntemi vardır: şablon tabanlı (*template-based*), şablon tabansız (*template-free; Ab-initio*) ve hibrit.

Şablon Tabanlı Modelleme

Şablon tabanlı modellemede modellenecek proteinin sekansına benzer sekanslar bulunduktan sonra bu sekansların sahibi proteinlerin üç boyutlu yapıları incelenir ve modellenecek proteinin modeli oluşturulur. Bu aşamalar homoloji modelleme ve katlanma tanıma (*threading/fold recognition*) olarak iki farklı yöntemle yapılabilir.

Homoloji Modelleme: Çoklu sekanslama sonucu sorgulanan proteine benzer protein sekansları bulunduktan sonra en yüksek benzerlik gösteren ve üç boyutlu yapısı bilinen protein sekansı seçilir. Bu sekans ile sorgulanan sekans arasında korunan bölgeler belirlenerek bu bölgelerin üç boyutlu yapıdaki karşılıklarında bulunan atomların koordinatları direkt sorgulanan sekansın ilgili bölgelerine atanır. Döngü modelleme algoritması ile kalan kısımlar da modellenir ve sorgulanan protein sekansının üç boyutlu yapısı elde edilmiş olur. Dolayısıyla evrimsel olarak aynı kökten gelen proteinler arasında arama yapıp homoloji gösteren proteinlerin üç boyutlu yapıları bilinmesi halinde kullanılabilen bir yöntemdir. Aynı kökten gelen proteinlerin yapıları birbirine benzediğinden doğruluğu çok yüksektir. Modellemesi yapılacak proteine homoloji gösteren uygun protein dizilerini saptamak için BLASTP arama motoru ile PDB veri tabanında arama yapılabilir.

Katlanma Tanıma: Bu yöntem homoloji modelleme yöntemine benzerlik gösterir. Ancak sorgulanan protein sekansına en benzer üç boyutlu yapısı bilinen

protein sekansını bulup oradan model çıkarmak yerine, evrimsel bağı olsa da olmasa da benzer tüm sekansları tarayarak her birinin katlanmasına bakar ve buradan öğrendiği ile sorgulanan sekansın üç boyutlu protein modelini çıkarır. Aynı evrimsel kökten gelen proteinlerin yanı sıra farklı evrimsel kökten gelmelerine rağmen aynı işi yapan yani aynı yapıya sahip proteinler olabileceğinden bunları da göz önünde bulundurarak doğruluğu artırır.

Şablon Tabansız Modelleme

Bu modellemede sadece fizik yasaları kullanılır. Bu yasalar klasik fizik yasaları yani Newton mekaniğinin öne sürdüğü yasalardır. Hesaplamalar Newton ikinci yasasının genişletilmiş hali olan Hamilton denklemini kullanır ve enerji hesabı yaparak bunu minimuma indirdiği konfigürasyonu model olarak sunar. Herhangi bir gözleme dayanmaz. Bu nedenle herhangi bir homoloji göstermeyen veya homologlarının üç boyutlu yapısının da bilinmediği proteinler için kullanılabilir. Ancak doğruluğu çok düşüktür.

Bu soruna belki kuantum fiziği çözüm bulabilir. Ancak günümüzde kuantum fiziğinde çoklu sistem çalışılmamaktadır. Kuantum fiziği yasaları tek parçacığı açıklamak üzerine kurulmuştur. Parçacık sayısı çoğaldıkça tek parçacığı açıklamak için kullanılan denklemler ne kadar parçacık varsa her biri için tek tek yazılarak hesaplaması yapılır. Ayrıca bu parçacıkların birbirine etkisinin de hesaplanması gerekir ki bu hesaplamaları daha da karmaşıktır ve uzatır. Dolayısıyla fizik yaklaşımlarının yetersizliğinden dolayı şablon tabansız modellemeler iyi sonuç vermemektedir. Ancak modellenecek protein için yeterince iyi bir şablon bulunamazsa en iyi çözüm yolu bu yöntemlere başvurmaktır.

Hibrit Modelleme

Bu yöntemde şablon tabanlı ve şablon tabansız modelleme yöntemleri birlikte kullanılır. Doğruluğu iki yöntemin arasında kalsa da yapay zeka ile geliştirilip artırılmaya çalışılmaktadır.

2.3.3. CASP Yarışması ve Günümüzde Kullanılan Modelleyiciler

CASP (*Critical Assessment of protein Structure Prediction*) protein yapısı modelleyicilerinin ne kadar doğru modelleme yaptığını ölçen ve iki senede bir yapılan bir yarışmadır. İlk CASP yarışması CASP1 adı ile 1994'te yapılmış ve CASP15 adı ile günümüze (2022) kadar sürmüştür. CASP16 yarışması ise 2024 yılının ağustos ayında yapılacaktır. Tüm yarışma sonuçları internet sitesinde bulunmaktadır. CASP1'den CASP7'ye (2006) kadar olan sonuçlar puanlamanın değişmesinden dolayı 2019 yılında tekrar hesaplanıp internet sitesinde paylaşılmaktadır.

CASP yarışmasında yarışmaya modelleyiciler ve henüz yayınlanmamış deneysel olarak çözülmüş proteinlerin üç boyutlu yapıları katılır. Modelleyicilere bu protein yapıları verilir ve modellemeleri beklenir. Ardından modelleyici modelleri ve deneysel model karşılaştırılıp modelleyicinin ne kadar iyi modelleme yaptığı ölçülür ve o sene yarışmaya katılan modelleyiciler arasında bu ölçüye göre sıralama yapıp yayınlanır. İki model GDT_TS değerlerinden Z değeri hesaplayarak karşılaştırılır.

Yarışmada modelleyiciler tüm modelleri modelledikten sonra her biri için sahip olduğu Z değerleri toplanır ve toplam Z değeri elde edilir. Toplam Z değeri ne kadar yüksekse modelleyici o kadar iyi modelleme yapıyor demektir. CASP, toplam Z değerlerine göre o sene yarışmaya katılan tüm modelleyicileri sıralar ve yayınlar. Bu sıralamaya göre en iyi modelleyici seçilebilir.

CASP yarışma sonuçları ve literatür taraması sonucunda tezde kullanılmasına karar verilen günümüz modelleyicileri alt başlıklarda kısaca açıklanmıştır.

Tablo 2.2. CASP13-15 yarışmalarında en yüksek değerleri alan modelleyiciler ve Z değerleri.

Yarışma Adı	Yarışma Yılı	Modelleyici Adı	Toplam Z Değeri	Z Değeri Sırası
CASP15	2022	Yang-server (trRosetta)	90.4273	1
		UM-TBM (i-TASSER)	89.2119	2
		Multicom	62.0521	7
		RaptorX	45.273	26
CASP14	2020	AlphaFold2	244.0217	1
		tFold_human	65.2157	6
		Zhang-server (i-TASSER)	62.9122	9
		Baker-Rosettaserver	55.2993	18
		ProQ2	52.3893	23
		Zhang-TBM (LOMETS)	51.2506	28
		RaptorX	44.0923	43
		Yang-server (trRosetta)	39.92	50
Baker-Robetta	10.7738	84		
CASP13	2018	A7D (AlphaFold)	128.0693	1
		Multicom	99.9886	3
		Zhang-server (i-TASSER)	89.7597	5
		ProQ2	61.8821	25
		Baker-Rosettaserver	55.7125	32
		RaptorX-Contact	54.7082	34
		Falcon	39.5169	44
		Yang-server (trRosetta)	36.9795	48
Yasara	16.4714	61		

AlphaFold ve AlphaFold2

Protein diziliminden üç boyutlu protein yapılarını tahmin etmek için kullanılan ve temelde fiziksel etkileşimler ve evrimsel korunmuşluğa odaklanan yaklaşımlarda kısmi başarı elde edilmesi üzerine biyoinformatik ve fiziksel yaklaşımların bir kombinasyonu olarak PDB sınırlı verilerinden öğrenen bileşenler oluşturma yoluna gidilmiştir. AlphaFold proteinlerin doğru bir şekilde katlanma yapılarını tahmin etmek için, çoklu dizi hizalaması kalıntı-kalıntı uzaklık tahmini ve parça montajı (fragment assembly) yöntemlerini içeren bir derin sinir ağı algoritması kullanmaktadır (26). Google'ın yapay zeka şirketi DeepMind tarafından geliştirilmiştir. Yapay zeka yöntemleri ile geliştirilen sinir ağı AlphaFold ve AlphaFold2 adı altında CASP13 ve CASP14 protein modelleme yarışmalarında en yüksek toplam Z skoruna sahip olarak en iyi yapı tahmin aracı olarak seçilmiştir (27).

ColabFold

ColabFold, AlphaFold2 ve MMseqs2 birlikte kullanan ve bu arama algoritması sayesinde AlphaFold2'den daha hızlı çalışmakta olan bir modelleme programıdır. Hem şablon tabansız hem de şablon tabanlı modelleme yapabilmektedir. Şablon tabanlı modellemede pdb100 veri tabanını kullanmakta olup isteğe bağlı olarak özel veri tabanları da tanımlanabilmektedir.

D-I-TASSER

D-I-TASSER Zhang laboratuvarında geliştirilen ve tamamen şablon tabansız modelleme yapan bir yazılımdır. Sadece küçük diziler halinde benzer dizileri alarak çoklu dizi hizalaması (Multiple Sequence Alignment, MSA) için geniş IMG/JGI metagenomik veri tabanını ve uzaklıklar için AlphaFold2 ile aynı uzaklık hesaplamalarını kullanmaktadır.

i-TASSER

i-TASSER da Zhang laboratuvarında geliştirilmiştir. D-I-TASSER programının şablon tabanlı modelleme yapan versiyonudur.

RoseTTAFold2

RoseTTAFold2 aracı homoloji aracılı modelleme yapmaktadır.

2.3.4. Homolog Protein Araştırması

Homolog protein araştırması BLASTP arama motoru ile PDB veri tabanında arama yapılması yolu ile uygulanır. Bu işlem hangi modelleme aracının kullanılacağını belirlemek için modelleme öncesinde yapılan bir işlemdir (28, 29). PDB'de homolog protein bulunmadığı zaman pLM-BLAST arama motoru ile uzak

homolog protein araması yapılarak hedef proteine benzeyen başka protein bulunmasına çalışılır (30, 31).

2.3.5. Protein Yapısının Görselleştirilmesi

Protein yapısının görselleştirilmesinde VMD (Virtual Molecular Dynamics) aracı ve Discovery Studio aracı en çok kullanılan ve iyi bilinen araçlardır. VMD ile RMSD değerlerini hesaplama, farklı modelleri üst üste bindirme, PSF üretme gibi çeşitli işlemler tamamlanabilirken Discovery Studio ile dizide aranan amino asit ve bölgeyi modelde detaylı inceleme ve yüksek pikseli görselleştirme yapılabilmektedir (32, 33).

2.3.6. Proteinlerde Yapı Kontrolü ve Yapıların Karşılaştırılması

Proteinleri karşılaştırmada genelde RMSD değeri kullanılır. RMSD değeri ne kadar düşüğe iki protein veya proteinler birbirine o kadar benzerdir. RMSD değeri Formül 2.2 ile hesaplanır.

$$RMSD_{(a,b)} = \sqrt{\frac{1}{n} \sum_{i=1}^n [(a_{ix} - b_{ix})^2 + (a_{iy} - b_{iy})^2 + (a_{iz} - b_{iz})^2]} \quad \text{Formül 2.2}$$

Eğer bir protein yapısından başka bir protein yapısı elde edilecekse, yani homoloji modelleme yapılacaksa, tüm atomların hesaba katılması gerekir.

MolProbity, yapı stereokimyasını inceleyerek amino asitlerin amino asitlerin konumu ve birbiri ile durumuna göre puanlama yapan bir araçtır (34). Yapı kontrol ve karşılaştırmasında kullanılabilir.

2.3.7. Minimizasyon

Termodinamik yasalarına göre bir molekül doğada, bulunabileceği en düşük global minimum enerji seviyesinde bulunur. Bundan dolayı elde edilen protein kuvvet alanı hesaplamaları ile minimum enerji düzeyine indirilip son hale getirilir. Bu işlem minimizasyon işlemi olarak adlandırılır. Minimizasyon için VMD görselleştirme aracı ile birlikte çalışan NAMD yazılımı kullanılmaktadır (35).



3. GEREÇ VE YÖNTEM

3.1. Yabancı Tip ALX4 Proteininin Üç Boyutlu Yapısının Araştırılması

PDB ve UniProt veri tabanlarından ALX4 için üç boyutlu yapı araştırması yapıldı.

3.2. Kullanılacak Modelleyicilerin Belirlenmesi

ALX4 veya ALX4'e benzer çözülmüş protein saptanmadığından hem güvenilirliği yüksek ve hibrit modelleyicilerle şablon tabansız hem de geniş veri tabanına sahip modelleyicilerle şablon tabanlı modelleme yapılmaya karar verildi. Bundan dolayı modelleyicilerin eriştikleri veri tabanının genişliğine, veri tabanında arama yaparken kullandıkları algoritmanın hassaslığına ve modelleme yaparken modelleme türüne göre diziyi benzeteceği kümeler kullanıp bunlarda iyileştirme yapmasına dikkat edilerek, son 10 yılda CASP yarışmalarında (CASP11 yarışmasından CASP15 yarışmasına kadar) Z skoru yüksek olan ve literatür taraması sonucu en çok kullanıldığı görülen modelleyiciler seçildi. Ayrıca CASP yarışmasında yüksek Z skoru elde ederek başarılı olan i-TASSER modelleyicisinin şablon tabansız modelleme yapan D-I-TASSER sunucusu da eklendi.

3.3. Yabancı Tip ALX4 Proteininin Üç Boyutlu Yapısının Tahmin Edilmesi

3.3.1. Dizin Belirlenmesi

Yabancı ALX4 protein dizisinin eldesi için NCBI veri tabanı kütüphanesinden Protein veri tabanında arama yapıp ve buradan referans dizi alındı.

3.3.2. Homolog Protein Araştırması

BLASTP arama algoritması ile PDB veri tabanında homolog protein araması yapıldı. Aynı zamanda pLM-BLAST arama motoru kullanılarak eğer varsa bu proteinlerin benzerini PDB veri tabanında aramak için uzak homolog protein arandı.

3.3.3. Belirlenen Modelleyicilerde Modelleme

Modelleyicilerde modelleme, her birinin altında belirtilen adımlar takip edilerek yapıldı. Her oluşturulan algoritmadan 5 model üretildi.

AlphaFold2

1. *sequence_1* kısmına 411 amino asit uzunluğundaki ALX4 dizisi girildi.
2. Rahatlatma ayarı (*run_relax*) kullanıldı.
3. En yüksek geri dönüşüm sayısı (*multimer_model_max_num_recycles*) en çok kullanılan değer olarak 3 seçildi.
4. Tüm algortima “Çalışma zamanı → Tümünü çalıştır” seçilerek çalıştırıldı. Bu adım, aynı parametrelere sahip 5 model üretmek için 5 kez tekrarlandı.
5. Tüm adımlar 20, 24 ve 48 geri dönüşümlü modelleri de test etmek için sadece *multimer_model_max_num_recycles* parametresi ilgili değere değiştirilerek tekrarlandı.

ColabFold

1. *query_sequence* kısmına 411 amino asit uzunluğundaki ALX4 dizisi girildi.
2. *num_relax* parametresi 5 seçilerek, oluşturulacak 5 modelin hepsinde Amber kuvvet alanı kullanılarak rahatlatma yapıldı.
3. Şablon tabanlı modelleme için *template_mode* parametresi *pdb100* seçildi.
4. Veri tabanında en geniş aramayı yapabilmesi için *msa_mode* parametresi *mmseqs2_uniref_env*, *pair_mode* parametresi *unpaired_paired* ve *pairing_strategy* parametresi *greedy* olarak seçildi.

5. Monomer protein modellemesi yapıldığından *model_type* parametresi *auto* olarak bırakıldı.

6. Geri dönüşüm sayısı *num_recycles* 3 olarak seçildi.

7. Erken durma noktasının belirlenen özelliklere göre en uygun nokta belirlenmesi için *recycle_early_stop_tolerance* parametresi *auto* seçildi.

8. *relax_max_iterations* parametresi modellemenin çok uzun sürmemesi için 200 olarak bırakıldı.

9. İstatistiksel olarak belirsizliğin en uygun şekilde belirlenmesi için *max_msa* parametresi *auto* olarak belirlendi.

10. Tezde model tekrar üretilmeyeceğinden *seed* belirlenmedi.

11. Tüm algoritma “Çalışma zamanı → Tümünü çalıştır” seçilerek çalıştırıldı.

12. Tüm adımlar 24 ve 48 geri dönüşümlü modelleri de test etmek için sadece *num_recycles* parametresi ilgili değere değiştirilerek tekrarlandı.

13. Şablon tabansız modelleme için sadece *template_mode* parametresi *none* olarak değiştirilecek şekilde tüm adımlar tekrarlandı.

D-I-TASSER

1. ALX4 dizisi FASTA formatında girdi kısmına yapııştırıldı.

2. MSA inşası için geniş IMG/JGI veri tabanının da dahil edilmesi seçildi.

3. Uzaklık hesaplamalarında AlphaFold2 uzaklıklarının kullanılması tercih edildi.

RoseTTAFold2

1. Kullanılacak parametrelerin paket tarihi için *params* parametresi *RF2_jun24* olarak seçildi.

2. *sequence* parametresine dizi yazıldı.

3. ALX4’ün bilinen bir simetrisi olmadığından *sym* parametresi *X* olarak, monomer modelleme yapılacağından *order* parametresi 1 olarak belirlendi.

4. *msa_concat_mode* parametresi *diag* olarak seçildi.

5. *msa_method* parametresi şablon tabanlı modelleme yapması için *mmseqs2* olarak seçildi.

6. *pair_mode* parametresi *unpaired_paired* olarak seçildi.

7. *num_recycles* parametresi 3 olarak seçildi.

8. *max_msa* sayısı 256 olarak seçildi.

9. Modeller yeniden üretilmeyeceğinden seed verilmedi.

10. 5 model üretmesi için *num_models* parametresi 5 olarak seçildi.

11. Tüm algortima “Çalışma zamanı → Tümünü çalıştır” seçilerek çalıştırıldı.

12. Farklı çoklu dizi hizalama yöntemlerini test etmek için sadece *msa_concat_mode* parametresi *default* ve *repeat* yöntemlerine değiştirilmek üzere tüm adımlar tekrarlandı.

13. Ardından 24 geri dönüşümlü modelleri de test etmek için sadece *num_recycles* parametresi 24 olarak değiştirilerek tüm adımlar tekrarlandı.

14. Ardından *max_msa* sayısının 512 olarak test edilmesi için sadece bu parametre değiştirilerek tüm adımlar tekrarlandı.

15. Son olarak 2023 tarihli parametre paketinin de test edilmesi için sadece *params* parametresi *RF2_apr23* olacak şekilde değiştirilerek tüm adımlar tekrarlandı.

i-TASSER modelleyicisinde varsayılan ayarlar kullanılarak modelleme yapıldı.

3.3.4. Eksik Atomların Tamamlanması

Modellemede oluşabilecek eksik atomların tamamlanması için VMD programının PSF yaratıcısı olan AutoPSF aracı kullanıldı. Aşağıdaki adımlar takip edilerek modelleyicilerden alınan tüm modellerin AutoPSF sonuçları alındı. Aşağıdaki adımlar takip edilerek tüm modeller için PSF yaratıldı.

1. VMD aracına “File → New Molecule” yolundan modelin PDB formatlı dosyası yüklendi.

2. “Extensions → Modeling → Automatic PSF Builder” yolundan PSF yaratıcısı açıldı.

3. Analize her şeyi dahil etmesi istenerek “I’m feeling lucky” tuşu ile PSF yaratıldı.

3.3.5. Modelleyicilerin En İyi Modelinin Belirlenmesi

2M0C’nin Eldesi ve Hazırlanması

NCBI Structure veri tabanında 2M0C aratılarak 75 amino asit uzunluğundaki domainin PDB formatlı dosyası indirildi. Proteinle işlem yapılabilmesi için ilk 3 amino asidi değiştirildiğinden 411 amino asitlik proteinden farklı amino asitlere sahiptir. Bundan dolayı bu 3 amino asit kesildi ve ardından modelleyicilerin ürettiği 411 amino asit uzunluğundaki modellerin 209. amino asidinden 280. amino asidine kadar olan kısım ile hizalanarak RMSD değeri alındı. Bu işlem için 2M0C önce VMD’ye “File → New Molecule” yolundan yüklenip “File → Save Coordinates” yolunda koordinatlara *resid 4 to 75* yazılarak ilk 3 amino asidinden kesildi. Ardından aşağıdaki Python komutları ile biopython paketinden PDB dosyasındaki amino asit numaraları 209’den başlayıp 280’de bitecek şekilde yeniden numaralandırıldı.

```
from Bio import PDB
pdb_io = PDB.PDBIO()
pdb_parser = PDB.PDBParser()
pdb = "2M0C(NCBI-Structure).pdb"
model = pdb_parser.get_structure(" ", pdb)
resno = [i+209 for i in range(281)]
for model in model:
    for chain in model:
        for i, res in enumerate(chain.get_residues()):
            res_id = list(res.id)
            res_id[1] = resno[i]
            res.id = tuple(res_id)
pdb_io.set_structure(model)
pdb_io.save(pdb)
```

RMSD Değerinin Alınması

VMD’de “File → New Molecule” yolu takip edilerek RMSD değeri alınacak model yüklendi. Ardından aynı sekmeden “Load files for → New Molecule” yolundan 2M0C yüklendi. “Extensions → Analysis → RMSD Trajectory Tool” yolundan RMSD aracına gidilerek seçim alanına *protein resid 209 to 280* yazıldı. *Selection Modifiers* kısmından *Backbone* seçilerek önce “Align” ardından “RMSD” tuşu ile RMSD değeri alınıp sonuç kaydedildi.

MolProbity Değerinin Alınması

MolProbity web adresine gidilerek “Choose File → Upload → Continue” yolu izlendi. MolProbity aracından işlem yapmadan sadece değerlendirmesi istendiğinden *Currently working on* kısmından *File uploaded by user* kısmı seçildi. *Analyze all-atom contacts and geometry* seçilerek istenilen analizlerin seçileceği sayfaya gidildi. Buradan *3-D kinemage graphics* kısmında kinemaj gösterimlerinde olması istenilen özellikler *Clashes, Hydrogen bonds, van der Waals contacts, Geometry evaluation, Ramachandran plots, Rotamer evaluation, C_β evaluation, Cis-Peptide evaluation, C_αBLAM backbone markup; Charts, plots, and tables* kısmında tablolar ve grafiklerde görmek istenilen özellikler *Clashes & clashscore, Geometry evaluation, Ramachandran plots, Rotamer evaluation, C_β deviations, Cis-Peptide evaluation, C_αBLAM backbone evaluation, Suggest/report on automatic structure fix-ups, Create html version of multi-chart* seçildi. Ardından “Run programs to perform these analyses” tuşu ile analiz başlatıldı ve tüm sonuçlar kaydedildi.

Yukarıdaki adımlar takip edilerek tüm modellerin AutoPSF sonuçlarıyla RMSD ve MolProbity değerleri alındıktan sonra her modelleyicinin farklı algoritmasının ürettiği modellerin sonuçlarının ortalaması alındı. Ardından R programlama dilinde AlphaFold2, ColabFold ve RoseTTAFold2 modelleyicilerinin her biri için RMSD-MolProbity grafiği oluşturuldu. RMSD değeri en düşük olup MolProbity değeri en yüksek olan algoritma, o modelleyicinin en iyi algoritması seçildi. Ardından bu algoritmanın ürettiği 5 modelden yine en düşük RMSD ve en yüksek MolProbity değerlerine sahip olan model en iyi model olarak seçildi. D-I-

TASSER ve i-TASSER modelleyicilerinin ürettiği 5 model için de bu işlem tekrarlanarak en iyi model seçildi. Ardından tüm modellerin sonuçları karşılaştırılarak yüksek MolProbity ve düşük RMSD değerlerine sahip model ve modelleyici ALX4 için sırasıyla en uygun model ve modelleyici olarak seçildi.

3.3.6. NAMD ile Minimizasyon Yapılması

Her modelleyiciden seçilen en iyi modeller aşağıdaki adımlar takip edilerek NAMD aracı ile minimize edildi.

1. Minimize edilecek modelin PDB formatlı dosyası VMD'ye yüklendi.
2. VMD'de "Extensions → Tk Console" yolu takip edilerek Tk yazım dilinin olduğu konsol açılıp aşağıdaki komutlarla molekül merkeze getirildi ve kaydedildi.

```
set molecule [atomselect top all]
measure center $molecule
$molecule moveby {}
$molecule writepdb molecule-center.pdb
```

{ } içerisinde merkez koordinatlarının – ile çarpılmış sonucu bulunmaktadır.

3. Geçmiş oturum kapatılıp merkeze alınmış molekül VMD'ye PDB formatında yükendi.

4. VMD'de "Extensions → Tk Console" yolundan konsol açılarak aşağıdaki komutlar yazılıp merkezinin sıfırda olup olmadığı kontrol edildi ve molekülün minimum ve maksimum değeri olarak en uzak atomlarının konumu elde edildi.

```
set molecule [atomselect top all]
measure center $molecule
measure minmax $molecule
```

5. Molekül "Extensions → Modeling → Add Solvation Box" yolu takip edilerek çözücü kutusuna koyuldu. Çıkan sekmede kutu büyüklüğü molekülün 4. adımda bulunan minimum ve maksimum x y z koordinatlarının 10 fazlası olarak belirlendi. Ardından "Solvate" tuşu ile protein su dolu çözücü kutu içerisine koyuldu.

6. Su moleküllerinin de eklenmesiyle yapı çok fazla atom sayısına ulaştığından 5. adım sonunda elde edilen PDB dosyası terminalde ilgili PDB dosyasının olduğu dizinde aşağıdaki ayırma komutu çalıştırıldı. Böylece PDB dosyası her birinde 29994 satır olacak birçok PDB dosyalarına ayrılması sağlandı.

```
split -l 29994 solvate.pdb
```

Çıktılar PDB formatında olmadığından aşağıdaki komut ile her bir çıktı PDB formatına dönüştürüldü.

```
mv cikti cikti.pdb
```

7. Çıktıların rezidü numaraları ayrılmadan dolayı sıralı olmadığından Tk programlama dilinde TCL uzantılı numaralandırma betiği yazıldı. Betiği çalıştırmak için terminalde betiğin olduğu dizinde aşağıdaki komut çalıştırıldı.

```
vm txt -e renumber.tcl
```

8. Sistemin eksik atomlarının tamamlanması için Tk programlama dilinde PSF oluşturacak TCL uzantılı betik yazıldı. Betiği çalıştırmak için terminalde betiğin olduğu dizinde aşağıdaki komut çalıştırıldı.

```
vm txt -e psf.tcl
```

9. 8. adım sonunda üretilen PDB formatlı dosya VMD'ye yüklendi.

10. VMD'de "Extensions → Modeling → Add Ions" yolu takip edilerek çıkan sekmede 8. adımda üretilen PSF ve PDB formatlı dosyalar girilip *Neutralize and set NaCl concentration to 0.15 mol/L* seçildi. *Autoionize* tuşu ile sisteme hem Na⁺ ve Cl⁻ iyonları eklendi hem de sistemin elektriksel yükü nötr hale getirildi.

11. 10. adımda elde edilen PDB formatlı dosya VMD'ye yüklendi.

12. Tk Console sekmesi açılarak aşağıdaki komutlarla merkez kontrol edilip minimizasyonun yapılacağı alan seçilen yapılandırma dosyasında ilgili yerlere yazmak üzere minimum ve maksimum değerleri alındı.

```
set molecule [atomselect top all]
measure center $molecule
measure minmax $molecule
```

13. 12. adımda bulunan x y z koordinatlarıdaki minimum ve maksimum değerleri yuvarlanıp her koordinat içerisinde toplanarak sonuçlar yapılandırma dosyasında *grid size* ve *cell size* kısımlarına girildi. Minimizasyon adım sayısı ise 5000 olarak belirlendi ve CHARMM kuvvet alanı kullanıldı. Topoloji dosyaları için NAMD’de kullanılan dosyalar kullanıldı.

14. Aşağıda yazılan komutları içeren dosyanın çalıştırılması ile yapılandırma dosyası (topoloji dosyalarını da içeren) çalıştırıldı ve yapı minimize edildi.

Dosyayı çalıştırma komutu aşağıdaki gibidir.

```
nohup ./dosya_adi
```

Minimizasyon sonucu minimize yapıya ulaşmak için aşağıdaki adımlar izlendi.

1. Sonuçların olduğu dizinde aşağıdaki komut çalıştırılarak iyonize edilen yapının PSF formatı ile minimizasyon çıktısı VMD’ye yüklendi.

```
vmd ionized.psf out_md000.restart.coor
```

2. VMD’de “Save Coordinates → selected atoms: protein → Save” yolu izlenerek sistemin sadece minimize protein yapısı kaydedildi.

3.3.7. ALX4 için En İyi Modelleyicinin ve Modelin Belirlenmesi

Her modelleyiciden seçilen en iyi model minimize edildikten sonra bu modellerin RMSD ve MolProbity değerleri alındı. Sonuçların R programlama dilinde RMSD – MolProbity grafiği oluşturularak RMSD değeri en düşük olup MolProbity değeri en yüksek olan model en iyi model ve bu modeli üreten modelleyici ile parametreleri ALX4 için en iyi modelleyici ve parametreler olarak seçildi.

Ancak kullanılan modelleyicilere ve minimizasyon ile proteinin iyileştirilmesine rağmen ALX4 proteini iyi bir doğrulukla modellenemedi. En iyi modellenen bölge ise dominin bölgesi oldu. Bu nedenle tezin ikinci kısmı olan fonksiyon kaybı ve kazanımının yapısal değişikliğinin araştırılması 2M0C ve bu bölgeye denk gelecek varyantlar ile yapıldı.

3.4. 2M0C'nin Yabancı Homeodomain Olarak Hazırlanması

2M0C'nin NCBI Structure veri tabanından indirilen PDB dosyası VMD'ye yüklenerek "File → Save Coordinates" yolu takip edilip sekmede *resid 3 to 75* yazılarak kaydedildi. Böylece ilk 3 amino asidi kesildi. Ardından biopython kütüphanesi kullanılarak aşağıdaki Python komutları ile rezidü numaraları 1'den başlayıp 72'de bitecek şekilde numaralandırıldı.

```
from Bio import PDB
pdb_io = PDB.PDBIO()
pdb_parser = PDB.PDBParser()
pdb = "2M0C(NCBI-Structure)-trimmed.pdb"
model = pdb_parser.get_structure("", pdb)
resno = [i+1 for i in range(73)]
for model in model:
    for chain in model:
        for i, res in enumerate(chain.get_residues()):
            res_id = list(res.id)
            res_id[1] = resno[i]
            res.id = tuple(res_id)
pdb_io.set_structure(model)
pdb_io.save(pdb)
```

3.5. Varyant Proteinlerin Seçilmesi

Sağlıklı varyant proteinler, 25 Nisan 2023 tarihinde gnomAD veri tabanından missense ve düşük alel frekanslı hastalık yapmayan varyantlar olarak seçildi. Toplumda görülme sıklığı yüksek olup sağlıklı olarak adlandırılmış varyantların seçilmesi amaçlandı. Aynı zamanda 2M0C dizisinin 411 amino asit uzunluğundaki dizide denk geldiği 209 ile 280 amino asitleri arasında olacak şekilde filtrelendi.

İşlev kaybına ve kazanımına neden olan varyantlar, 2016 yılında yayınlanmış ve varyant sonuçları bilinen “The ALX Homeobox Gene Family and Frontonasal Dysplasias” makalesinden 209 ile 280 amino asitleri arasına denk gelen varyantlar olarak seçildi. Tüm varyantlar konumlarıyla birlikte kaydedildi.

3.6. Varyant Proteinlerin Modellenmesi

3.6.1. Dizilerin Üretilmesi

2M0C dizisi ile 411 amino asit uzunluğundaki dizinin Excel programında alta koyularak oluşturulan tablo takip edilerek varyant proteinlerin 72 amino asit uzunluğundaki dizide denk geldiği konumlar kaydedildi. Ardından her bir varyant için aşağıdaki Python komutları çalıştırıldı.

```
2m0c =
"SNKGKRRNRRTTFTSYQLEEELEKVFQKTHYPDVYAREQLAMRTDLTEARVQVWFQ
NRRAKWRKRERFGQMQQ"
varyant_adi = list(2m0c)
    varyant_adi[34] = "V" # 34 denk gelen konumu, V
ise varyant amino asidi
varyant_adi = "".join(varyant_adi)
with open("varyant_adi.txt","w") as f:
    f.write(varyant_adi)
```

3.6.2. Modelleme ve PSF Oluşturma

Varyantlar hem proteinin tümü hem de domain kısmı için en iyi modelleyici olduğu görülen şablon tabanlı ColabFold modelleyicisinde 48 geri dönüşüm sayısı ile modellendi. Parametreler tezin 3.2.4. Belirlenen Modelleyicilerde Modelleme başlığı altında gösterildiği gibi seçilip farklı olarak *template_mode* parametresi *pdb100*, *num_recycles* parametresi 48 olarak seçildi. Ardından tezin 3.2.5. Eksik Atomların Tamamlanması başlığında belirtildiği gibi elde edilen tüm modellerin VMD'nin AutoPSF aracından olası eksik atomları tamamlandı.

3.6.3. En İyi Modellerin Belirlenmesi

Her varyantın 5 modelinin pLDDT ve MolProbity deęeri alındı. pLDDT deęeri, modelleme sonucu oluřan log dosyasındaki ortalama pLDDT deęeri olarak kaydedildi. Ardından her varyant modelleri iin R programlama dilinde pLDDT – MolProbity grafięi ıkarıldı. pLDDT ve MolProbity deęeri en yksek model en iyi model olarak belirlendi.

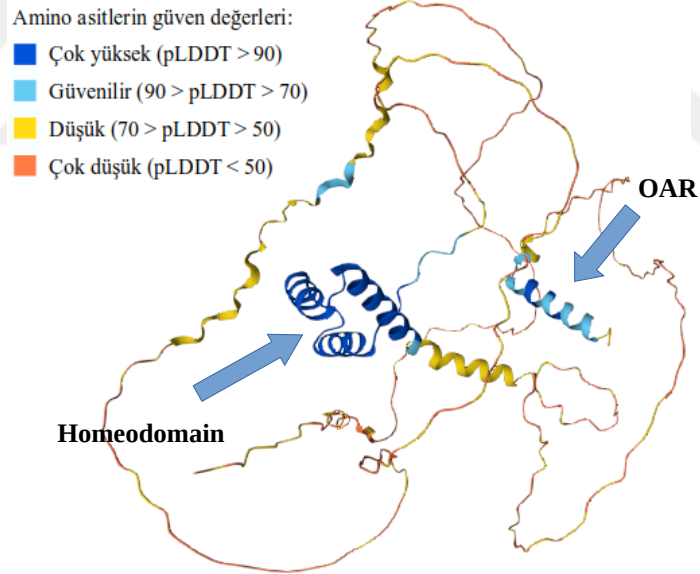
3.7. Yabarı Homeodomain ile Varyant Proteinlerin Karřılařtırılması

Yabarı tip homeodomain ile varyant proteinler nce VMD aracından RMSD hesaplanarak karřılařtırıldı. Ardından Discovery Studio aracından deęiřen amino asidin etkileřimde olduęu amino asitler ve etkileřtięi amino asitlerle yaptıęı baęlar karřılařtırılarak deęerlendirildi. Deęerlendirmeler saęlıklı, iřlev kaybı olan ve iřlev kazanımı olan varyantlar olarak  grup altında incelendi. Sonular AlphaMissense aracının sonuları ile de karřılařtırıldı.

4. BULGULAR

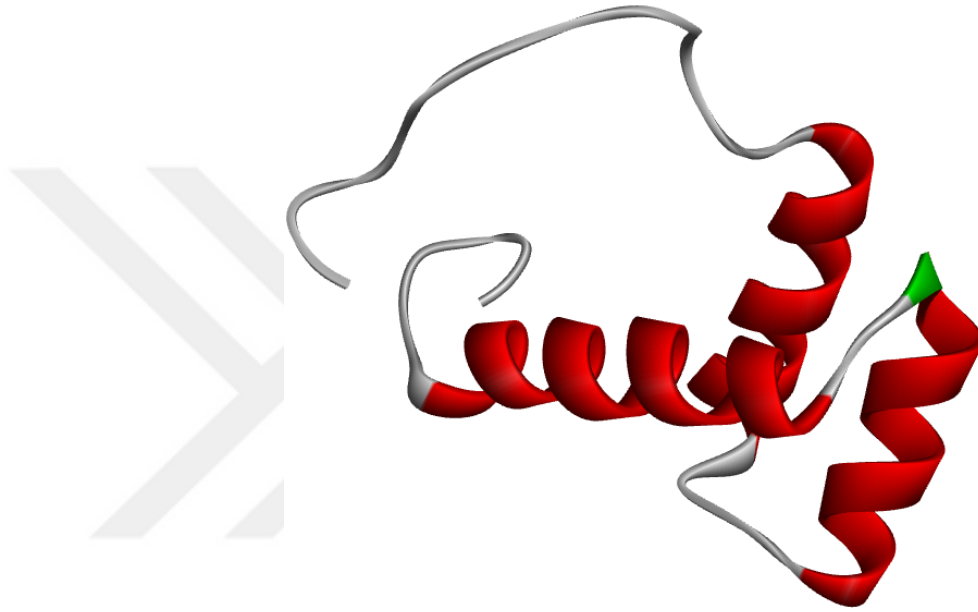
4.1. Yabancı Tip ALX4 Proteininin Üç Boyutlu Yapısı ve Dizisi

ALX4'ün UniProt veri tabanında *Homo sapiens* için üç boyutlu yapısı 411 amino asit uzunluğunda AlphaFold tarafından tahmin edilmiş şekilde Q9H161 kimliği ile bulunmaktadır (Şekil 4.1). Bu kayıt, 411 amino asit uzunluğundaki ALX4 proteininin tamamını kapsayan tek yapıdır. ALX4'ün yapısı Şekil 4.1'de görülmektedir. Yapıda özellikle homeodomain ve OAR bölgelerindeki ikincil yapının yüksek güvenilirlikle tahmin edildiği göze çarpmaktadır. Bu bölgeler dışındaki kısımların amino asit konumları güvenilir olmayıp, düzensiz (ikincil yapı içermeyen) katlanma içermektedir.



Şekil 4.1. ALX4 proteininin tahmini yapısı. Yapı UniProt veri tabanındaki AlphaFold tahmin sonucunu göstermekte olup ChimeraX programında .jason dosyası ile pLDDT değerlerine göre renklendirilmiştir. Koyu mavi renk pLDDT > 90 (çok güvenilir), açık mavi renk 90 > pLDDT > 70 (yüksek güvenilirlik), sarı renk 70 > pLDDT > 50 (düşük güvenilirlik) ve turuncu renk pLDDT < 50 (çok düşük güvenilirlik) değerlerine karşılık gelmektedir.

Buna ek olarak, UniProt veri tabanında NMR yöntemi ile çözülmüş ve PDB veri tabanında da bulunan 2M0C kimlikli, ALX4 proteininin sadece homeodomain kısmına ait ve 75 amino asit uzunluğunda bir yapı bulunmaktadır (36). Şekil 4.2’de 2M0C’nin Discovery Studio programında ikincil yapısına göre renklendirilmiş görüntüsü gösterilmektedir.



Şekil 4.2. ALX4 proteini homeodomain bölgesinin NMR yapısı. PDB veri tabanında 2M0C kimlikli olarak bulunan protein ikincil yapılarına göre renklendirilmiştir. Kırmızı renk alfa sarmal, yeşil renk beta pileli ve gri renk ise düzensiz bölgeleri temsil etmektedir.

2M0C, yabancı ALX4’ün tamamında 206.-280. amino asitler arasında denk gelmektedir. Ancak ilk üç amino asidinin yabancı ALX4 ile uyumlu olmadığı, kalan 72 amino asidin uyumlu olduğu görülmüştür. Uyumlu bölge yabancı ALX4’ün tamamında 209.-280. amino asitler arasında denk gelmektedir. ALX4 proteininin tahmini yapısı (Q9H161) ile homeodomain bölgesinin NMR yapısı (2M0C) Şekil 4.3’te üst üste çakıştırılarak gösterilmiştir.

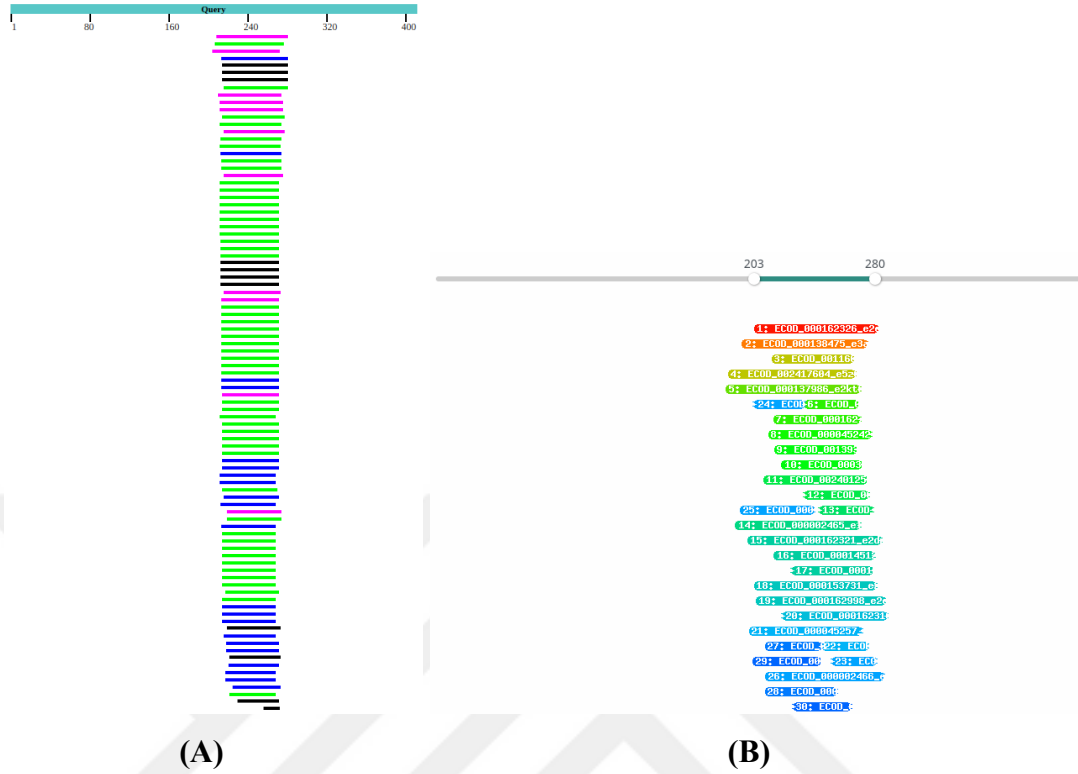

```
>2M0C_1|Chain A|Homeobox protein aristaless-like 4|Homo sapiens (9606)  
SHMSNKGKKRRNRRTTFTSYQLEELEKVFQKTHYPDVYAREQLAMRTDLTEARVQVWFQNRRA  
KWRKRERFGQMQQ
```

4.2. Yabanıl Tip ALX4 Üç Boyutlu Yapısının Modellenmesi

Yapılan değerlendirmelerde ALX4'ün tamamı için deneysel sonuç olmadığı ve AlphaFold tahmini düşük güvenilirlikli olduğundan mevcut AlphaFold modelinin kullanılması yerine yeniden ALX4 modellemesi yapılmasına karar verildi. ALX4 protein dizisinin eldesi ardından farklı algoritmalarla bu diziye göre üç boyutlu yapılar modellenmiş, ardından da modeller birbiri arasında karşılaştırılarak gerçeğe en yakın olduğu düşünülen model en iyi model seçilmiştir.

4.2.1. Homolog Protein

ALX4 dizisi için BLASTP arama motoru ile PDB veri tabanında homoloji arama ve pLM-BLAST arama motoru ile uzak homoloji arama sonuçları Şekil 4.4'te gösterilmektedir. Sonuçlarda homeodomain bölgesi dışında üç boyutlu yapısı bilinen bir proteine veya uzak homolog olup yabanıl ALX4'e homeodomain bölgesi dışında benzeyen başka bir proteine rastlanmamıştır.



Şekil 4.4. ALX4 için homolog arama sonuçları. (A) BLASTP arama motoru ile PDB veri tabanında ALX4 dizisi için elde edilen sonuçlar. (B) pLM-BLAST arama motorunda ALX4 dizisi için elde edilen sonuçlar. Her iki şekilde de yatay çizgiler homoloji gösteren protein dizilerine ait bölgeleri göstermektedir.

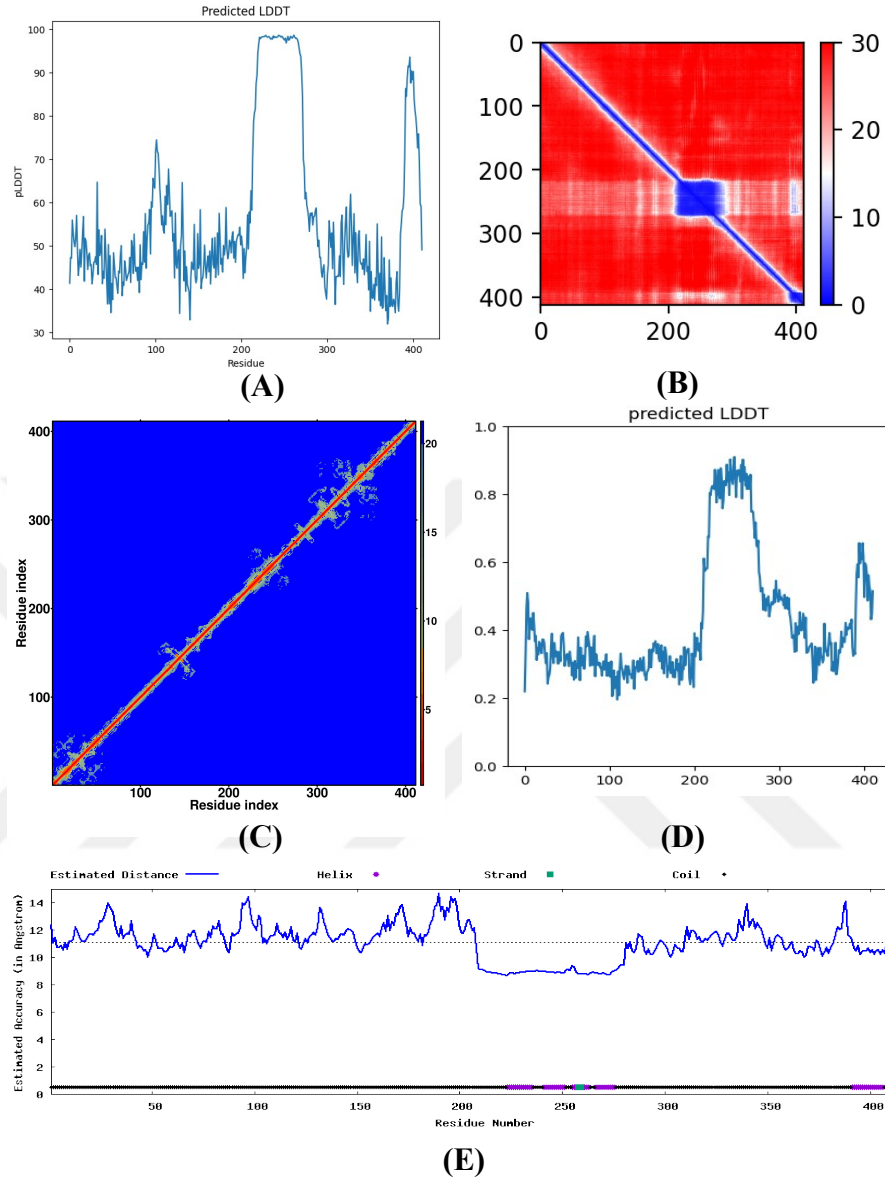
4.2.2. Yabancı ALX4 Modellerinin Karşılaştırılması ve Minimizasyon

Modelleme için son 6 yılın CASP sonuçlarında toplam Z değerleri yüksek AlphaFold2, i-TASSER ve RoseTTAFold2 modelleyicileri seçildi. Nadiren kullanılan ve Tablo 2.2’de görülen tFold ve Multicom araçları önerilen koşullar kullanılarak çalıştırılmadığı için kullanılmadı. Bu araçlara ek olarak, şablon tabansız modelleme yapan D-I-TASSER aracı, ALX4’ün tümü ile homoloji gösteren ve yapısı bilinen bir protein bulunmadığından eklendi. Her bir modelleyici için farklı parametreler arasından en iyisini seçmek amacıyla birden fazla parametre ile modelleme yapıldı.

Her modelleyicinin ürettiği modellerin eksik atomları tamamlandıktan sonra MolProbity ve 2M0C’nin 72 amino asit uzunluğundaki bölgesi ile yapısal olarak

çakıştırılıp RMSD hesaplandı. Her modelleyicide farklı parametrelere dayanan algoritmalarla üretilen 5'er modelin MolProbity ve RMSD değerlerinin ortalaması alınmış ve RMSD-MolProbity grafikleri Şekil 4.5'te sunulmuştur. Ortalama MolProbity değeri yüksek olup RMSD değeri düşük olan parametre kombinasyonu her bir modelleyici için en iyi algoritma seçilmiştir. Seçilen algoritmaların en iyi modeli ise tekrar aynı kriterlere göre seçilmiştir. En iyi modellerin doğruluk grafikleri Şekil 4.6'da gösterilmektedir. Model başarısını yansıtan ayrıntılı MolProbity değerleri RMSD değerleri ile birlikte Tablo 4.1'de sunulmaktadır. Doğruluk grafiklerinden görüldüğü üzere homeodomain dışındaki doğruluk değerleri oldukça düşüktür. Bununla birlikte izinsiz bölgede ve kötü açıda bulunan amino asit sayısı göz ardı edilemeyecek kadardır. MolProbity ve RMSD değerleri de oldukça yüksektir. Modelleme ve 2MOC ile çakıştırma sonuçları Şekil 4.7'de gösterilmektedir. Modelleri iyileştirmede minimizasyonun etkisini test etmek için her modelleyicinin en iyi sonucu ile 5000 adımlı ve CHARMM kuvvet alanlı minimizasyon işlemi yapılmıştır. Hesaplamalar sonucu elde edilen üç boyutlu protein yapıları ve 2MOC ile çakıştırma sonuçları Şekil 4.8'de verilmiştir. Minimize modellerin RMSD ve ayrıntılı MolProbity değerleri Tablo 4.1'in sonunda sunulmaktadır. Minimizasyon öncesi sonuçlar ile kıyaslandığında, minimizasyon işlemi ile MolProbity değerlerinin düşüp RMSD değerini arttığı görülmüştür. Ayrıca, minimizasyon sonucunda homeodomain kısmında da deneysel model ile karşılaştırıldığında ikincil yapılarda belirgin bozulmalar saptanmıştır. Bu sonuçlardan yola çıkarak, minimizasyon işleminin ALX4 modellerini kötüleştirdiği sonucuna varılmış ve sonraki işlemlerde minimizasyon kullanılmamasına karar verilmiştir.

Tüm sonuçlar değerlendirildiğinde, en düşük RMSD ve en yüksek MolProbity değerlerine sahip olan şablon tabanlı ColabFold (48 geri dönüşümlü) en başarılı modelleyici olarak ortaya çıkmaktadır. Ancak en başarılı modelde dahi homeodomain dışındaki modellemenin düzensiz ve ikincil yapılardan fakir olduğu görülmektedir. Bu nedenle, homeodomain dışında iyi bir modelleme yapılamadığı sonucu ortaya çıkmaktadır.

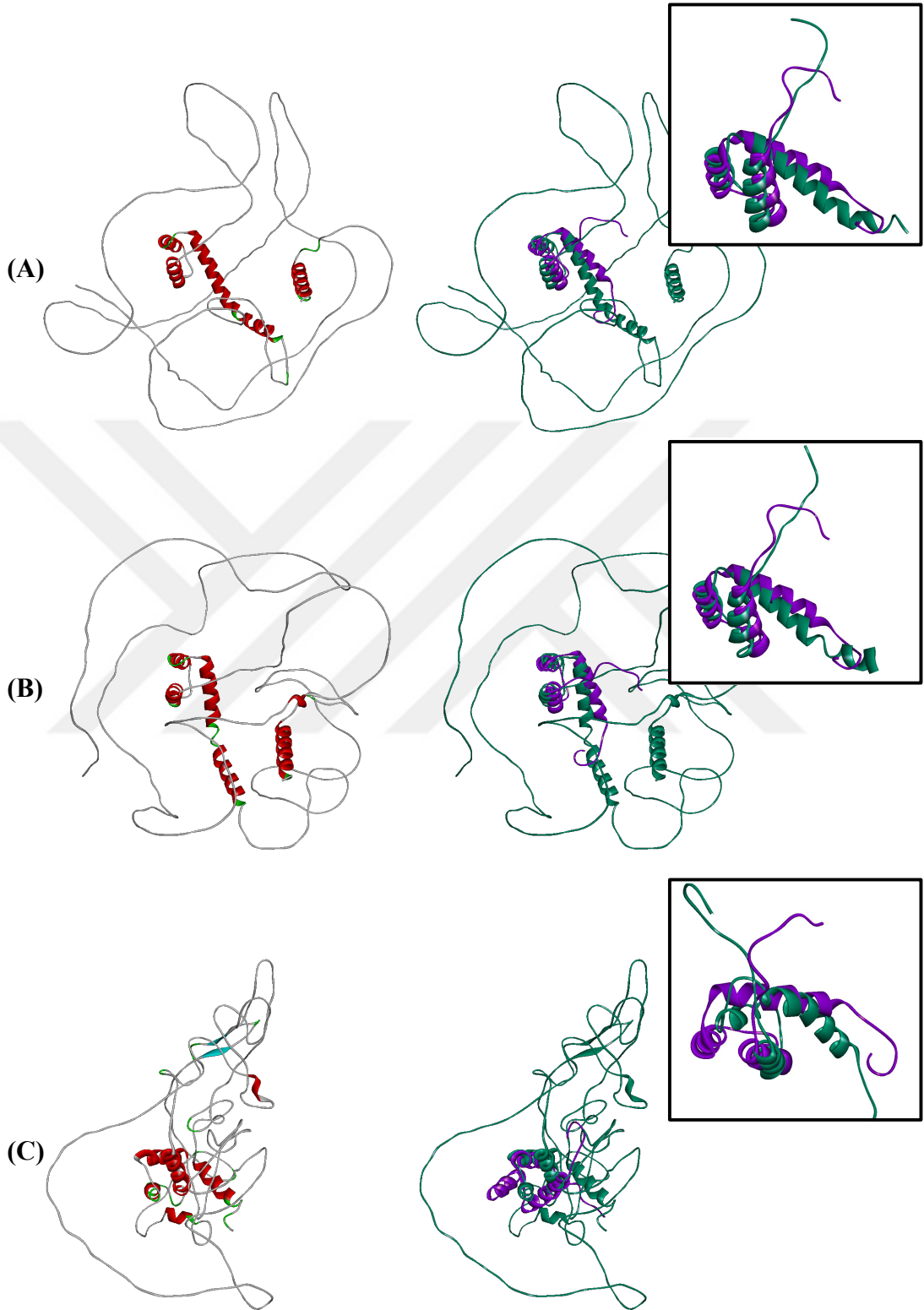


Şekil 4.6. Modelleyicilerde üretilen en iyi modellerin doğruluk grafikleri. (A) sekmesinde AlphaFold2’de üretilen en iyi modelin pLDDT grafiği sunulmaktadır. Yatay ekseninde amino asit numaraları, düşey ekseninde pLDDT değerleri gösterilmektedir. pLDDT değeri 70’in üzerinde olan amino asitlerin konumları güvenilir kabul edilmektedir. ALX4 homeodomini dışındaki yerler yüksek güvenilirlikle modellenememiştir. (B) sekmesinde ColabFold’da üretilen en iyi modelin PAE grafiği gösterilmektedir. Hem yatay hem de düşey ekseninde amino asit numaraları bulunmaktadır. Amino asitlerin birbiri ile uzaklıkları maviden kırmızıya doğru artmaktadır. En yakın amino asitler homeodomain bölgesine denk gelmekte olup kalan amino asitler kompakt bir yapıya sahip olmak için birbirinden oldukça

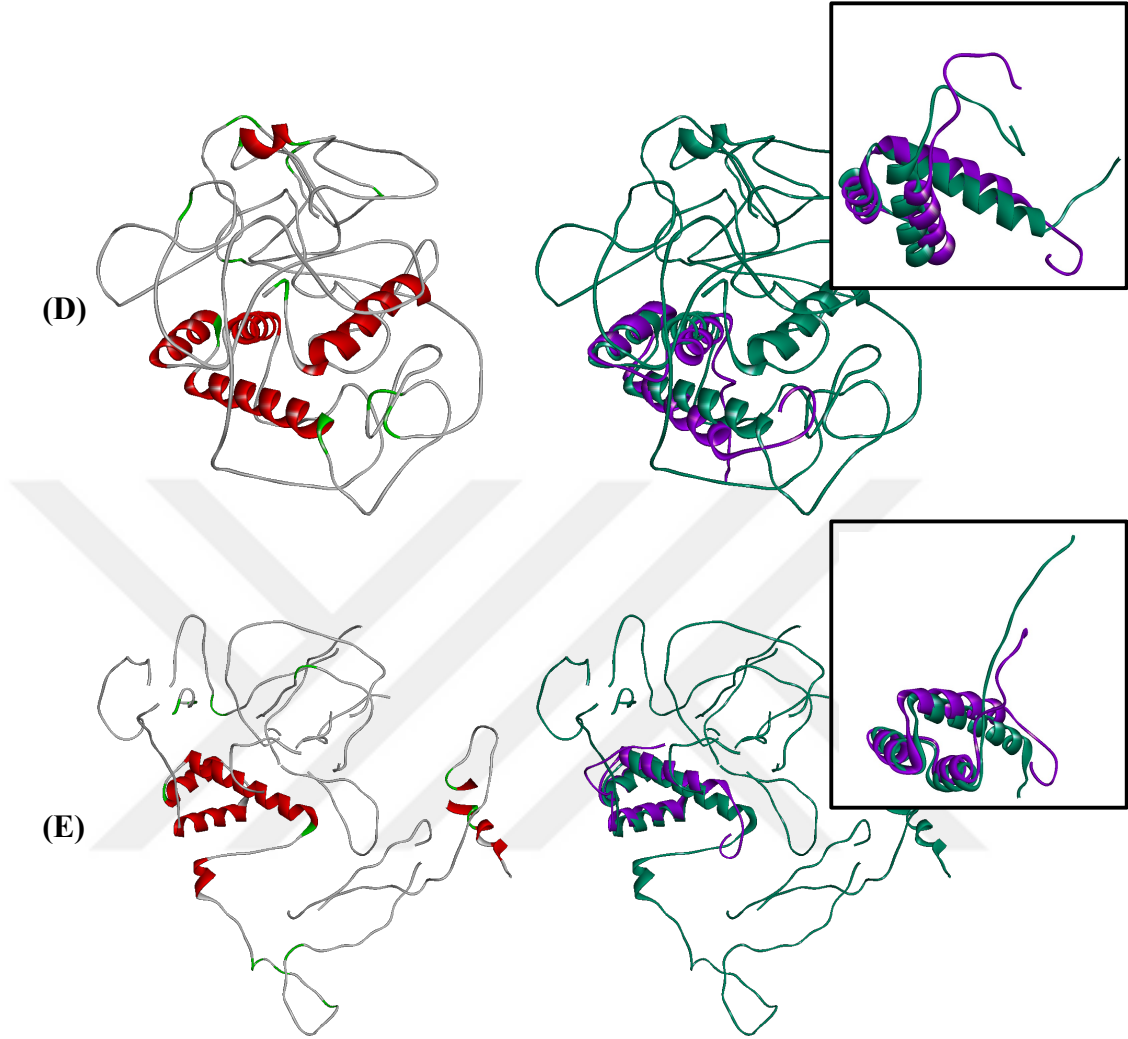
uzaktır. (C) sekmesinde D-I-TASSER'da üretilen modelin PAE grafiđi gösterilmektedir. Kırmızıdan maviye amino asitlerin uzaklıđı artmaktadır. (D) sekmesinde RoseTTAFold2'nin pLDDT grafiđi görölmektedir. Grafikte yatay eksen amino asit numaraları, dişey eksende ondalık ölçekte pLDDT deđerleri bulunmaktadır. 0,7 ve üzeri pLDDT deđerine sadece homeodomain bölgesindeki amino asitler sahiptir. (E) sekmesinde i-TASSER'ın dođruluk grafiđi sunulmaktadır. Grafikte yatay eksen amino asit numaralarını, dişey eksen şablonların amino asitleri ile modelin amino asitleri arasındaki uzaklıđı angstrom cinsinden belirtmektedir. Grafiđin en altında ise oluşturulan modelin ikincil yapısı özetlenmektedir. Yeşil kısım alfa sarmalı, pembe kısım sarmallar arasındaki zinciri ve gri kısım düzensiz katlanmaları temsil etmektedir. İkincil yapıya homeodomain ve OAR domain ulaşmış olup, şablona en çok homeodomain bölgesi benzemektedir.

Tablo 4.1. En iyi modeller ve bu modellerin minimize yapılarının RMSD ve ayrıntılı MolProbability değerleri. En iyi modeller ve bu modellerin ayrıntılı MolProbability değerleri ve 2M0C ile örtüşmeyi gösteren RMSD değerleri modelin algoritma parametreleri ile sunulmuştur.

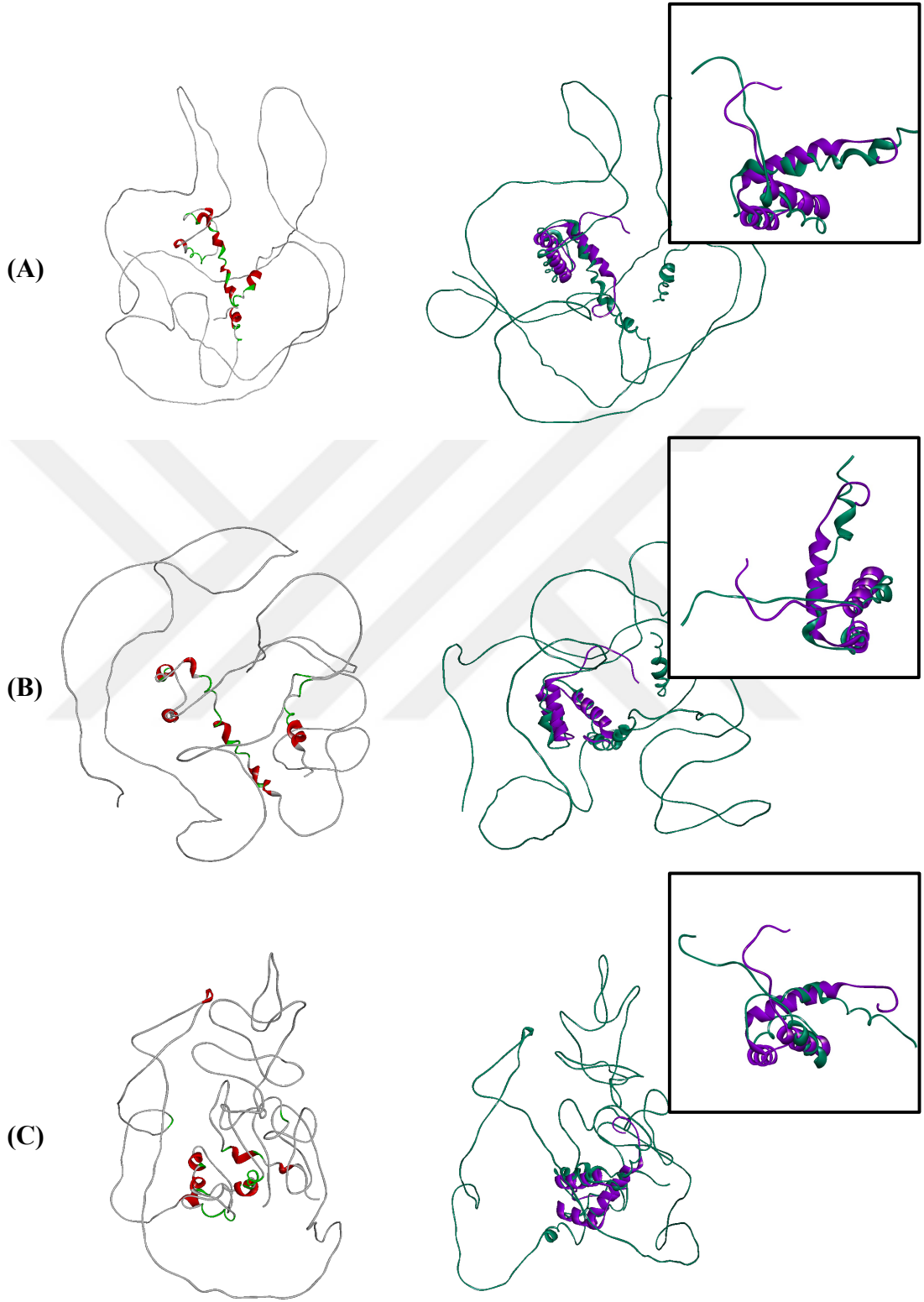
Modelleyici	Algoritma Parametreleri	MolProbability	Clash Puam	Zayıf Rotamer Sayısı (%)	Tercih Edilir Rotamer Sayısı (%)	İzinsiz Bölgedeki AA Sayısı (%)	İzinli Bölgedeki AA Sayısı (%)	Sapma > 0,25Å Olan C β Sayısı (%)	Kötü Bağ Sayısı (%)	Kötü Açık Sayısı (%)	RMSD
AlphaFold2	3 GD	2,02	0,19	9,44	78,31	20,11	63,76	5,87	0,06	2,63	5,23
ColabFold	Tabanlı, 48 GD	2,24	0,39	8,12	80,94	14,88	61,62	2,12	0,06	3,35	5,08
D-I-TASSER	-	3,21	67,59	0,31	99,69	12,53	60,31	0,00	0,29	0,53	8,47
i-TASSER	-	3,02	3,83	14,42	73,35	17,80	51,57	7,96	0,10	1,46	7,30
RoseTTAFold2	Tabanlı, Yeni, 512 MSA, 24 GD	4,35	68,96	27,50	51,25	13,71	54,57	80,16	10,72	34,60	5,39
AlphaFold2	3 GD	1,94	0,00	5,94	84,38	14,36	68,93	6,88	0,00	1,77	5,52
ColabFold	Tabanlı, 48 GD	1,93	0,00	5,31	86,56	15,93	64,75	9,52	0,00	2,61	5,46
D-I-TASSER	-	1,82	0,00	4,06	91,25	12,79	68,41	3,44	0,00	1,46	8,94
i-TASSER	-	2,22	0,00	11,60	76,80	11,26	61,26	8,49	0,00	1,27	7,64
RoseTTAFold2	Tabanlı, Yeni, 512 MSA, 24 GD	1,91	0,00	5,62	85,31	9,14	69,71	3,44	0,00	1,15	5,78



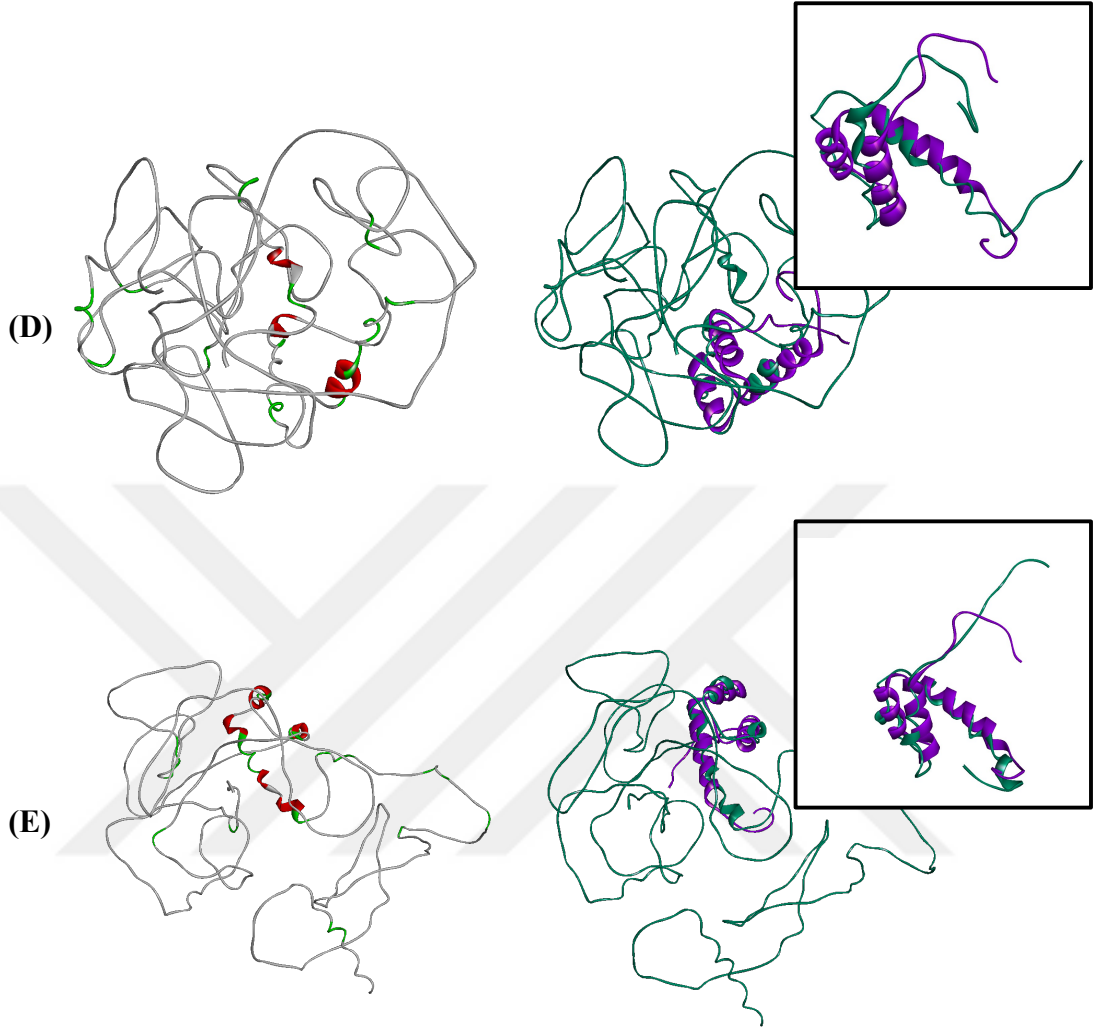
Şekil 4.7. Modelleme sonuçları.



Şekil 4.7. (devamı) Modelleme sonuçları. (A) sekmesinde AlphaFold2'nin, (B) sekmesinde ColabFold'un, (C) sekmesinde D-I-TASSER'in, (D) sekmesinde i-TASSER'in, (E) sekmesinde RoseTTAFold2'nin ürettiği en iyi model sunulmaktadır. Solda model ikincil yapısına göre renklendirilmiş, sağda ise 2M0C ile karşılaştırılmış görüntüsü gösterilmektedir. Sağdaki görüntü yeşil model, mor 2M0C olacak şekilde renklendirilmiştir. Kutu içerisinde karşılaştırmanın yakın görüntüsü gösterilmektedir.



Şekil 4.8. Minimizasyon sonuçları.



Şekil 4.8. (devamı) Minimizasyon sonuçları. (A) sekmesinde AlpaFold2'nin, (B) sekmesinde ColabFold'un, (C) sekmesinde D-I-TASSER'in, (D) sekmesinde i-TASSER'in, (E) sekmesinde RoseTTAFold2'nin ürettiği en iyi modele uygulanan minimizasyon sonucunda elde edilen model sunulmaktadır. Solda model ikincil yapısına göre renklendirilmiş, sağda ise 2M0C ile karşılaştırılmış görüntüsü gösterilmektedir. Sağdaki görüntü yeşil model, mor 2M0C olacak şekilde renklendirilmiştir. Kutu içerisinde karşılaştırmanın yakın görüntüsü gösterilmektedir.

4.3. ALX4 Homeodomain Varyant Modelleri

Yabanıl ALX4 modellerinin sadece homeodomain için başarılı modelleme yapması nedeniyle tezin bundan sonraki kısmında homeodomain kısmına denk gelen 72 amino asit uzunluğundaki (ALX4 proteininin 209.-280. amino asitler arası) kısım varyant modellemelerinde kullanılmıştır. Tüm karşılaştırmalarda bu amino asitlere denk gelen 2M0C'nin son 72 amino asit uzunluğundaki kısmı kullanılmış olup karşılaştırılacak varyantlar da ALX4'ün bu kısmından seçilmiştir.

4.3.1. Modelleme için Kullanılacak Varyant Proteinler ve Dizileri

Varyant bilgilerine daha önceden bilinen ALX4 literatüründeki patojenik varyantlar, ClinVar verileri ve gnomAD 4.1 veri tabanından ulaşılmıştır. Bu kaynaklarda yer alan varyantlardan homeodomain (214-273) ve deneysel (206-280) bölgeye denk gelen ve *missense* olan 66 varyanta ulaşılmıştır (37). Bu varyantlardan 63 tanesi gnomAD veri tabanında bulunmakta olup sadece bir tanesinin sıklığı yüksek ($>10^{-5}$) ve ClinVar'da muhtemel benign varyant olarak bildirilmiştir (p.Ala243Val). Bunun dışında, p.Gln225Glu, p.Arg218Gln ve p.Arg272Pro olmak üzere 3 varyant daha önceden patojenik olarak işlev kaybettiren ALX4 ilişkili hastalıklarda bildirilmiştir (16). Son olarak, homeodomain bölgesinin içerisine denk gelmese de çok yakınında olan p.Lys211Glu varyantı işlev kazandıran bir varyant olarak bildirilmiş olup bu tezde modellenmektedir (16). Varyant dizileri aşağıda değişmiş amino asit kalın işaretlenmiş olarak verilmektedir.

> Ala243Val

SNKGKKRRNRRTTFTSYQLEEELEKVFQKTHYPDVY**V**REQLAMRTDLTEARVQVWFQN
RRAKWRKRERFGMQQ

> Gln225Glu

SNKGKKRRNRRTTFTSY**E**LEEELEKVFQKTHYPDVYAREQLAMRTDLTEARVQVWFQN
RRAKWRKRERFGMQQ

> Arg218Gln

SNKGKKRRN**Q**TTFTSYQLEEELEKVFQKTHYPDVYAREQLAMRTDLTEARVQVWFQN
RRAKWRKRERFGQM**Q**Q

> Arg272Pro

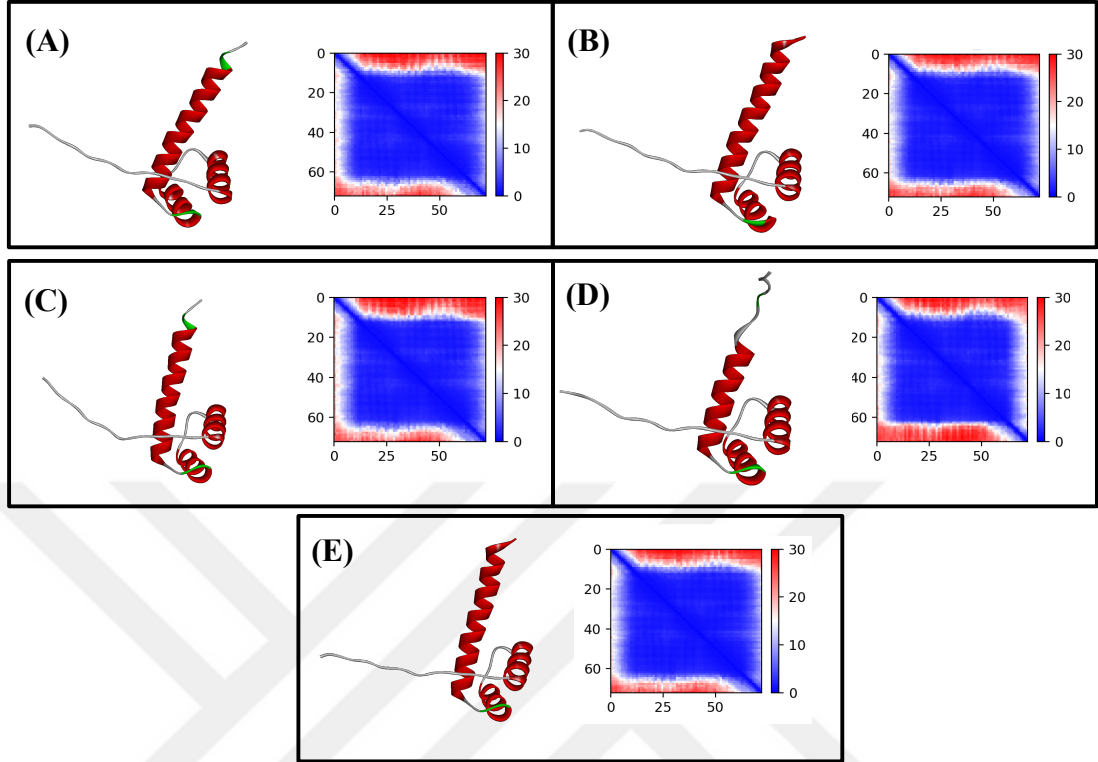
SNKGKKRRN**R**TTFTSYQLEEELEKVFQKTHYPDVYAREQLAMRTDLTEARVQVWFQN
RRAKWRK**P**ERFGQM**Q**Q

> Lys211Glu

SN**E**GKKRRN**R**TTFTSYQLEEELEKVFQKTHYPDVYAREQLAMRTDLTEARVQVWFQN
RRAKWRKRERFGQM**Q**Q

4.3.2. Varyant Protein Modelleri

Homeodomain bölgesi için ColabFold'un ürettiği 5 modelden MolProbitiy ile pLDDT değerlerine göre en iyisi seçilen varyantların modelleme sonuçları ve PAE grafiklerine bakıldığında varyantların güvenilir şekilde modellenebildiği görülmüştür (Şekil 4.9). Değerlerin ayrıntılı tablosu Tablo 4.2'de gösterilmektedir.



Şekil 4.9

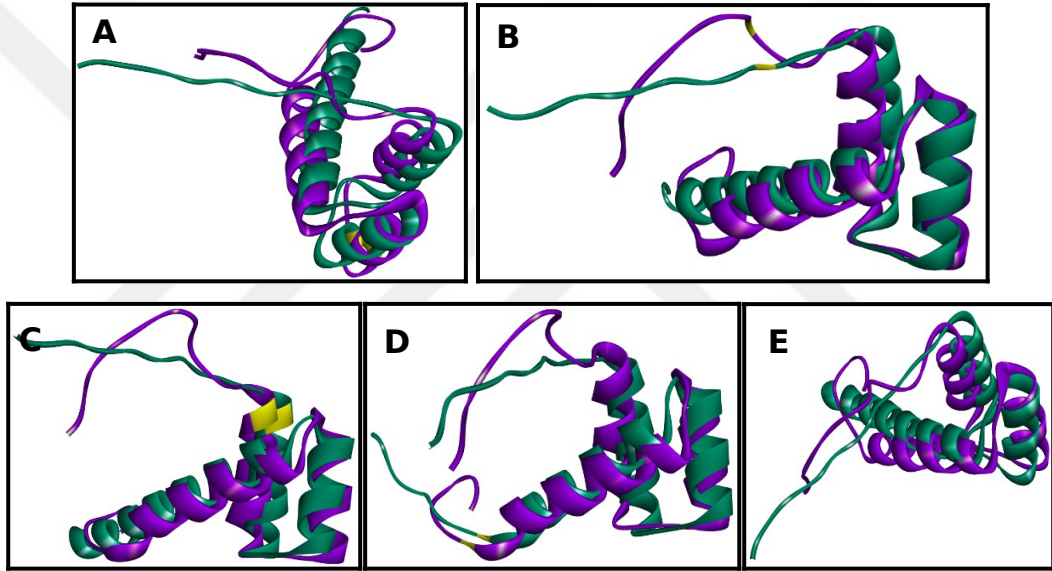
Varyant modelleri ve PAE grafikleri. (A) sekmesinde Ala243Val, (B) sekmesinde Lys211Glu, (C) sekmesinde Arg218Gln, (D) sekmesinde Arg272Pro, (E) sekmesinde Gln225Glu varyantları sağda Discovery Studio aracında ikincil yapılarına göre renklendirilmiş olarak, solda PAE grafikleri ile gösterilmektedir. Grafikte model içerisindeki amino asitler arası uzaklık maviden kırmızıya doğru artmaktadır.

Tablo 4.2. Varyantların pLDDT ve detaylı MolProbability değerleri

Varyant	MolProbability	Clash Puanı	Zayıf Rotamer Sayısı (%)	Tercih Edilir Rotamer Sayısı (%)	İzinsiz Bölgedeki AA Sayısı (%)	İzinli Bölgedeki AA Sayısı (%)	Sapma > 0.25Å Olan C _β Sayısı (%)	Kötü Bağ Sayısı (%)	Kötü Açı Sayısı (%)	pLDDT
Ala243Val	0,97	0,00	0,00	98,48	1,49	92,54	0,00	0,31	0,83	87,4
Lys211Glu	1,59	2,80	0,00	98,46	1,49	91,04	0,00	0,32	0,71	88,2
Arg218Gln	1,02	0,00	0,00	98,46	1,49	91,04	0,00	0,32	0,83	86,2
Arg272Pro	1,13	0,00	0,00	98,46	4,48	86,57	1,43	0,32	1,19	85,6
Gln225Glu	1,59	2,80	0,00	100,00	1,49	91,04	0,00	0,32	0,47	88,7

4.4. Yabanıl Homeodomain ile Varyantların Karşılaştırılması

Yabanıl homeodomain ile 5 varyantın karşılaştırılma sonucu her varyant için Şekil 4.10'de verilmektedir. Burada yeşil renk varyantı, mor renk yabanıl tipi temsil etmektedir. Tüm karşılaştırmalarda homeodomain bölgeleri için yabanıl tip ile mutant modellerin benzer RMSD değerlerine (3,90-4,64 Å) sahip olduğu ve bu tür bir karşılaştırma ile farklı varyantlar arasında ayırım yapmanın mümkün olmadığı görülmektedir (Tablo 4.3).



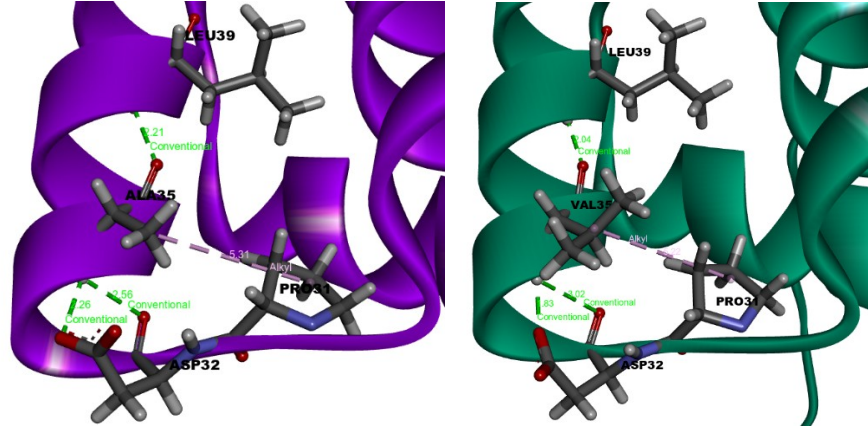
Şekil 4.10. ALX4 protein varyant modellerinin homeodomain bölgesi için yabanıl ALX4 ile üst üste karşılaştırılması. Tüm şekillerde mor modeller yabanıl ALX4'ü yeşil modeller ise varyantlar için oluşturulan modelleri göstermektedir. Değişen amino asit ise tüm modeller üzerinde sarı ile işaretlidir. (A) p.Ala243Val, (B) p.Arg218Gln, (C) p.Gln225Glu, (D) p.Arg272Pro, (E) p.Lys211Glu.

Her bir varyantın amino asitler arası etkileşim üzerine etkileri incelendiğinde ise *missense* varyant dışındaki diğer varyantlarda belirgin farklılıklar görülmektedir. *Benign* p.Ala243Val varyantı homeodomain içerisinde bulunmaktadır. Bu varyant hidrofobik bir amino asidin başka bir hidrofobik amino aside değişimine sebep olmasına rağmen AlphaMissense aracı tarafından patojenik bir etkiye sahip olduğu tahmin edilmiştir. Bu tezde yapılan incelemeye göre ise bu *missense* değişiklik

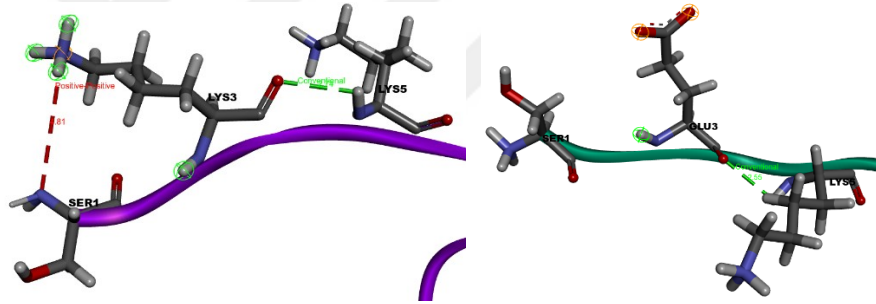
homeodomain içerisinde amino asit interaksiyonlarında belirgin bir değişikliğe sebep olmamaktadır (Şekil 4.11). Öte yanda, işlev kazandıran p.Lys211Glu değişikliği yabancı proteindeki Lys211 ile Ser209 arasında bulunan etkileşimi bozmaktadır (Şekil 4.12). Ser209 homeodomain dışında bulunan bir amino asit olduğu için bu değişimin homeodomain dışındaki bir interaksiyon üzerine etki ettiği söylenebilir. İşlev kaybettiren p.Gln225Glu ve p.Arg272Pro değişiklikleri homeodomain içerisinde yer almakta ve homeodomain içerisindeki diğer amino asitlerle etkileşime girmektedir. AlphaMissense bu değişiklikleri patojenik etkiye sahip olarak tahmin etmiştir. Bu değişikliklerin homeodomain içerisindeki diğer amino asitlerle etkileşimde bozulmalara sebep olduğu görülmektedir (Şekil 4.13). İşlev kaybettiren bir diğer değişiklik olan p.Arg218Gln için ise etkileşim üzerine nasıl bir etkisi olduğu konusunda tahmin yürütmek mümkün olmamıştır (Şekil 4.14).

Tablo 4.3. Modellenen ALX4 varyantlarının özellikleri.

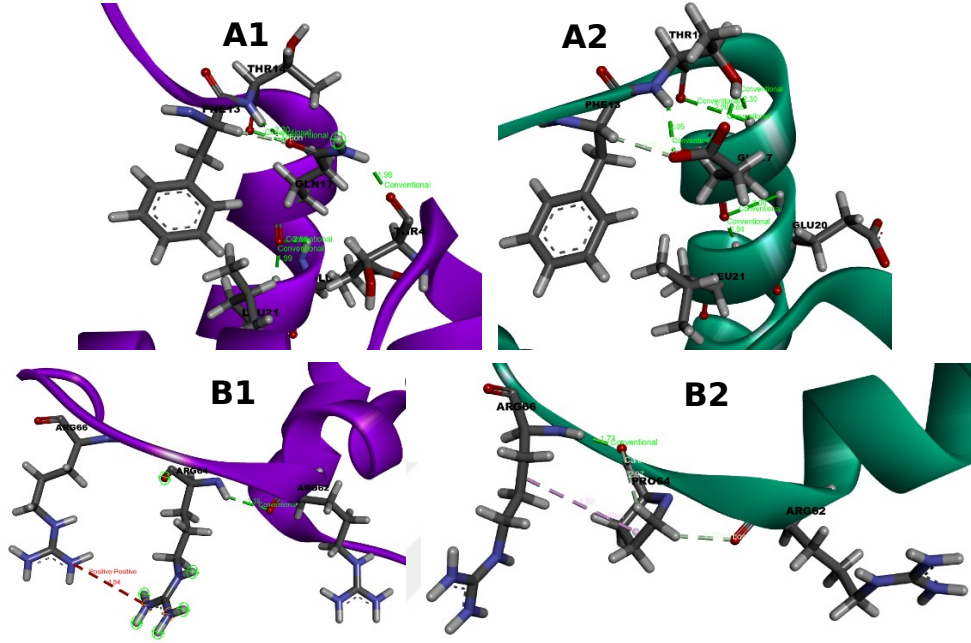
Varyant	Bilinen Etkisi	RMSD (Å)	Farklı Etkileşim	Tahmin Edilen Etkisi	AlphaMissense
p.Ala243Val	<i>Benign</i>	4,42	Yok	<i>Benign</i>	Patojenik
p.Arg218Gln	<i>LoF</i>	3,90	?	Patojenik	Patojenik
p.Gln225Glu	<i>LoF</i>	4,44	Homeodomain içi	Patojenik	Patojenik
p.Arg272Pro	<i>LoF</i>	4,64	Homeodomain içi	Patojenik	Patojenik
p.Lys211Glu	<i>GoF</i>	4,35	Homeodomain dışı	Patojenik	Patojenik



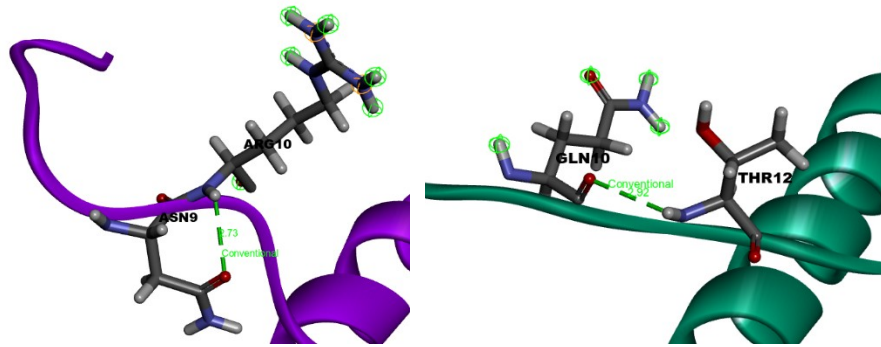
Şekil 4.11. p.Ala243Val varyantının etkileşimleri. Yabani homeodomainde (mor) Ala243 (homeodomain parçasında Ala35) ile Pro239 (Pro31) arasında bulunan etkileşim p.Ala243Val varyantında (yeşil) korunmaktadır.



Şekil 4.12. p.Lys211Glu varyantının etkileşimleri. Yabani homeodomainde (mor) Lys211 (homeodomain parçasında Lys3) ile Ser209 (Ser1) arasında bulunan etkileşim p.Lys211Glu varyantında (yeşil) kaybolmuştur. Ser209 homeodomain dışında yer almaktadır.



Şekil 4.13. p.Gln225Glu ve p.Arg272Pro varyantlarının etkileşimleri. Yabancıl ALX4 proteininde (mor, A1) Gln225 (homeodomain parçasında Gln17) ile Ser251 (Thr43) arasında bulunan etkileşim p.Gln225Glu varyantında (yeşil, A1) kaybolmuştur. Yabancıl homeodomainde (mor, B1) Arg272 (Arg64) ile Arg270 (Arg62) arasında bulunan etkileşim p.Arg272Pro varyantında (yeşil, B2) kaybolmuştur. Ser251 ve Arg270 homeodomain içerisinde yer almaktadır.



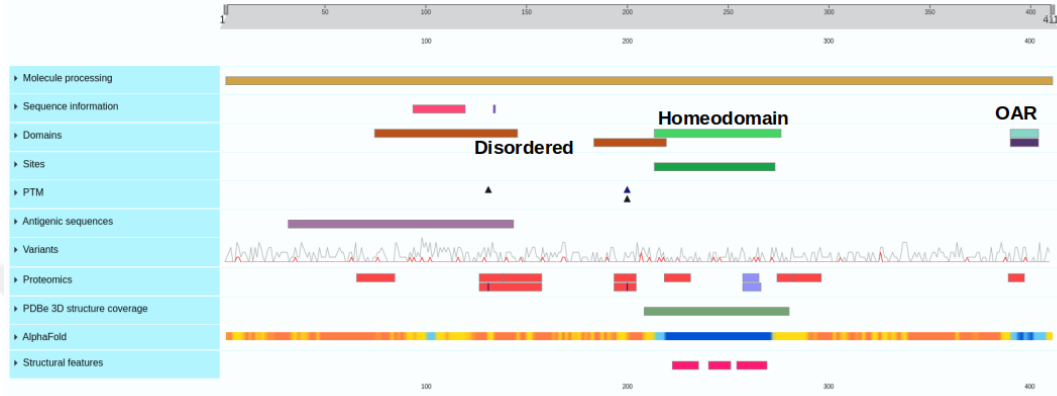
Şekil 4.14. p.Arg218Gln varyantının etkileşimleri. Yabancıl homeodomainde (mor) Arg218 (homeodomain parçasında Arg10) amino asidine ait yan zincirin belirgin bir etkileşimi bulunmamış olup p.Arg218Gln varyantının (yeşil) yan zinciri için de belirgin bir etkileşim saptanmamıştır.

5. TARTIŞMA

Homeobox genlerinin bir sınıfını oluşturan *ALX* gen ailesi, insanlarda *ALX1*, *ALX3* ve *ALX4* olmak üzere 3 gen içermektedir. Bu genlerin tamamı özellikle yüzün embriyonik gelişiminde rol oynayan transkripsiyon faktörlerini kodlamaktadırlar. Bu genlerin işlev kaybettiren mutasyonlarında farklı şiddetlerde Frontonazal Displazi adı verilen ve embriyonik orta yüz ve burun gelişiminin tam tamamlanamadığı hastalıklar ortaya çıkmaktadır. Bu 3 gen arasında *ALX4* işlev kazandıran mutasyonlarının da kraniyosinostoz adı verilen bir hastalık ile ilişkilendirilmesi ile diğerlerinden ayrılmaktadır. *ALX4* geninin bu şekilde hem işlev kaybının hem de kazancının fenotipik açıdan birbirine zıt görünen hastalıklarla ilişkili olması kalıtsal hastalıklar arasında oldukça nadir bir özelliktir. Bu tezde, *ALX4* genindeki farklı genetik değişikliklerin etkisinin *ALX4* yapısal protein modelleri üzerinde farklı etkileri olabileceği öngörülerek bu etkiler ile protein işlevinin ilişkilendirilmesi amaçlanmıştır. Elde edilen veriler ışığında *ALX4* proteininin tamamı için sağlıklı bir modelleme yapılamamış olsa da DNA bağlanmasından sorumlu homeodomain için farklı varyantlara yönelik modeller yapılmış, özellikle işlev kaybettiren *ALX4* varyantları için anlamlı değişiklikler ortaya çıkmıştır.

ALX4 proteininin tamamının modellenmesi için günümüzde yaygın kullanılan 6 farklı araçla elde edilen modeller, proteinin homeodomain ve OAR domain bölgeleri dışında başarılı olamamış, bu bölgeler dışında yüksek pLDDT değerine sahip amino asitler elde edilememiştir. Bu durumun en önemli nedeninin *ALX4* ile homoloji proteinler için sadece bu 2 bölgeye yönelik deneysel protein yapılarının bulunmasıdır. Bu durumun altında *ALX4* transkripsiyon faktörünün çok sayıda amino asit tekrarları içeren bölgeler içermesi yatmalıdır. Bu tür bölgeler, intrinsik düzensiz bölgeleri (IDR) işaret etmekte olup UniProt veri tabanına göre de *ALX4* proteini 75.-145. amino asitler ile 184.-219. amino asitler arasında olmak üzere toplam 107 amino asit uzunluğunda bir IDR içermektedir (Şekil 5.1). Bu bölge tüm proteinin dörtte birlik kısmını oluşturmakta olup bu türden bölgelerin oldukça dinamik yapılar olduğu ve sağlıklı modellerinin oluşturulamadıkları bilinmektedir

(38). Ayrıca bu bölgelerin transkripsiyon faktörlerinde zengin oldukları da bilinmektedir (39). Bu nedenle çalışmada kullanılan protein modelleme yöntemlerinin ve minimizasyon işleminin başarıya ulaşmadığı düşünülmüştür.



Şekil 5.1. ALX4 için UniProt veri tabanından elde edilen protein domainleri.

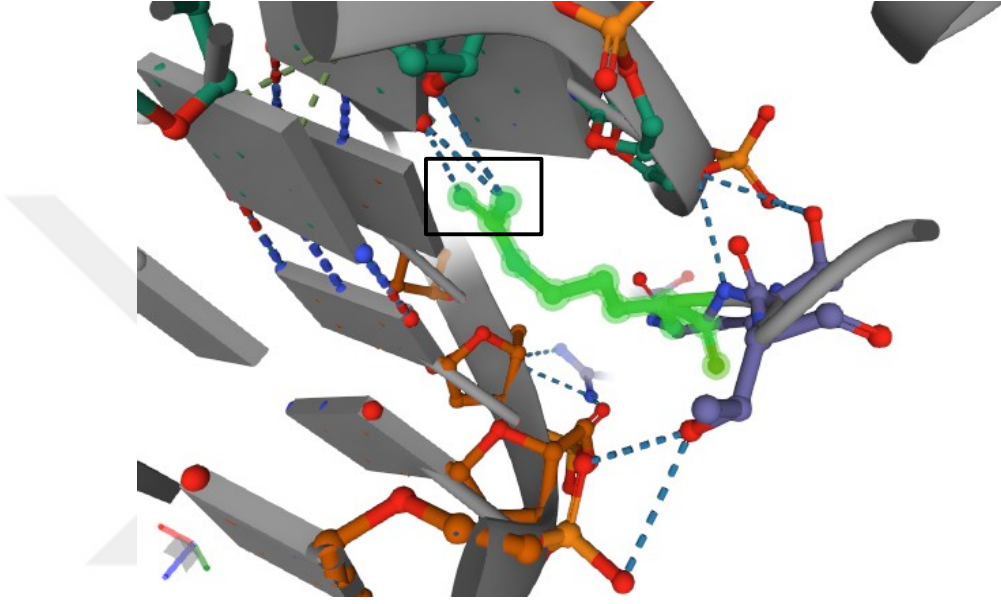
Modelleyicilerde modeller elde edildikten sonra yapıların iyileştirilmesi için modelleyicinin ürettiği en iyi model ile 5000 adımlı ve CHARMM kuvvet alanlı minimizasyon yapılmıştır. Ancak bu işlem yapıların hem belkemiğini hem de yan zincirlerini kötüleştirilmiş, bazılarında homeodomain bölgesi bozulmuştur. Bu durum yine IDR bölgenin varlığı ve bu bölgenin etrafındaki katlanmaların düzensiz olmasından kaynaklanabilir (40).

Homeodomain bölgesi yüksek güvenilirlikle elde edilmiş ve deneysel sonucu bulunduğundan bu bölgeye denk gelen *benign*, işlev kaybı ve işlev kazanımı olan varyantlar modellenmiş ve yapısal olarak incelenmiştir. İncelemede RMSD hesaplaması yapılmış ve varyantın yabancı homeodomainden değişen amino asidinin etrafındaki amino asitlerle etkileşimine bakılmıştır. RMSD hesaplamasında *benign* varyant ile yabancı homeodomain arasında bir değişim olmaması ve işlev kaybı ve kazanımı olan varyantlarda bir değişim olması beklenirken göz ardı edilemeyecek bir değişim meydana gelmemiştir. Bunun nedeni yabancı homeodomain olarak deneysel sonucun kullanılmış olması olabilir. Deneysel sonucun protein dizisi varyant

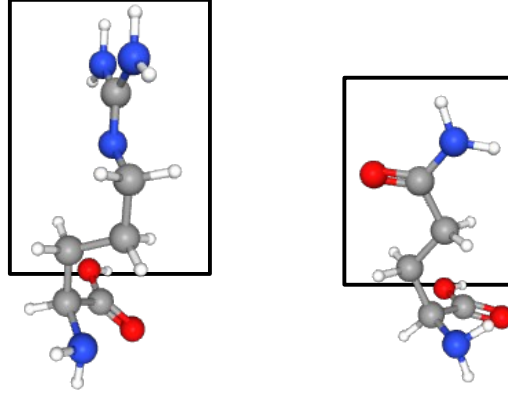
modellemesinde kullanılmış ColabFold’da modellenmiş olsaydı RMSD değerlerinin *benign* ve diğer varyantlarda daha çok farklılık gösterebileceği düşünülmüş olmakla birlikte tez çalışmasında bu henüz tamamlanmamıştır. Amino asit etkileşimlerine bakıldığında ise beklendiği gibi *benign* varyant ile yabancı homeodomain arasında bir fark görülmemiştir. Bölge yüksek güvenilirlikle modellenbildiği için, yabancı homeodomain için deneysel ya da hesaplamalı yöntemle elde edilmiş yapıdan hangisinin kullanılacağı amino asit etkileşimleri incelenirken önemsiz olduğu düşünülmektedir. İşlev kazanımı olan varyantta yabancı homeodomainin etkileşimlerinden farklı olarak homeodomain bölgesinin dışındaki amino asitle etkileştiği görülmektedir. p.Arg218Gln varyantı dışındaki işlev kaybı olan varyantların ise homeodomainin etkileşimlerinden farklı olarak homeodomain bölgesindeki amino asitle etkileştiği görülmektedir. ALX4 DNA’ya bağlanma işlevini homeodomain bölgesi ile gerçekleştirdiğinden, değişen amino asidin proteine işlev kazandıran özellikte olunca bu bölgenin dışı ile, işlev kaybettiren özellikte olunca bu bölgede bir amino asit ile etkileşmesi ilgili varyantın işlev kaybettiren ya da kazandıran olmasına neden olmuş olabilir. Bu durumda homeodomainde meydana gelen işlev kazandıran ya da kaybettiren tek amino asit değişikliği varyant proteinin değişen amino asidinin diğer amino asitlerle etkileşimi sırasıyla domain bölgesi dışında veya içinde olabilir. Ancak bu hipotezin test edilmesi için daha fazla veriye ihtiyaç duyulmaktadır.

p.Arg218Gln varyantında, yabancı homeodomaindeki Arg amino asidi ile varyant proteindeki değiştiği Gln amino asidinin etkileştiği amino asitler farklıdır, ancak belkemiği bağı yapmaktadır. Belkemiği her amino asitte aynı olduğundan bu değişiklik önemli bulunmamıştır. Ancak tez teslimine yakın tarihte yayınlanan PDB veri tabanındaki 8OSB kimlikli ALX4 homeodomain bölgesi, bu varyantın etkileşimlerinin tekrar değerlendirilmesini sağlamıştır (41). 8OSB proteini ALX4’ün tamamında 216-277 amino asitleri arasına denk gelmekte ve PDB veri tabanında TWIST1 genine TCF4 proteini ile bağlı şekilde bulunmaktadır (Şekil 5.2). Arjinin, DNA ile bağ yaparken iki nitrojenini de kullanmaktadır. Burada Arg ve Gln kimyasına bakıldığında glutaminin arjinininden daha kısa olduğu ve bir nitrojeninin

olduđu grlmektedir (Şekil 5.3). Bu durum glutaminin DNA'ya ulaşmasını ve nitrojen eksikliğinden bağ yapmasını engelleyebilir. Bundan dolayı p.Arg218Gln varyantı homeodomainin DNA'ya bağlanma işlevini etkiliyor olabilir ve etkisi patojenik olarak tahmin edilebilir.



Şekil 5.2. ALX4 homeodomaininde arjininin DNA'ya bağlanması. Sol gri basamaklar DNA'nın nükleotitlerini temsil ederken sağda ALX4 homodomaininin ilk 3 amino asidi gözükmektedir. Yeşil ile işaretlenen arjinindir. Kutu içine alınan bölge arjininin DNA ile bağ yaptığı iki nitrojenini göstermektedir.



Şekil 5.3. Arjinin ve glutaminin kimyasal yapıları. Yapılar NIH PubChem kimyasal yapı veri tabanından yakın ölçeklerde alınmıştır. Kutu içine alınan kısım yan zinciri göstermektedir.

Sonuç olarak bu tez çalışmasında ilk kez ALX4 proteininin deneysel modellerinde işlev kazandıran ve fonksiyon kaybettiren mutasyonların fenotip çeşitliliğine etkisi çalışılmış olup homeodomain bölgesinde yer alan mutasyonların amino asit etkileşimleri ile işlev kazandırma ve kaybettirme etkilerinin tahmin edilebileceği gösterilmiştir.

6. SONUÇ VE ÖNERİLER

1. ALX4'ün homeodomaini dışındaki bölgelere benzer ve yapısı bilinen bir protein bulunamamıştır.
2. Yabancı ALX4'ün tamamı günümüzde kullanılan araçlarla modellenmeye çalışılmış, ancak başarılı olunamamıştır. Bunda en büyük etken düzensiz domain bölgeleridir. Buna karşın, homeodomaini başarılı olarak modellenebilmiştir.
3. ALX4 için en iyi modellemeyi yapan modelleyicinin 48 geri dönüşümlü parametresi ile şablon tabanlı ColabFold olduğu bulunmuştur.
4. ALX4 için şablon tabanlı veya şablon tabansız modelleme pek farklılık yaratmamıştır. Ancak i-TASSER'ın şablon tabansız modelleme yapan modelleyicisi D-I-TASSER, i-TASSER'dan daha kötü performans sergilemektedir.
5. 5000 adımlı ve CHARMM kuvvet alanlı minimizasyon modelleri iyileştirmede yeterli olamamıştır.
6. Farklı fenotipik etkileri olan homeodomain bölgesindeki ALX4 varyantlarının yapısal değişikliğini RMSD hesaplamasını kullanarak saptamak mümkün değildir.
7. Homeodomaindeki ALX4 varyant proteinlerinin değişen amino asidindeki yabancı tipten farklı olan etkileşimler, varyantların protein sonuçları ve fenotipik etkileri ile örtüşmektedir. Kısıtlı sayıda incelemede bu yöntem AlphaMissense'ten daha başarılı bulunmuştur.
8. Bu yöntem ile etkisi tahmin edilemeyen etkileşim, yine yabancı ve varyant amino asitlerin kimyasal yapıları incelenerek tahmin edilebilmiştir.
9. ALX4 homeodomain-DNA etkileşiminin modellenmesi ve moleküler dinamik simülasyonlarının yapılması varyant ekisinin anlaşılmasına katkı sağlayabilir.

7. KAYNAKLAR

1. Alberts B. Molecular biology of the cell. Sixth edition. New York, NY: Garland Science, Taylor and Francis Group; 2015.
2. Alberts B. Essential cell biology. Fourth edition. New York, NY: Garland Science; 2013.
3. Anfinsen CB. Principles that govern the folding of protein chains. *Science*. 1973;181(4096):223-30.
4. Lodish HF. Molecular cell biology. Eighth edition. New York: W.H. Freeman-Macmillan Learning; 2016.
5. Blue Book [Internet]. 2023 [Erişim Tarihi 16 Nisan 2024]. Erişim adresi: <https://iupac.qmul.ac.uk/BlueBook/>
6. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*. 1963;7(1):95-9.
7. Petrucci, R.H., William, S., Herring, F.G., Madura, J.D. General Chemistry: Principles and Modern Applications. 9th Edition. Upper Saddle River:Pearson Education;2007.
8. Brändén CI, Tooze J. Introduction to protein structure. 2nd ed. New York: Garland Pub; 1999.
9. Goodsell DS, Olson AJ. Structural Symmetry and Protein Function. *Annu Rev Biophys Biomol Struct*. 2000;29(1):105-53.
10. Bowley R, Sánchez M. Introductory statistical mechanics. 2nd ed. Oxford:Oxford University Press; 1999.
11. Scherer SW, Lee C, Birney E, Altshuler DM, Eichler EE, Carter NP ve ark. Challenges and standards in integrating surveys of structural variation. *Nat Genet*. 2007;39(7 Suppl):S7-15.
12. Holland PW, Booth HAF, Bruford EA. Classification and nomenclature of all human homeobox genes. *BMC Biology*. 2007;5(1):47.
13. McGonnell IM, Graham A, Richardson J, Fish JL, Depew MJ, Dee CT ve ark. Evolution of the *Alx* homeobox gene family: parallel retention and independent loss of the vertebrate *Alx3* gene. *Evol Dev*. 2011;13(1): 343-351.
14. Kayserili H, Uz E, Niessen C, Vargel I, Alanay Y, Tuncbilek G ve ark. ALX4 dysfunction disrupts craniofacial and epidermal development. *Hum Mol Genet*. 2009;18(22):4357-66.

15. Uz E, Alanay Y, Aktas D, Vargel I, Gucer S, Tuncbilek G ve ark. Disruption of ALX1 causes extreme microphthalmia and severe facial clefting: expanding the spectrum of autosomal-recessive ALX-related frontonasal dysplasia. *Am J Hum Genet.* 2010;86(5):789-96.
16. Çetinkaya A, Akarsu NA. The ALX Homeobox Gene Family and Frontonasal Dysplasias. Erickson RP, Wynshaw-Boris AJ, editors. *Epstein's Inborn Errors of Development.* 3rd edition. Oxford University Press; 2016. s. 747-52.
17. Ring CS, Cohen FE. Modeling protein structures: construction and their applications. *FASEB j.* 1993;7(9):783-90.
18. Anand, Namrata, Possu Huang. Generative Modeling for Protein Structures. *Advances in Neural Information Processing Systems.* 2018;31.
19. Fox BG, Goulding C, Malkowski MG, Stewart L, Deacon A. Structural genomics: from genes to structures with valuable materials and many questions in between. *Nat Methods.* 2008;5(2):129-32.
20. Clark DP, Pazdernik NJ, McGehee MR. *Molecular biology.* Third edition. London San Diego Cambridge Oxford: Academic Press; 2019.
21. PDB [Internet]. 1995 [Erişim Tarihi 3 Temmuz 2024]. Erişim adresi: <https://www.rcsb.org>
22. Xu D, Xu Y, Uberbacher EC. Computational Tools For Protein Modeling. *Current Protein and Peptide Science.* 2000;1;1-21.
23. Mariani V, Biasini M, Barbato A, Schwede T. IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics.* 2013;29(21):2722-8.
24. Kryshtafovych A, Malhotra S, Monastyrskyy B, Cragolini T, Joseph AP, Chiu W, vd. Cryo-EM targets in CASP13: overview and evaluation of results. *Proteins.* 2019;87(12):1128-40.
25. Pereira J, Simpkin AJ, Hartmann MD, Rigden DJ, Keegan RM, Lupas AN. High-accuracy protein structure prediction in CASP14. *Proteins.* 2021;89(12):1687-99.
26. Tekin B., Gurbanov, R. AlphaFold: Derin Öğrenme ve Sinir Ağları Yoluyla Protein Katlanmasında Devrim Yaratmak. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi.* 202; 22(44), 445-466.
27. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596(7873):583-9.

28. McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* 2004;32(Web Server issue):W20-5.
29. BLASTP [Internet]. 2024 [Erişim Tarihi 3 Temmuz 2024]. Erişim adresi: <https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>
30. Kaminski K, Ludwiczak J, Pawlicki K, Alva V, Dunin-Horkawicz S. pLM-BLAST: distant homology detection based on direct comparison of sequence representations from protein language models. *Bioinformatics.* 2023; 39(10):btad579.
31. pLM-BLAST [Internet]. 2024 [Erişim Tarihi 3 Temmuz 2024]. Erişim adresi <https://toolkit.tuebingen.mpg.de/tools/plmblast>
32. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics.* 1996;14(1):33-8.
33. Umi Baroroh SS, Muscifa ZS, Destiarani W, Rohmatullah FG, Yusuf M. Molecular interaction analysis and visualization of protein-ligand docking using Biovia Discovery Studio Visualizer. *Indonesian Journal of Computational Biology (IJCB).* 2023;2(1):22-30.
34. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, vd. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr.* 2010;66(Pt 1):12-21.
35. James C. Phillips, David J. Hardy, Julio D. C. Maia, John E. Stone, Joao V. Ribeiro, Rafael C. Bernardi ve ark. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *Journal of Chemical Physics,* 2020;153(4):044130.
36. PDB 2M0C [Internet]. 2012 [Erişim Tarihi 3 Temmuz 2024]. Erişim adresi: <https://www.rcsb.org/structure/2m0c>
37. gnomAD SNV 11-44275397-G-A (GRCh38). 2018 [Erişim Tarihi: 16 Nisan 2023]. Erişim adresi: https://gnomad.broadinstitute.org/variant/11-44275397-G-A?dataset=gnomad_r4
38. Fuxreiter M, Simon I, Bondos S. Dynamic protein-DNA recognition: beyond what can be seen. *Trends Biochem Sci.* 2011;36(8):415-23.
39. Cermakova K, Hodges HC. Interaction modules that impart specificity to disordered protein. *Trends Biochem Sci.* Mayıs 2023;48(5):477-90.
40. Feig M, Mirjalili V. Protein Structure Refinement via Molecular-Dynamics Simulations: What works and what does not? *Proteins.* 2016;84(Suppl 1):282-92.

41. PDB 8OSB [Internet]. 2024 [Eriřim Tarihi 26 Mayıs 2024]. Eriřim adresi: <https://www.rcsb.org/structure/8OSB>



8. EKLER

EK-1: Tez Çalışması ile İlgili Etik Kurul İzni



T.C.
HACETTEPE ÜNİVERSİTESİ
Girişimsel Olmayan Klinik Araştırmalar Etik Kurulu

Sayı : 16969557 - 3202

Konu :

04.10.2022

Prof. Dr. Nurten AKARSU
Tıp Fakültesi
Tıbbi Genetik Anabilim Dalı
Öğretim Üyesi

Sayın Prof. Dr. AKARSU,

Kurulumuza değerlendirilmek üzere sunduğunuz GO 22/916 kayıt numaralı ve "*Fenotip Çeşitliliğini Açıklamada ALX4 Varyantlarının Protein Modellerinde İncelenmesi*" başlıklı proje Kurulumuzun 04.10.2022 tarihli toplantısında değerlendirilmiş olup, çalışmanın erişime açık veri tabanlarından veri toplanması yolu ile yapılacağı görülmüştür. Gönüllü insanlar üzerinde gerçekleştirilecek nitelikte olmayan bu tip çalışmalar Etik Kurulların kapsamı dışında kalmaktadır.

Bu yazı ilgili protokolün bilimsel ve etik açıdan incelendiğini belirtmek için Etik Kurul kararı yerine geçmek üzere hazırlanmıştır.

Prof. Dr. M. Özgür UYANIK
Başkan Vekili

EK
1. Toplantı Katılım Tutanağı (1 Adet)



T.C.

HACETTEPE ÜNİVERSİTESİ


Girişimsel Olmayan Klinik Araştırmalar Etik Kurulu

04/10/2022 Tarih ve 2022/15 Numaralı Toplantı

KATILIM LİSTESİ

S. No	Adı Soyadı	İmza
1.	Prof. Dr. Nüket Paksoy ERBAYDAR (Başkan)	İZİNLİ
2.	Prof. Dr. G. Burça AYDIN	
3.	Prof. Dr. Özgür UYANIK (Başkan V.)	
4.	Prof. Dr. Ayşe Kin İŞLER	
5.	Prof. Dr. Sibel PEHLİVAN	
6.	Prof. Dr. Tolga YILDIRIM	
7.	Doç. Dr. H. Tuna Çak ESEN	İZİNLİ
8.	Doç. Dr. Betül Çelebi SALTIK	
9.	Doç. Dr. Hânde Güney DENİZ	İZİNLİ
10.	Doç. Dr. Merve BATUK	
11.	Doç. Dr. Gülten KOÇ	
12.	Dr. Öğr. Üyesi Müge DEMİR	
13.	Av. Buket ÇINAR	

EK-2: Tez Çalışması Orijinallik Raporu



Dijital Makbuz

Bu makbuz ödevinizin Turnitin'e ulaştığını bildirmektedir. Gönderiminize dair bilgiler şöyledir:

Gönderinizin ilk sayfası aşağıda gönderilmektedir.

Gönderen:	Bengü Yaren Beyaz
Ödev başlığı:	FENOTİP ÇEŞİTLİLİĞİNİ AÇIKLAMADA ALX4 VARYANTLARININ...
Gönderi Başlığı:	FENOTİP ÇEŞİTLİLİĞİNİ AÇIKLAMADA ALX4 VARYANTLARININ...
Dosya adı:	tuntin.docx
Dosya boyutu:	9.41M
Sayfa sayısı:	78
Kelime sayısı:	12,479
Karakter sayısı:	87,478
Gönderim Tarihi:	22-Ağu-2024 06:26ÖS (UTC+0300)
Gönderim Numarası:	2436155575

ÖZET

Kıyık, K.Y., Fenotip Çeşitliliği Açıklamada ALX4 Varyantlarına Pratikte Modellerde İncelenen, Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü Biyobiyoteknoloji Programı Yüksek Lisans Tezi, Ankara, 2024. ALX4, ALX gen ailesi, embriyonik gelişimde önemli rol oynayan ve manyetik olarak inaktivite kullanan ve diğer mallelasyonlarda gösteren heterodimeri içeren transkripsiyon faktörüdür. Adenin (peptidomimetik ALX) ve ALX4 materyalleri potansiyel işlev kaybi etkilerini araştırılmaya çalışılmıştır. ALX4'ün işlev kaybi ve kısıtlama sonucu ortaya çıkan diğer fenotipler değerlendirilmiştir. ALX4'ün tamamen işlevsiz yapısı ve bu durumunun nedeni hakkında bilimsel bilgilerdir. Bu tez çalışmada, ALX4 geninin modifikasyonu yapılmıştır. Çiçekli bir modelde, ALX4'ün heterodimeriyle ilgili etkileri araştırılmaya çalışılmıştır. ALX4'ün işlev kaybi ve kısıtlama sonucu ortaya çıkan diğer fenotipler değerlendirilmiştir.

Araştırma Kilitleri: ALX4, genetik modifikasyon, genetik inaktivite modifikasyonu, işlev kaybi, işlev kısıtlama

Copyright 2024 Turnitin. Tüm hakları saklıdır.

Bengü Yaren Beyaz

FENOTİP ÇEŞİTLİLİĞİNİ AÇIKLAMADA ALX4 VARYANTLARININ PROTEİN MODELLERİNDE İNCELENMESİ

 FENOTİP ÇEŞİTLİLİĞİNİ AÇIKLAMADA ALX4 VARYANTLARININ PROTEİN MODELLERİNDE İNCELENMESİ
 Bengü Yaren Beyaz Tez
 Hacettepe University

Belge Ayrıntıları

Gönderi Kimliği
trnoid::1:2988130694

Gönderi Tarihi
22 Ağu 2024 18:25 GMT+3

İndirme Tarihi
22 Ağu 2024 18:33 GMT+3

Dosya Adı
turnitin.docx

Dosya Boyutu
9.4 MB

78 Sayfa

12.479 Sözcük

87.478 Karakter




2% Genel Benzerlik

Her veri tabanı için çakışan kaynaklar da dâhil tüm eşleşmelerin kombine toplamı.

Rapordan Filtrelenen


- Bibliyografya
- Alıntılanan Metin

Ön Sıradaki Kaynaklar

- 2%  İnternet kaynakları
- 1%  Yayınlar
- 1%  Gönderilen çalışmalar (Öğrenci Makaleleri)

Bütünlük Bayrakları

İnceleme için 1 Bütünlük Bayrağı

-  **Değiştirilen Karakterler**
14 sayfada 33 şüpheli karakter
Harfler başka bir alfabeden benzer karakterlerle değiştirilir.

Sistemimizin algoritmaları bir belgede, onu normal bir gönderiden ayıracabilecek her türlü tutarsızlığı derinlemesine inceler. Tuhaf bir şey fark edersek incelemeniz için bayrak ekleriz.

Bir Bayrak mutlaka bir sorun olduğunu göstermez. Ancak daha fazla inceleme için dikkatinizi vermenizi öneririz.

9. ÖZGEÇMİŞ