



T.C.

SELÇUK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

İNSAN KOLLAJEN DİZİ MUTASYONLARININ OSTEOGENEZİS İMPERFEKTA HASTALIĞI ÜZERİNDEKİ İLİŞKİSİNİN PYTHON PROGRAMLAMA DİLİ KULLANILARAK ANALİZ EDİLMESİ

Gülsüm TİRAŞ

YÜKSEK LİSANS TEZİ

Biyomedikal Mühendisliği

Ocak-2021
KONYA
Her Hakkı Saklıdır

TEZ KABUL VE ONAYI

Gülsüm TİRAŞ tarafından hazırlanan "İnsan kollajen dizi mutasyonlarının Osteogenezis İmperfekta hastalığı üzerindeki İlişkisinin Python programlama dili kullanılarak analiz edilmesi" adlı tez çalışması 11/02/2021 tarihinde aşağıdaki jüri tarafından oy birliği ile Selçuk Üniversitesi Fen Bilimleri Enstitüsü Biyomedikal Mühendisliği Anabilim Dalı'nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri	**İmza**
Başkan Dr. Öğr. Üyesi Adem GÖLCÜK	……..
Danışman Dr. Öğr. Üyesi Esma ERYILMAZ DOĞAN	……..
Üye Prof. Dr. Ercan TÜRKKAN	……..

Yukarıdaki sonucu onaylarım.

Prof. Dr. SAİT GEZGİN
FBE Müdürü

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Gülsüm TİRAŞ
11.02.2021

ÖZET

YÜKSEK LİSANS TEZİ

İNSAN KOLLAJEN DİZİ MUTASYONLARININ OSTEOGENEZİS İMPERFEKTA HASTALIĞI ÜZERİNDEKİ İLİŞKİSİNİN PYTHON PROGRAMLAMA DİLİ KULLANILARAK ANALİZ EDİLMESİ

Gülsüm TİRAŞ

Selçuk Üniversitesi Fen Bilimleri Enstitüsü
Biyomedikal Mühendisliği Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Esma ERYILMAZ DOĞAN

2021, 39 Sayfa

Jüri
Dr. Öğr. Üyesi Esma ERYILMAZ DOĞAN
Dr. Öğr. Üyesi Adem GÖLCÜK
Prof. Dr. Ercan TÜRKKAN

Hücre dışı matris, hücre-hücre etkileşimlerine aracılık eden önemli bir bileşen olarak ortaya çıkmakta olup, hücre desteği için bir iskele görevi görmektedir. Hücre dışı matrisin temel bileşeni kollajendir. Kollajenler, insanlarda 28 üyeden oluşan geniş bir aileye ait karmaşık protein yapılarıdır. Genellikle alfa zincirleri olarak bilinen üç farklı polipeptit zinciri ile üçlü bir sarmal yapı oluştururlar. Ayrıca, bu üçlü sarmalın kendine özgü bir "Gly-Xaa-Yaa" dizisine sahip olduğu bulunmuştur. Her dizide ki glisin (Gly) varlığı, en küçük amino asit olma özelliği nedeniyle sarmal yapının stabilitesini açıklar. Xaa ve Yaa herhangi bir amino asit olabilir ancak çoğunlukla prolindir. Bu nedenle, kollajenin glisin ve prolin bakımından zengin bir protein olduğu bilinmektedir. Kollajen, bir dizi patolojik durumda rol oynamaktadır. Üçlü sarmal alanlardaki bir aminoasit diğer aminoasitlerle yer değiştirdiğinde kollajen yapısı tahrip olur. Kollajen yapısındaki bozulma çeşitli kalıtsal hastalıklara ve işlev bozukluklarına sebep olmaktadır.

Bu çalışmada, insan vücudunda en fazla bulunan tip I kollajenin alfa-1 zincirinde meydana gelen mutasyonlar incelenmiştir. Daha önce yapılan çalışmalardan yola çıkarak tip-I kollajen alfa-1 zincirlerindeki mutasyonlardan kaynaklı hastalıklardan biri olan kırılgan kemik hastalığı(OI) tip 2 üzerinde durulmuştur. Bodian ve arkadaşlarının yaptığı çalışmada farklı kişilerden elde edilen mutasyonlar UniprotKB veri tabanından elde edilmiştir. Mutasyona uğramış aminoasitler değiştirilerek yeni alfa-1 zincirleri elde edilmiştir. PyCharm programında Python programlama diliyle yazılan kodlar ile hesaplamalar yapılmıştır. Elde edilen mutasyonlar bir alfa-I zincirinde birleştirilmiş ve en çok mutasyona uğrayan amino asitin glisin(Gly) olduğu görülmüştür. Glisin amino asiti kollajen alfa-I zincirinin sarmal yapısının stabilitesini etkilediğinden kollajen kaynaklı hastalıklara etkisi olduğu düşünülebilir. İncelenen mutasyonlarda en sık tekrar eden mutasyonlar (glisin(G)>arjinin(R), glisin(G)>serin(S), glisin(G)>aspartat(D)) tespit edilmiştir. Mutasyona uğrayan amino asitlerin değişen molekül ağırlıkları, izoelektrik değerleri, hidropati indeksleri ve fiziksel özellikleri incelenmiştir. Karşılaştırma sonucunda ise molekül kütlesinde artış, izoelektrik noktasında değişim, hidropati indeksinde küçülme, yük durumunda ve asidik-bazik özelliklerinde değişim gözlenmektedir. Değişen bu özelliklerin kırılgan kemik hastalığına etkisi yorumlanmıştır.

Anahtar Kelimeler: Alfa-1 zinciri, Hücre dışı matris, Kollajen, Kırılgan kemik hastalığı, Python.

ABSTRACT

MS THESIS

ANALYSIS OF THE RELATIONSHIP OF HUMAN COLLAGEN SEQUENCE MUTATIONS ON OSTEOGENESIS IMPERFECTAL DISEASE USING PYTHON PROGRAMMING LANGUAGE

Gülsüm TİRAŞ

The Graduate School Of Natural And Applıed Scıence Of Selçuk Unıversıty
The Degree Of Master Of Scıence
In Bıomedical Engıneerıng

Advisor: Asst. Prof. Dr. Esma ERYILMAZ DOĞAN

2021, 39 Pages

Jury
Asst. Prof. Dr. Esma ERYILMAZ DOĞAN
Asst. Prof. Dr. Adem GÖLCÜK
Prof. Dr. Ercan TÜRKKAN

The extracellular matrix is emerging as an important component mediating cell-cell interactions, along with its role as a scaffold for cell support. The main component of the extracellular matrix is collagen. Collagens are complex protein structures belonging to a large family consisting of 28 members in humans. They form a triple helical structure with three distinct polypeptide chains, commonly known as alpha chains. Also, this triple helix is found to possess a characteristic fragment of 'Gly-Xaa-Yaa'. The presence of glycine (Gly) in every sequence accounts for the stability of the helical structure owing to its property of being the smallest amino acid. Xaa and Yaa can be any amino acid but are mostly proline. Thus, collagen is known to be a glycine and proline rich protein. Collagen has been implicated in a number of pathological conditions. When an amino acid in triple helix areas is replaced with other amino acids, the collagen structure is destroyed. The deterioration in the collagen structure causes various hereditary diseases and dysfunctions.

In this study, the mutations that occur in the alpha-1 chain of type I collagen, which is the most common in the human body, were examined. Based on previous studies, brittle bone disease (OI) type 2, which is one of the diseases caused by mutations in type-I collagen alpha-1 chains, has been focused on. In the study of Bodian et al., Mutations obtained from different individuals were obtained from the UniprotKB database. New alpha-1 chains have been obtained by changing the mutated amino acids. In the PyCharm program, calculations were made with the codes written in the Python programming language. The mutations obtained were combined in an alpha-I chain and it was seen that the most mutated amino acid was glycine (Gly). Since glycine amino acid affects the stability of the helix structure of the collagen alpha-I chain, it can be considered to have an effect on collagen-induced diseases. The most frequently recurring mutations (glycine (G)> arginine (R), glycine (G)> serine (S), glycine (G)>aspartate (D)) were detected in the mutations examined. The varying molecular weights, isoelectric values, hydropathy indexes and physical properties of the mutated amino acids were examined. As a result of comparison, increase in molecular mass, change in isoelectric point, decrease in hydropathy index, change in charge state and acid-base properties are observed. The effect of these changed features on brittle bone disease has been interpreted.

Keywords: Alpha-1 chain, Collagen, Extracellular matrix, Osteogenesis imperfecta, Python.

ÖNSÖZ

Bu çalışmanın gerçekleştirilmesinde, bana destek veren ve yol gösteren değerli danışman hocam; Esma ERYILMAZ DOĞAN ’a yaşadığım tüm her şeyde beni bir an olsun yalnız bırakmayan, desteklerini esirgemeyen ve sevgilerini her daim yanımda hissettiğim Annem ’e, ve Babam ’a sonsuz teşekkürlerimi sunarım.

Gülsüm TİRAŞ
KONYA-2021

İÇİNDEKİLER

ÖZET ..... **iv**

ABSTRACT ..... **v**

ÖNSÖZ ..... **vi**

İÇİNDEKİLER ..... **vii**

SİMGELER VE KISALTMALAR ..... **viii**

1. GİRİŞ **1**

2. KAYNAK ARAŞTIRMASI **3**

3. MATERYAL VE YÖNTEM **12**

3.1. Kollajen 12
3.2. Kırılgan Kemik Hastalığı (Osteogenezis Imperfecta) 14
3.3. PyCharm programı ve Python programlama dili 14
3.4. PyCharm Program Kurulumu 16
3.5. UniProtKB / SWISS-PROT 17
3.6. Amino asitler 17
3.7. İzoelektrik Noktası (pı) 19
3.8. Hidropati İndeksi 19

4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA **21**

4.1. Elde edilen alfa-1 zincirlerinin PyCharm programında analizi 21
4.2. Alfa-I zincirinde birleştirilen mutasyonlardan elde edilen zincir ile normal alfa-I zincirinin amino asit dizilimlerinin karşılaştırılması 21
4.3. Tekrarlanan mutasyonların incelenmesi 22
4.4. Glisin ile glisin yerine ikame edilen aminoasitlerin özelliklerinin karşılaştırılması 23

5. SONUÇLAR VE ÖNERİLER **26**

5.1 Sonuçlar 26
5.2 Öneriler 26

KAYNAKLAR **27**

ÖZGEÇMİŞ ..... Hata! Yer işareti tanımlanmamış.

SİMGELER VE KISALTMALAR

Kısaltmalar

COL1A1 :Kollajen Tip 1 alfa-I zinciri

COL1A2 :Kollajen Tip 1 alfa-II zinciri

CRTAP :Kıkırdak İlişkili protein

LEPRE :Lösin proteinle zenginleştirilmiş proteglikan

OI 1 :Osteogenez imperfekta tip 1 hastalığı

OI 2 :Osteogenez imperfekta tip 2 hastalığı

Alfa-I :Proteinlerin alfa-I zinciri

HPLC :Yüksek performanslı sıvı kromatografisi

1. GİRİŞ

İnsan vücudu karmaşık yapıya sahip fakat uyum içinde çalışan sistemleri barındıran mükemmel bir organizmadır. İnsanın sağlıklı şekilde yaşamını sürdürebilmesi için, vücuttaki sistemler birbiriyle bağlantılı olarak çalışır. Her birinin ayrı görevi bulunmaktadır ve bu görevler sistem içerisinde sürekli olarak devam eder.

Hücre, canlının yaşamsal özelliklerini taşıyan, yapı ve görev bakımından en küçük parçasıdır. Hücreler dokuları, dokular organları, organlar sistemleri, sistemler de organizmayı oluşturur. Protein ise insan vücudundaki her hücrede yer alan önemli bir yapıtaşıdır. Vücutta sudan sonra en bol miktarda bulunan maddedir.

İnsan vücudunda çok sayıda bulunan protein Kollajen'dir. Kollajen deri, bağ dokusu, tendon, eklem, tırnak, diş, saç, kıkırdak ve kemik gibi vücudumuzun önemli bileşenlerinde bulunur. Vücudumuzun temel yapı taşı olan proteinlerin çoğu kollajen kaynaklıdır ve kollajen vücudumuzda birçok kritik göreve sahiptir. Kollajen, insanların dokularında bulunan, fibroblastlar ve hücreler tarafından oluşturulan bir protein türüdür. Temel görevi bağ dokusunu güçlendirmek ve vücudumuzun bütünlüğünü korumaktır. Cilde esneklik ve sıkılık kazandırmak, eklem, tendon ve organlarımızı bir arada tutmak, diğer görevleri arasında sıralanabilir.

Kollajen, biyolojik olarak parçalanabilirlik, zayıf antijenite ve kendi kendine birleşme gibi çeşitli özellikleri nedeniyle, biyolojik olarak uyumlu ve güvenlidir. Kollajen, kendi kendine birleşme ve çapraz bağlanma ile ekstra mukavemet ve stabiliteye sahip lifler oluşturabilir. Bu nedenle tıbbi uygulamalarda da sıklıkla kullanılan bir biyomalzemedir. Kollajen'in yapısındaki değişiklikler ve kollajen proteinlerini kodlayan genlerdeki mutasyonlar sonucu çeşitli hastalıklar oluşabilir. Bu nedenle vücudumuzda önemli bir role sahip, tıbbi uygulamalarda sıklıkla kullanılan kollajen proteinini daha iyi analiz etmek gerekmektedir.

Bu bilgilerden yola çıkarak yapılan literatür taramasında kollajen genlerindeki mutasyon ve bunun sonucunda meydana gelen hastalıklarla ilgili araştırmalar yapılmıştır. Biz bu çalışmada, kollajenin yapısında bulunan amino asitlerin mutasyonları sonucu değişen fiziksel özelliklerini ve değişen aminoasit sayılarını inceledik. Çalışmamızda, aminoasitlerin mutasyonu sonucu değişen fiziksel özelliklerin yorumlanması ve değişen aminoasit sayılarının Pycharm programı aracığıyla hesaplanması sonucu elde edilen sayısal verilere bakılarak kollajen kaynaklı hastalıklar üzerindeki etkisinin yorumlanması amaçlanmaktadır. Bu amaçla veri tabanında kayıtlı

daha önce yapılan çalışmalardan elde edilen mutasyonlar kullanılmıştır. Amino asitlerin bilinen fiziksel özellikleri kullanılarak mutasyon sonucu oluşan değişimin hastalığa etkisi olabileceği değerlendirilmiştir. Elde edilen tüm mutasyonlar incelenerek en çok mutasyona uğrayan amino asit sayısı hesaplanmıştır. Hesaplama sonucunda kollajen yapısındaki değişim yorumlanmıştır.

2. KAYNAK ARAŞTIRMASI

Yapılan literatür taramalarında kollajen molekülünün yapısı ve yapısındaki mutasyonlardan kaynaklanan hastalıklarla ilgili çalışmalar incelenmiştir. İncelenen çalışmalara, kollajen yapısındaki farklılıklardan kaynaklı hastalıklarla ilgili hesaplamalı teorik katkı sağlamak amacıyla tez konusu oluşturulmuştur.

Nassa vd. Biyo-hesaplama araçları kullanarak insan kollajeninin alfa-1 dizilerinin karşılaştırmalı bir karakterizasyonunu yapmışlardır. Kollajen ailesi, insanlarda bulunan 28 genetik olarak farklı üyeden oluşan büyük ve karmaşık bir ailedir (Cen, Liu, Cui, Zhang, & Cao, 2008). Kollajen ’de bulunan üç alfa zincirinden, 28 insan kollajeninin hepsinde alfa-1 zinciri gözlenmiştir. Bu nedenle, kollajen proteinlerinin daha ileri analizi için alfa-1 zinciri kullanılmıştır. İnsan kollajeninin alfa-1 sekanslarının fiziko-kimyasal, ikincil yapısal, fonksiyonel ve filogenetik analizi yapılmıştır. Her insan kollajeninin alfa-1 zincirinin fonksiyonel rolü analiz edilmiştir. MultAlin aracı, artan sayıda girdi dizisiyle azaltılmış bir dizi benzerliğine tanık olunan protein dizisi hizalamasını gerçekleştirmek için kullanmışlardır. Alfa-1 zincirinin 28 üyenin hepsinde protein dizisi boyunca çeşitli pozisyonlarda benzerlikler ve farklılıklar gösterdiği anlaşıldı. ExPASY’ın ProtParam aracı kullanılarak insan kaynaklı her bir alfa-1 kollajen sekansının aminoasit kompozisyonunun hesaplanması, diğer aminoasitlere kıyasla çok yüksek glisin ve prolin yüzdelerini göstermiştir. Tüm kollajenler de glisin içeriği, % 9.2, %10.9 ve % 11.5'lik bir değerle kollajen 12, 14 ve 20 hariç, % 12'den fazla olduğu görülmüştür. Büyük aminoasitlerin dizideki varlığı sterik engellemeye neden olabileceğinden, yüksek glisin yüzdesi kollajen üçlü sarmal yapının stabilitesini açıklar. Ayrıca, prolin içeriği, sırasıyla % 8.7, 8.6, 8.7 ve 10 değerleriyle Kollajen 6, 12, 14 ve 20 hariç, çoğu kollajen de % 10'dan fazlaydı. Prolin kalıntıları, helezonu dışarıya doğru işaret etmek, stabilize etmek ve ayrıca ikincil yapısal elemanların yapısal bozucusu olarak hareket etmek için aynı derecede önemlidir (Nimni, 1973). İnsanlarda bolca bulunan kollajen protein ailesinin alfa-1 zincirinin karakterizasyonu yeni anlayışlar verdi. Bu karşılaştırmalı karakterizasyona dayanarak, Kollajen 12, 14 ve 20'yi atipik bir davranışla birlikte birçok özellikte benzerlik gösteren potansiyel bir protein kümesi olarak varsaymışlardır. Bu araştırma, kollajen proteinlerinin çeşitli protein özelliklerine bir bakış açısı sağlamayı ve bu büyük aileyi karakterize etmeyi amaçlamaktadır. Aynı zamanda, kollajen ailesinin nispi muayenesine veya kollajen moleküllerinde herhangi

bir atipik özelliğin varlığına dayanarak hastalık koşullarında belirtilen potansiyel üyeleri önermeyi amaçlamaktadır (Nassa et al., 2012).

Kollajen, organizmalara mekanik stabilite, elastikiyet ve mukavemet sağlayan insan proteomunun üçte birini oluşturur. Tip-I kollajen molekülü, iki alfa-1 zinciri ve bir alfa-2 zincirinden oluşan bir heterotrimer üçlü sarmal moleküldür. Üç α-1 zincirinden oluşan tip I kollajen'in homotrimerik izoformu, sadece insanlarda fetal dokularda, fibroziste ve kanserde bulunur. Genetik kırılgan kemik hastalığının fare modeli, osteogenez kusurlu, oim, α -2 zincirinin bir α -1 zinciri ile değiştirilmesiyle karakterize edilir, bu da bir homotrimer kollajen molekülü ile sonuçlanır. Chang vd. tarafından Oim fare tendonu ve kemiği üzerinde yapılan deneysel çalışmalar, normal farelere kıyasla azalmış mekanik kuvvet göstermiştir. Gerçek dizi fare kollajen moleküllerinin (kollajen, tip I, α -1 ve α -2 zincir öncüsü (Mus musculus)) tip I heterotrimer ve homotrimeri arasındaki mekanik ve yapısal farklılıkları incelemek için moleküler simülasyon yaklaşımı kullanılmıştır. Kalıcı uzunluklar hesaplanır ve heterotrimer ve homotrimer kollajen molekülleri arasındaki farkların tam olarak anlaşılmasını sağlamak için yapısal analizler yapılır. Kalıcı uzunluk, uzun polimer zincirlerinin sertliğini ölçerek kollajen sertliğini tahmin etmek için uygun bir parametre haline getirir. Bununla birlikte, kalıcılık uzunluğu hesaplamaları zincirin homojen olduğunu varsayar ki bu kollajen molekülleri için doğru değildir. Hem deneysel çalışmalar (Arnold et al., 1998; Steplewski et al., 2004) hem de hesaplamalı modeller (Bodian, Madhan, Brodsky, & Klein, 2008; Bodian, Radmer, Holbert, & Klein, 2011), kollajen moleküllerinin üçlü sarmal alanının homojen olmadığını ve uzunluğu boyunca çeşitli yapısal ve biyolojik özelliklere sahip olduğunu göstermiştir (Chang, Shefelbine, & Buehler, 2012).

Atom düzeyindeki moleküler dinamikler (MD), uyarlanabilir Poisson-Boltzmann çözücü (APBS) hesaplamaları ve kollajen fibrilleri modeli mezoskal (orta ölçekli) bir molekül kullanılmıştır. Saf kollajen dokuya odaklanmış bir dizi sistematik molekül ölçeğine dayalı aşağıdan yukarıya hesaplama çalışmaları rapor edilmektedir. Gautieri ve arkadaşlarının yaptığı çalışmanın temeli bir referans tropokollajen peptid ve yedi tip glisin mutasyonuna sahip yedi modifiye versiyonuna dayanır. Analizin temel amacı, bu OI (Osteogenez imperfekta) mutasyonlarının tek molekül özellikleri, lateral moleküller arası etkileşimdeki değişiklikler ve ayrıca kollajen fibrillerin mekanik özelliklerindeki değişiklikleri araştırmaktır. Ayrıca kemik oluşumu sırasında daha büyük ölçekli

özellikler ve mineralizasyon üzerindeki olası etkileri tartışılmaktadır (Gautieri, Uzel, Vesentini, Redaelli, & Buehler, 2009).

2:1 oranında alfa1 (I) ve alfa2 zincirleri içeren tip I kollajenin ve alfa (III) zincirlerinden oluşan tip III kollajenin amino asit sekansları bilinmektedir. Bu alfa zincirlerinin dizilerinin istatistiksel bir analizi sunulmaktadır. Alfa-1(I) zincirinin ilk karşılaştırma matrisi analizi, D uzunluğuna (234 kalıntı) sahip önemli yinelenen dizi segmentleri bulunan ve bunu bir uzunluk d segmentinin gen çoğaltılması için kanıt olarak alan McLachlan (1976) tarafından gerçekleştirilmiştir. Zincirler arası karşılaştırma, üç alfa zinciri arasında yüksek düzeyde bir homoloji göstermiştir. Moleküllerin birleşmesinden sorumlu olan polar yüklü ve hidrofobik kalıntıların bir kısmı gibi etkileşimli amino asitler güçlü bir şekilde korunur. Zincir içi analiz, alfa zincirlerinin her biri 234 kalıntı uzunluğunda dört ilişkili D birimine bölündüğünü ortaya çıkardı. Alfa-1 zincirinin içyapısı da Fourier analizi ve diğer yöntemler ile analiz edildi ve D, D / 5, D / 6, D / 11 dalga boyuna sahip benzer amino asit gruplarının periyodik dağılımları bulundu. Bir zincir içindeki D birimleri arasında, polar kalıntılar hidrofobik amino asitlerden daha yüksek bir değişkenlik gösterir. D birimleri, D / 3 (78 kalıntı), D / 6 (39 kalıntı), D / l 1 (21 kalıntı) ve D / 13 (18 kalıntı) gibi diğer periyodiklikler, özellikle alfa1 (l) ve alfa1 (III) 'de gözlenmiştir. D ünitesi, etkileşimli polar yüklü ve hidrofobik artıkların oluşturduğu ve moleküllerin kümeleşmesini belirleyen fonksiyonel bir tekrardır. D / 3 birimi esas olarak prolin ve alanin gibi etkileşimli olmayan kalıntılar tarafından telaffuz edilir ve ona bir ilkel genin bir anımsaması gibi görünmektedir. Daha küçük periyodik tekrar eden birimler, ek genetik birimler olarak veya üçlü helisel adımı ve dolayısıyla moleküllerin yanal kümelenmesini belirleyen yapısal birimler olarak düşünülebilir. Alfa1 (1) ve alfa1 (III) 'ün aksine, alfa2 zinciri içyapısında daha az düzenlilik göstermiştir (Hofmann, Fietzek, & Kühn, 1980).

Kollajenler, çeşitli organ ve dokuların yapısal bütünlüğünün korunmasında baskın bir rol oynayan ve bir takım başka önemli biyolojik fonksiyonlara sahip olan bir hücre dışı matris proteinleri ailesidir. Kritik rolleri, şimdiye kadar 20'den fazla kollajen tipinden 12'si için 22 gende karakterize edilen 1000'den fazla mutasyona sahip hastalarda bulunan çok geniş klinik belirtilerle açıkça gösterilmiştir. Bu mutasyonların

sonuçları erken ölümcül fenotiplerden bazı yaygın hastalıklara yatkınlıklara kadar uzanmaktadır. Bu hastalıklar arasında osteogenez imperfekta, birçok kondrodisplazi, Ehlers-Danlos sendromunun çeşitli alt tipleri, Alport sendromu, Bethlem miyopati, epidermolizis bülloza, Knobloch sendromunun bazı alt tipleri ve ayrıca bazı osteoporoz, arteriyel anevrizmalar, osteoartroz ve intervertebral disk hastalığı vakaları bulunur (Prockop & Kivirikko, 1995). Genetiği değiştirilmiş kollajen mutasyonları olan fareler, çeşitli kollajenlerin işlevlerinin tanımlanması ve ilgili hastalıkların birçok yönünü incelenmesi için değerli olduklarını kanıtlamıştır. Ayrıca, bu tür fareler, kollajen mutasyonlarının sonuçlarını incelemek, kollajen mutasyonlarının neden olduğu ek hastalıkları tanımlamak ve hatta karşılık gelen insan hastalıkları için potansiyel tedavileri test etmek için son derece yararlıdır. Protein intiharı veya baskın negatif etki olarak bilinen fenomen nedeniyle, transgenik farelerde mutant bir kollajen geninin ekspresyonunun, karşılık gelen mutasyona sahip hastalarda görülene benzer bir hastalık fenotipi üretmesi beklenebilir. Birçok fare hatları da nakavt fareler üretmek için bir genin hedeflenen inaktivasyonu kullanılarak oluşturulmuştur. Bu yaklaşımlar, gerçekten çeşitli insan hastalıklarının fenotiplerini yeniden üreten bir dizi fare çizgisine yol açmıştır. Bazı durumlarda bu yaklaşım, insanlarda henüz bir kusur tespit edilmediği kollajen genlerindeki mutasyonların sonuçlarını incelemek için de kullanılmıştır (Myllyharju & Kivirikko, 2001).

Kollajen çapraz bağlanmasının temel mekanizması ve ana yolları 1960'ların sonlarında, 1970'lerin ve 1980'lerin başlarında tanımlanmıştır. Farklı doku tipleri farklı çapraz bağlanma kimyası modelleri gösterdi, ancak hepsi temelde özel lizin ve hidroksilizin yan zincirlerinden oluşan peptide bağlı aldehitlerin, kollajen alt birimlerinin fibrillere montajı sırasında lizil oksidaz (lar) ın etkisiyle reaksiyonlarına dayanıyordu. Tüm fibril oluşturan kollajenler (tip I, II, III, V ve XI) ve fibril ile ilişkili tip IX kollajen, doku yapısal bütünlüğü ve materyal işlevi sağlamak için bu çapraz bağlanma mekanizmasına dayanır. HPLC(Yüksek performanslı sıvı kromatografisi) ile nispeten standart peptid izolasyon yöntemleri ve ardından sekans analizi ile diğer ayırma teknikleri kullanılarak, ana doku kolajenlerindeki çapraz bağlanma etkileşim alanlarının çoğu üzerinde çalışılmıştır (Eyre, Paz, & Gallop, 1984; Hyry). Bununla birlikte, herhangi bir doku için çapraz bağlanan olgunlaşma ürünlerinin tam tamamlayıcısının tanımlanıp tanımlanmadığı ve çapraz bağlamadaki evrimleşmiş doku farklılıklarının tarz ve kapsamının işlevsel avantajlar sağlayıp sağlamadığı dâhil olmak

üzere, hala birçok soru ve bilgi eksikliği vardır. Eyre ve arkadaşları bunları ve kollajen biyolojisindeki ilgili soruları araştırmak için, yeni bir yaklaşım sağlamak için protein kütle spektrometrik yöntemlerini geliştirmişlerdir. Kütle spektrometrik analizinin, geleneksel yöntemlerle mümkün olmayan ve örnek boyutu açısından büyük bir hassasiyetle, kollajen'in post-translasyonel çapraz bağlanma kimyasının ayrıntılarını ortaya çıkarma potansiyelini gösteriyoruz. İlk başta yorumlanması ürkütücü olsa da, karmaşık çapraz bağlı peptitlerden gelen kütle spektrumları, yapısal tanımlama olarak yardımcı olabilecek karakteristik parçalanma modellerini ve yük zarfındaki farklılıkları (örneğin, esasen aynı bağlantılı peptitler için pirol ve piridinolin bazlı yapılar arasında) ortaya çıkarabilir. Glikosile edilmiş ve Glikosile edilmemiş çapraz bağları ayırt etme yeteneği, artan olgunlaşma yaşı ve kollajen yapısını etkileyen kalıtsal hastalıklarda doku tipleri arasında ilginç farklılıklar ortaya çıkarmaktadır. Burada verilen örneklerden, bu tür yaklaşımların çözülmemiş yapısal sorular hakkında ve daha önceki, dolaylı yöntemlere dayalı çapraz bağlantı oluşumunun önerilen mekanizmalarını ve moleküler bölgelerini yeniden değerlendirmede yeni anlayışlar sağlayabileceği açık olmalıdır. Sert hidroliz koşullarından kaçınarak ve peptidlerle çalışarak hafif proteolitik koşullar altında üretilen, artık yerel çapraz bağlantı ortamlarını ve yerel çeviri sonrası kimyayı daha güvenilir bir şekilde inceleyebiliyoruz. Bu, sonraki etkileşim kimyasını, tercih edilen zincir katılımlarını ve dolayısıyla polimerin özelliklerini değiştirmede yer alması muhtemel çapraz bağlama lokuslarındaki herhangi bir translasyon sonrası varyasyonu ortaya çıkarma potansiyeli sunmaktadır (Eyre, Weis, & Wu, 2008).

Flier vd. birçok hastalığın bağ dokusunun dejenerasyonunun, normal dokuların gücünün çoğunu sağlayan kollajenlerdeki genetik kusurun sonucu olduğu şeklindeki nispeten basit hipotezi test etmişlerdir. Hipotezi ilk önce kemik, kıkırdak ve diğer dokuların yaşamın erken dönemlerinde yaşamı tehdit edecek kadar kırılgan olabileceği nispeten nadir genetik hastalıklarda incelemişlerdir. Daha yakın zamanlarda, aynı hipotezin osteoporoz, osteoartrit ve aort anevrizmaları gibi daha yaygın hastalıklar için geçerli olup olmadığını görmeye çalışmışlardır. Tip I prokollajen genlerindeki mutasyonların neden olduğu osteoporozlu kadın alt grubu için erken teşhis, kemik kaybını azaltmak ve kemik kütlesini kırık eşiğinin üzerinde tutmak için östrojen ve diğer ajanlarla tedaviye başlamayı mümkün kılmalıdır. Tip III prokollajen genindeki mutasyonların neden olduğu aort anevrizması olan ailelerde, genetik analiz, aile

üyelerini mutasyonsuz ve dolayısıyla anevrizma riski artmadan belirleyecektir. Ayrıca, ultrasonografi ve ilgili tekniklerle sık aralıklarla anevrizma açısından incelenmesi gereken yüksek risk altındaki kişileri de belirleyecektir. Onlarca yıllık deneyim, bir aort anevrizmasının profilaktik olarak sentetik bir damarla değiştirilmesi durumunda yırtılan tedavi edilmemiş anevrizmalar için ölüm oranlarının yaklaşık yüzde 90'dan yüzde 10'un altına düştüğünü göstermektedir. Test ettikleri hipotezin, dokuların parçalandığı bazı yaygın hastalıklar için geçerli olduğuna artık ikna olmuş olsalar da, osteoporoz, osteoartrit veya aort anevrizması olan milyonlarca insanın ne kadarının kollajen genlerinde mutasyonlara sahip olduğu net değildir. Bu nedenle, teknolojideki ilerlemelerin yakında bu tür analizleri genel kullanıma sunacağını beklemek makul görünmektedir (Prockop, 1992).

Birçok musküler distrofinin (ilerleyici güçsüzlüğe ve kas kütlesi kaybına neden olan bir grup hastalık) ve miyopatinin (kas hastalığı) altında yatan moleküler temelin keşfi, iskelet kasında hücre iskeleti ile hücre dışı matris arasındaki bağlantının önemini vurguladı. Hücre dışı matriks ve bazal lamina moleküllerinin iskelet kası stabilitesi, rejenerasyonu ve kas hücresi matriks yapışması için kritik olduğu bilinmektedir. Kollajen VI, kas hücreleri etrafında bazal membran ile yakın ilişki içinde bir mikrofibriler ağ oluşturan ve diğer birçok matris bileşeniyle etkileşime giren, her yerde bulunan hücre dışı bir matris proteinidir. Yapılan çalışmada Lampe vd. Bethlem miyopatisinin (BM) ve Ullrich konjenital kas distrofisinin (UCMD) klinik fenotiplerini, bu bozukluklardaki kas patolojisini, bunların neden olduğu kollajen VI 'nın rolünü tartışmışlardır. Kollajen VI ile ilgili kas distrofilerinin genetiği karmaşıktır. İki farklı klinik fenotip olan Bethlem miyopatisi ve Ullrich konjenital musküler distrofinin bağımsız tanımını, bu bozuklukların her ikisinin de kollajen VI'daki mutasyonlardan kaynaklandığının keşfi izlemiştir. Farklı mutasyonların, protein toplanması, salgılanması ve bunun işleyen bir hücre dışı ağ oluşturma yeteneği üzerinde çeşitli etkilere sahip olduğu gösterilmiştir. Bu çok sayıda olası mutasyonel mekanizma ve kalıtım biçimi, hastaların ve ailelerinin genetik danışmanlığı ve kişiselleştirilmiş tıp uygulamaları için önemli sonuçlara sahiptir. Şu anda hiçbir iyileştirici tedavi yaklaşımı klinik bir düzeye ulaşmamıştır, ancak destekleyici tedavinin uzun ömürlülüğü ve yaşam kalitesini önemli ölçüde artırabileceği düşünülmektedir (Lampe & Bushby, 2005).

Kırılgan kemik hastalığı olarak da bilinen Osteogenezis imperfekta (OI), esas olarak kırılmaya yatkınlıkla karakterize edilen, klinik ve genetik olarak heterojen bir hastalıktır. OI genellikle tip I kollajen genleri olan, COL1A1 ve COL1A2'deki mutasyonlardan kaynaklansa da, genotip ve fenotip arasındaki ilişki henüz tam olarak anlaşılmamıştır. Genotip-fenotip analizleri için ek veri sağlamak ve ölümcül OI formlarına sahip denekler arasında tip I kollajen genlerindeki mutasyonların oranını belirlemek için, COL1A1 ve COL1A2'nin kodlama ve ekson-yan bölgelerini OI tip II, hastalığın perinatal ölümcül formunu 63 denekten oluşan bir kohortta sıralamışlardır. Tip I kollajen ’de 61 farklı heterozigot mutasyon tespit etmişlerdir. Kollajen mutasyonları olmayan üç örnekte, CRTAP (Kıkırdak İlişkili protein) ve LEPRE1'de(Lösin proteinle zenginleştirilmiş proteglikan) inaktive edici mutasyonlar bulmuşlar, bu resesif mutasyonların OI tip II'de % 5 sıklığını düşündürmüştür. Kollajen alfa-I zincirlerinin üçlü sarmal alanı içindeki glisin ikamelerinin sonucunu tahmin eden bir hesaplama modeli, ölümcüllüğü% 90 doğrulukla tahmin etmiştir. Sonuçlar, genotipi fenotiple ilişkilendiren mevcut modelleri değerlendirmek ve iyileştirmek için veri sağlayarak ve OI ile ilişkili genlerdeki mutasyonların göreceli sıklığının tarafsız bir göstergesini sağlayarak OI etiyolojisinin anlaşılmasına katkıda bulunmuştur (Bodian et al., 2009).

Genom dizileme projelerinin temel amaçlarından biri, hastalığın genetiğinin daha iyi anlaşılmasını sağlamak ve hastalıkla ilişkili genlerin tanımlanmasına yardımcı olmaktır (Green, 1997; Hogenesch et al., 2001; Lander et al., 2001). Bununla birlikte, hastalık hastalarında bulunan genetik varyasyonu içeren birçok insan mutasyon veri tabanı zaten mevcuttur ve yeni veri tabanları ve veri tabanı girişleri hızla birikmektedir (Chen et al., 2001; Cotton, 2000). Bu iki tür bilginin eşzamanlı analizi, hastalıkla ilişkili insan genetik varyasyonunun kendine özgü özelliklerini tanımlamak için benzersiz fırsatlar sağlar ve bu da, mutasyonlar ve hastalık fenotiplerinin gelişimi arasındaki ilişkinin daha iyi anlaşılmasına yol açar. Evrim tarihi boyunca bu tür pozisyonlarda ortaya çıkan varyasyonlar, açıkça güçlü saflaştırma seleksiyonu altında olmuştur ve popülasyonlardan elimine edilmiştir, bu da değişmez pozisyonlardaki mevcut amino asit kalıntılarının uygun gen fonksiyonu için kritik olduğunu düşündürmektedir. Bu nedenle, spesifikler arası hizalamalardan elde edilen bilgiler, insanlarda mutasyona uğradığında

hastalık oluşturması muhtemel olan gen ürünlerindeki amino asit kalıntılarını gösterebilir. Benzer şekilde, protein dizilerindeki bazı pozisyonlar, türler arasında farklılık gösterir ve bu tür değişken alanlar, daha az şiddetli seçici kısıtlamalar altında olan pozisyonları gösterebilir. Kumar ve arkadaşları, araştırmacılar tarafından, gözlemlenen bir amino asit değişikliğinin insanlarda hastalık üretme potansiyelini belirtmek için sıklıkla kullanıldığından, türler arası dizi hizalamalarının insanlarda gözlenen zararlı mutasyonların, ortak özellikleri ortaya koyma derecesini doğrudan değerlendirmek için bir çalışma yürütmüşlerdir. Hastalık mutasyon verilerini ve hastalıkla ilişkili yedi genden homolog gen dizilerini kullanarak üç tür analiz gerçekleştirmişlerdir. İlk olarak, hastalık mutasyonlarının yaygınlığı ile diğer türlerdeki karşılık gelen amino asit bölgelerinin metazoanların evrimsel tarihi boyunca ne ölçüde korunduğu arasındaki ilişkiyi belirlemişlerdir. İkinci olarak, hastalık hastalarında belirli bir amino asit değişikliğinin sıklığını, spesifikler arası karşılaştırmalardan elde edilen frekanslarla karşılaştırmışlardır. Son olarak, türler ve hasta olmayan insanlar arasında görülen amino asit değişikliklerinin kimyasal özellik farklılıklarını, hastalık hastalarında gözlemlenenlerle karşılaştırmışlardır. İnsanlarda hastalıkla ilişkili amino asit değişikliklerinin, türler arasında görülen insan polimorfik değişikliklerinden veya varyasyonundan çeşitli şekillerde farklı olduğunu göstermişlerdir. Bu yaygın olmayan hastalıkla ilişkili amino asit değişikliklerinin kimyasal farklılıkları, insanlarda yaygın olarak karşılaşılan veya evrim tarihi boyunca doğal seçilimin izin verdiği polimorfik amino asit varyasyonundan daha radikaldir. Türler arasında gözlemlenen türler arası hastalıkla ilişkili genetik varyasyon ve varyasyonun karşılaştırılması, birkaç ortak eğilimi ortaya çıkarmakta ve karşılaştırmalı moleküler analizlerin, insanlarda bulunan zararlı genetik varyasyona doğrudan uygulanabilir olduğunu göstermektedir (Miller & Kumar, 2001).

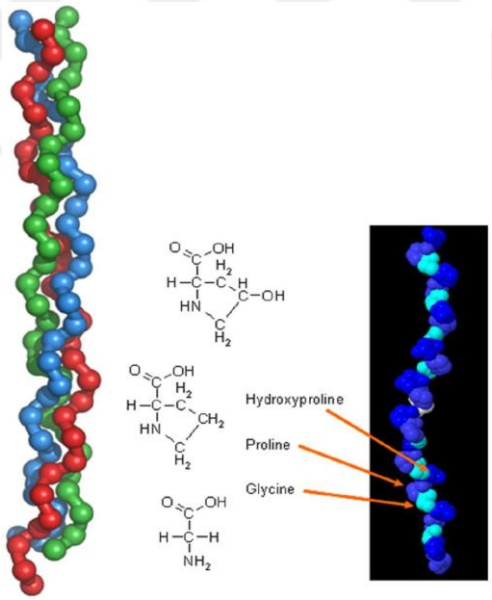
Amino asit sekansında bir değişiklikle sonuçlanan genetik varyasyonlar, stabilite, hidrojen bağ ağı, konformasyonel dinamikler, aktivite ve proteinlerin diğer birçok fizyolojik olarak önemli özelliği üzerinde dramatik bir etkiye sahip olabilir. Yanlış anlamlı mutasyonlar olarak adlandırılan bir protein sekansındaki sadece bir kalıntının ikameleri, birçok patolojik durumla ilişkili olabilir ve hastalığa ve ilaç tedavisine yatkınlığı etkileyebilir.

Yanlış anlam mutasyonlarının makul etkileri, makromoleküler stabiliteyi etkilemekten makromoleküler etkileşimleri ve hücresel lokalizasyonu bozmaya kadar uzanır. Stefl vd. yanlış anlam mutasyonları ile hastalıklar arasındaki ilişkiyi gösteren bireysel vakaları ve genom çapında çalışmaları gözden geçirmişlerdir. Ek olarak, hastalığın kökenini anlamak için mutasyonların etkilerinin moleküler mekanizmalarının ortaya çıkarılması gerektiğini vurgulamışlardır. Son olarak, mutasyonların protein stabilitesi, hidrojen bağ ağı, pH bağımlılığı, konformasyonel dinamikler ve protein işlevi üzerindeki etkilerini tahmin eden mevcut son teknoloji metodolojileri rapor etmişlerdir. Mutasyonlar ve kronik hastalığa yatkınlık arasındaki bağlantıyı belirleme çabası içinde, hastalığa neden olan nadir ve yaygın mutasyonlar için literatürden birçok veri toplamışlardır. Adenom hastaları ve kontrol gruplarının varyant sıklıkları karşılaştırıldığında, hastalarda varyant sıklığının kontrollere göre daha yüksek olduğu gösterilmiştir. Bu, hastalığa yatkınlığın kısmen birçok nadir varyantın etkilerine bağlı olabileceğini öne sürdü. Bir protein dizisindeki sadece bir kalıntının ikamesinin Alzheimer, Parkinson ve Creutzfeldt-Jakob hastalıkları gibi bir dizi patolojik durumla ilişkili olabileceği bilinmektedir (Capriotti, Calabrese, & Casadio, 2006). İnsan yanlış anlam mutasyonlarının çeşitli biyofiziksel özellikler, kararlılık, hidrojen bağ ağı, dinamikler ve aktivite üzerindeki etkileri bu çalışmada gözden geçirilmiştir. Mutasyonların sıklıkla birkaç biyofiziksel özelliği aynı anda etkileyebileceği ve hastalıklara neden olabileceği veya olmayabileceği gösterilmiştir. Mutasyonların zarar verici etkilerinin tahminleri, artmış protein aktivitesi, daha yüksek stabilite veya bağlanma afinitesinin hücre için zorunlu olarak avantajlı olmadığı ve hastalığa neden olabileceği gözlemleriyle daha da karmaşık hale gelir (Stefl, Nishi, Petukh, Panchenko, & Alexov, 2013).

3. MATERYAL VE YÖNTEM

3.1. Kollajen

Kollajen hareket sisteminin yapı taşlarını, özellikle kemik, kıkırdak, lif ve eklemleri oluşturan proteindir. Kollajenin ana molekülü tropokollajen'dir. Tropokollajenler de hücrenin içinde üretilen prokollajenlerden oluşur. Kollajenler, insanlarda 28 üyeden oluşan geniş bir aileye ait karmaşık protein yapılarıdır. Genellikle alfa zincirleri olarak bilinen üç farklı polipeptit zinciri ile üçlü bir sarmal yapı oluştururlar. Ayrıca, bu üçlü sarmalın kendine özgü bir "Gly-Xaa-Yaa" dizisine sahip olduğu bulunmuştur. Her dizide ki glisin (Gly) varlığı, en küçük amino asit olma özelliği nedeniyle sarmal yapının stabilitesini açıklar. Xaa ve Yaa herhangi bir amino asit olabilir ancak çoğunlukla prolindir. Bu nedenle, kollajenin glisin ve prolin bakımından zengin bir protein olduğu bilinmektedir (Cen, Liu, Cui, Zhang, & Cao, 2008). Şekil 3.1 'de kollajenin yapısı gösterilmiştir.



Şekil 3.1 Kollajenin yapısı (Ileana & Francois, 2014).

İnsan vücudunda kollajen en bol bulunan proteinlerden biridir. Yapısına göre sınıflandırılır ve yaygın olan tipleri fibril oluşturan kollajenler olarak adlandırılır; bunlar arasında tip I, tip II, tip III, tip V ve tip XI en çok vücut dokusunda bulunur. Fibril oluşturan kollajenlerin insan vücudunda bulunduğu yerler;

Tip I: En bol bulunan kollajen türüdür. Genellikle deri, kemik, dermis, bağ, kornea ve tendon gibi bağ dokusunda bulunur.

Tip II: Kıkırdakta, embriyonik epitelyal mezenkimal geçişlerde, kornea epitelinde, notokordda, intervertebral disklerin nükleus pulposusunda ve vitröz gövdede bulunur.

Tip III: Genellikle ciltte, damar duvarında, bağırsaklarda ve retiküler liflerde genellikle Tip I ile birlikte bulunur.

Tip V: Çoğunlukla kemik, akciğer, kornea, fetal membranlarda bulunan tip I kollajenle bulunur.

Tip XI: Genellikle tip XI ve tip II kollajenler vitröz vücutta, intervertebral diskte ve kıkırdakta bir arada bulunur (Gelse, Pöschl, & Aigner, 2003).

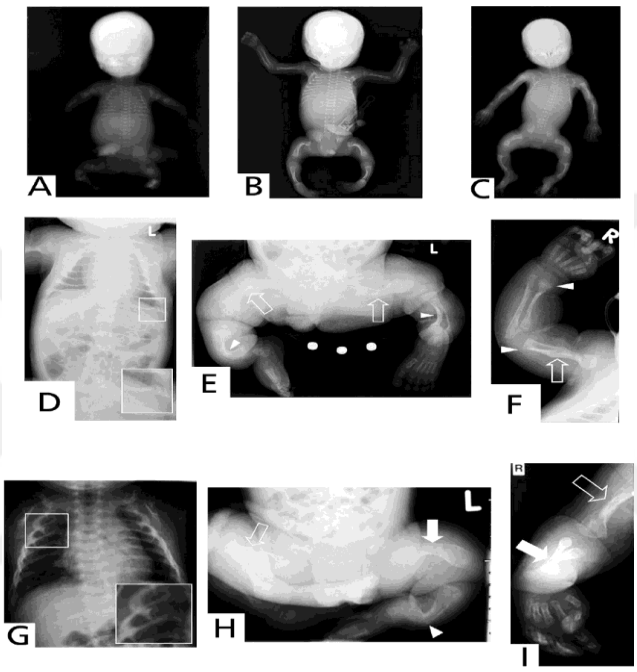
Tablo 3.1 'de kollajen tipleri ve sınıflandırması verilmiştir.

Çizelge 3.1 Kollajen tipleri ve sınıflandırması

Fibril oluşturan kollajenler	Type I, II, III, V, XI
FACIT (Kesikli üçlü sarmallı fibril ile ilişkili kollajenler)	Type IX, XII, XIV, XIX, XX, XXI
Ağ oluşturan kollajenler	Type IV, VIII, X
Multipleksin kollajenler	Type XV, XVI, XVIII
Transmembran kollajenler	Type XIII, XVII
Mikrofibriler kollajenler	Type VI
Fibril oluşturan kollajenleri sabitleme	Type VII

3.2. Kırılgan Kemik Hastalığı (Osteogenezis Imperfecta)

Osteogenezis imperfekta (Ol), artmış kemik kırılganlığı, düşük kemik kütlesi ve diğer bağ dokusu belirtilerinin genetik bir bozukluğudur. Osteogenez kusurunun klinik tanısı olan hastaların çoğunda, kollajen tip 1'in alfa zincirlerini (COL1A1 ve COL1A2) kodlayan iki genden birinde bir mutasyon görülür (Rauch & Glorieux, 2004). Hastalığın şiddeti hafif (OI tip I) ile perinatal ölümcül (OI tip II) arasında değişir ve hem otozomal dominant hem de resesif kalıtım paternleri sergiler (Bodian et al., 2008).



Şekil 3.2 Kollajen tip I mutasyonları nedeniyle Osteogenez imperfekta (OI) olan hastaların radyografileri (van Dijk et al., 2011).

3.3. PyCharm programı ve Python programlama dili

Proje de kollajen alfa-1 zincirlerindeki aminoasit sayıları Python programlama diliyle yazılmış kodlar kullanılarak PyCharm programı aracılığıyla hesaplanmıştır. PyCharm, özellikle Python dili için bilgisayar programlamada kullanılan entegre bir geliştirme ortamıdır. Python, nesne yönelimli, yorumlamalı, birimsel (modüler) ve etkileşimli yüksek seviyeli bir programlama dilidir. Şekil 3.3 ’de PyCharm programında yazılan kodların ekran görüntüsü ve Şekil 3.4 ’de sonuç kodları gösterilmiştir.

```
collagen_alphal_chain=("""MFSFVDLRLLLLLAATALLTHGQEEGQVEGQDEDIPPITCVQNGLRYHDRDVWKPEPCRICVCDNGKVLCDDVICDETKNCPGAEVPEGECCPV
glycine_amino_acid=['G']
alanine_amino_acid=['A']
arginine_amino_acid=['R']
asparagine_amino_acid=['N']
asparticacid_amino_asid=['D']
cysteine_amino_acid=['C']
glutamine_amino_acid=['Q']
glutamicacid_amino_acid=['E']
histidine_amino_acid=['H']
isoleucine_amino_acid=['I']
leucine_amino_acid=['L']
lysine_amino_acid=['K']
methionine_amino_acid=['M']
phenylalanine_amino_acid=['F']
proline_amino_acid=['P']
serine_amino_acid=['S']
threonine_amino_acid=['T']
tryptophan_amino_acid=['W']
tyrosine_amino_acid=['Y']
valine_amino_acid=['V']

print("total alanine amino acid number:",len(alanine_amino_acid_number))
print("total arginine amino acid number:",len( arginine_amino_acid_number))
print("total asparagine amino acid number:",len( asparagine_amino_acid_number))
print("total asparticacid amino acid number:",len(asparticacid_amino_asid_number))
print("total cysteine amino acid number:",len( cysteine_amino_acid_number))
print("total glutamicacid amino acid number:",len( glutamicacid_amino_acid_number))
print("total glutamine amino acid number:",len( glutamine_amino_acid_number))
print("total histidine amino acid number:",len( histidine_amino_acid_number))
print("total isoleucine amino acid number:",len( isoleucine_amino_acid_number))
print("total leucine amino acid number:",len( leucine_amino_acid_number))
print("total lysine amino acid number:",len( lysine_amino_acid_number))
print("total methionine amino acid number:",len( methionine_amino_acid_number))
print("total phenylalanine amino acid number:",len( phenylalanine_amino_acid_number))
print("total proline amino acid number:",len( proline_amino_acid_number))
print("total serine amino acid number:",len( serine_amino_acid_number))
print("total threonine amino acid number:",len( threonine_amino_acid_number))
print("total tryptophan amino acid number:",len( tryptophan_amino_acid_number))
print("total tyrosine amino acid number:",len( tyrosine_amino_acid_number))
print("total valine amino acid number:",len( valine_amino_acid_number))
```

Şekil 3.3 PyCharm programında yazılan kodların ekran görüntüsü

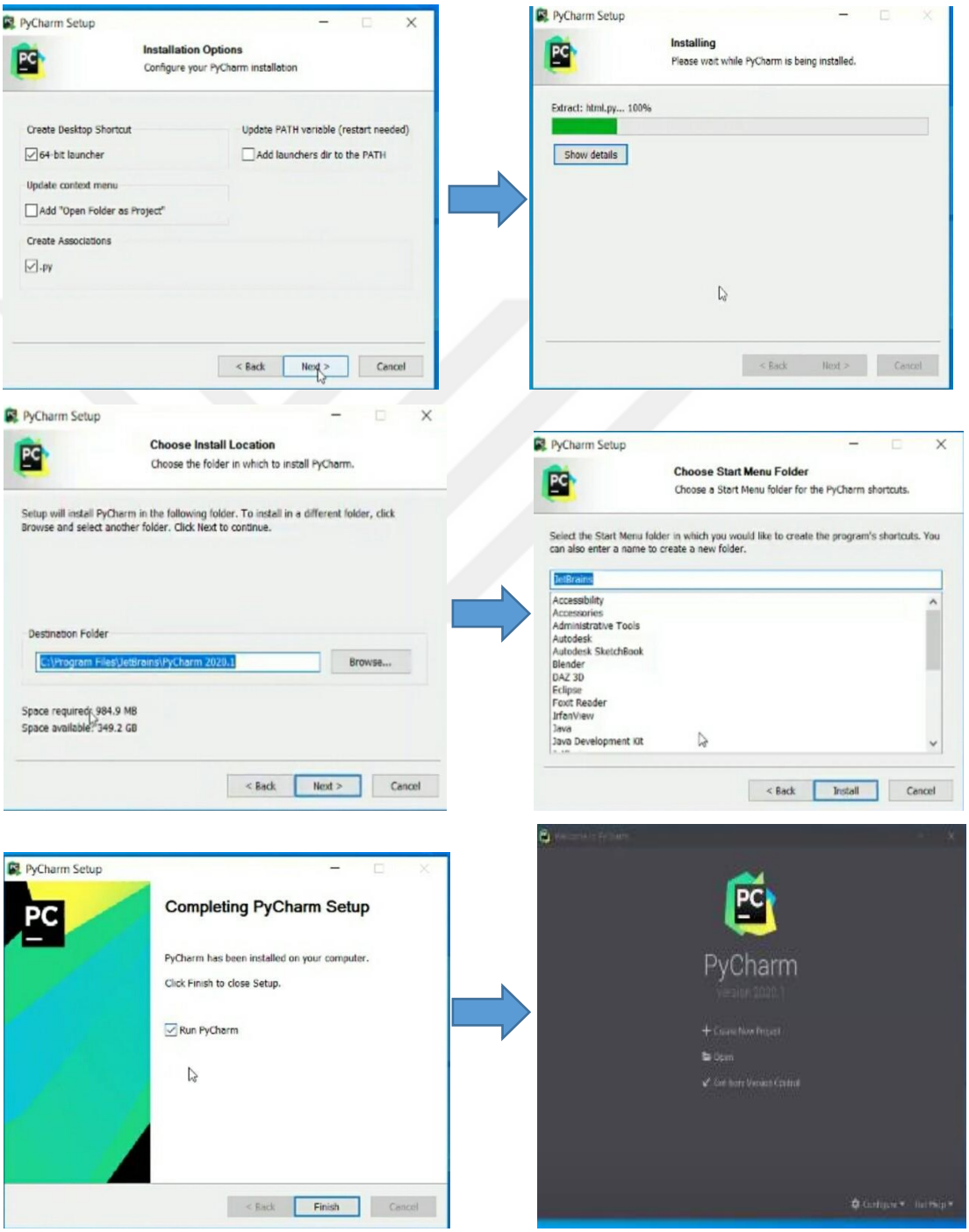
```
total glycine amino acid number: 391
total alanine amino acid number: 139
total arginine amino acid number: 71
total asparagine amino acid number: 28
total asparticacid amino acid number: 66
total cysteine amino acid number: 18
total glutamicacid amino acid number: 75
total glutamine amino acid number: 49
total histidine amino acid number: 9
total isoleucine amino acid number: 24
total leucine amino acid number: 48
total lysine amino acid number: 57
total methionine amino acid number: 13
total phenylalanine amino acid number: 27
total proline amino acid number: 278
total serine amino acid number: 60
total threonine amino acid number: 45
total tryptophan amino acid number: 6
total tyrosine amino acid number: 13
total valine amino acid number: 47
```

```
toplam glisin amino asit kütlesi: 29250
toplam alanine amino asit kütlesi: 12371
toplam asparagine amino asit kütlesi: 3696
toplam arginine amino asit kütlesi: 12354
toplam asparticacid amino asit kütlesi: 8778
toplam cysteine amino asit kütlesi: 2178
toplam glutamine amino asit kütlesi: 7154
toplam glutamicacid amino asit kütlesi: 11025
toplam histidine amino asit kütlesi: 1395
toplam isoleucine amino asit kütlesi: 3144
toplam leucine amino asit kütlesi: 6288
toplam lycine amino asit kütlesi: 8322
toplam methionine amino asit kütlesi: 1937
toplam phenylalanine amino asit kütlesi: 4455
toplam proline amino asit kütlesi: 31970
toplam serine amino asit kütlesi: 6405
toplam threonine amino asit kütlesi: 5355
toplam tryptophan amino asit kütlesi: 1224
toplam tyrosine amino asit kütlesi: 2353
toplam valine amino asit kütlesi: 5499
toplam moleküler ağırlık: 165153
```

Şekil 3.4 PyCharm programında yazılan sonuç kodlarının ekran görüntüsü

3.4. PyCharm Program Kurulumu

Projede kullanılan PyCharm programının bilgisayara yüklemek için jetbrains (https://www.jetbrains.com/pycharm/download/#section=windows) web sitesiden program indirilmiştir. Şekil 3.5 ’de kurulum aşamaları gösterilmiştir.

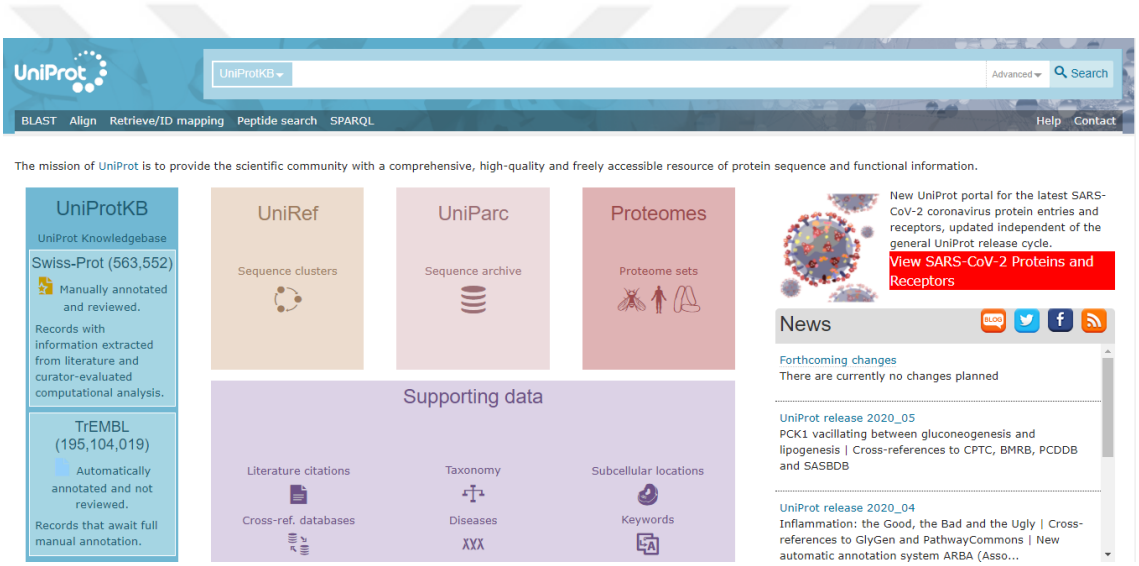


Şekil 3.5 PyCharm programı kurulum aşamaları

Program belirtilen web sitesinden bilgisayara yüklenir. Şekil 3.5 'de gösterilen aşamalar izlenerek bilgisayara kurulumu gerçekleştirilir. Kurulum tamamlandıktan sonra açılan PyCharm penceresinden yeni bir proje dosyası açılarak gereken kodlar yazılır.

3.5. UniProtKB / SWISS-PROT

UniProt serbestçe erişilebilir bir protein dizisi ve fonksiyonel bilgi veri tabanıdır, birçok giriş genom dizileme projelerinden türetilmiştir. Araştırma literatüründen türetilen proteinlerin biyolojik fonksiyonları hakkında büyük miktarda bilgi içerir. Şekil 3.6 'da UniProt veri tabanının ana sayfa ekran görüntüsü gösterilmektedir. Veri tabanına erişim için http://www.uniprot.org/ adresi kullanılır.

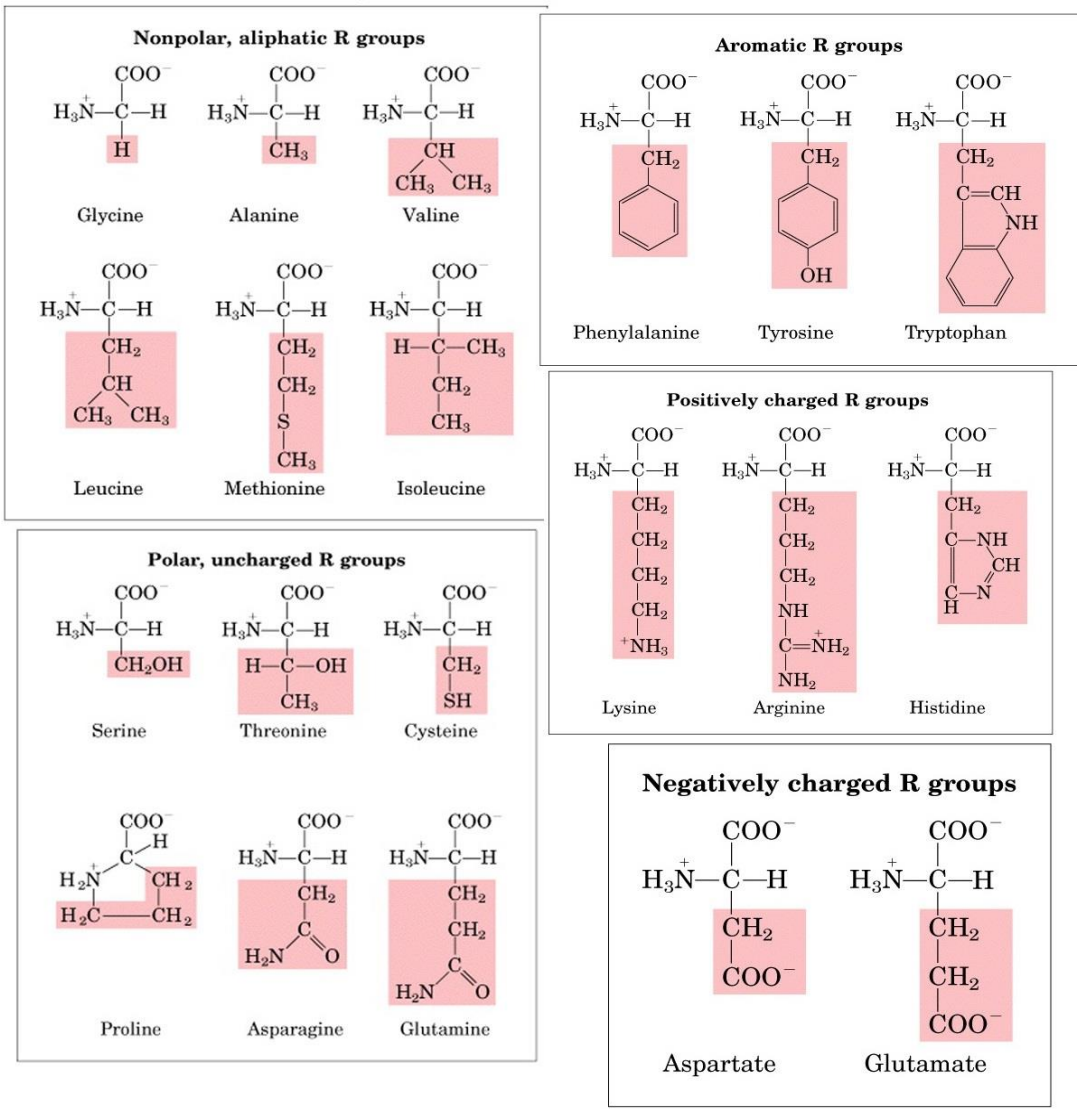


Şekil 3.6 UniProt veri tabanı

3.6. Amino asitler

Amino asitler, proteinleri oluşturan temel yapı taşlarıdır. Amino asitler, amin (–NH2) ve karboksil (–COOH) fonksiyonel grupları ile her bir amino aside özgü bir yan zincir (R grubu) içeren organik bileşiklerdir (Arjadi, Nauta, Chowdhary, & Bockting, 2015; Bahadıroğlu, Büyükberber, Gürbüz, Nakkaş, & Şahin). Bir amino asidin temel unsurları karbon (C), hidrojen (H), oksijen (O) ve nitrojendir (N), ancak bazı amino asitlerin yan zincirlerinde başka elementler de bulunur. Doğal olarak oluşan yaklaşık 500 amino asit bilinmektedir (genetik kodda sadece 20 tane görünmesine rağmen) ve birçok şekilde sınıflandırılabilir (Wagner & Musso, 1983). Aminoasitler, polarite, ph

seviyesi ve yan zincir grubuna göre sınıflandırılabilirler. 20 çeşit amino asit ’in yapısı sınıflandırılmış olarak Şekil 3.7 ’de verilmiştir.



Şekil 3.7 Aminoasitlerin yapısı ve sınıflandırılması (Aryal, August 9, 2018).

3.7. İzoelektrik Noktası (pı)

İzoelektrik nokta (pI, pH (I), IEP), bir molekülün hiçbir net elektrik yükü taşımadığı veya istatistiksel ortalamada elektriksel olarak nötr olduğu pH'dır (Fforde, 2019). Birçok molekül, hem pozitif hem de negatif yükler içeren zwitteryonlardır. Molekül üzerindeki net yük, çevreleyen ortamın pH'ı tarafından yönetilir. PI, molekülün net elektrik yükü taşımadığı ve dolayısıyla negatif ve pozitif yüklerin eşit olduğu pH değeridir (Gargaud, Amils, & Cleaves, 2011).

Çizelge 3.2 Aminoasitlerin izoelektrik noktaları (Idrees, Mohammad, Karodia, & Rahman, 2020).

Amino asitler	**pı (İzoelektrik Noktası)**
Glycine(G)	5,97
Alanine(A)	6,01
Arginine(R)	10,76
Asparagine(N)	5,41
Aspartic acid(D)	2,77
Cysteine(C)	5,07
Glutamic acid(E)	3,22
Glutamine(Q)	5,65
Histidine(H)	7,59
Isoleucine(I)	6,02
Leucine(L)	5,98
Lysine(K)	9,74
Methionine(M)	5,74
Phenylalanine(F)	5,48
Proline(P)	6,48
Serine(S)	5,68
Threonine(T)	5,87
Trptophan(W)	5,89
Tyrosine(Y)	5,66
Valine(V)	5,97

3.8. Hidropati İndeksi

Hidropati (hidrofobiklik vs hidrofiliklik veya lipofilik vs lipofiliklik) genellikle -7.5 ila 3.1 arasında sayılar ile karakterize edilirken, hidrofobiklik yan zincirlerin sudan ne kadar güçlü bir şekilde itildiğinin bir ölçüsüdür. Sayı ne kadar pozitifse, amino asit

artığı o kadar sulu bir ortamda olmama eğiliminde olacaktır (Biro, 2006). Aminoasitlerin hidropati indeksleri Şekil 3. 9'da verilmiştir.

Çizelge 3.3 Aminoasitlerin hidropati indeks değerleri (Mitaku, Hirokawa, & Tsuji, 2002)

Amino Acid	**Hydropathy Index**
Lysine(K)	-3.9
Arginine(R)	-4.5
Histidine(H)	-3.2
Glutamic Acid(E)	-3.5
Glutamine(Q)	-3.5
Aspartic Acid(D)	-3.5
Asparagine(N)	-3.5
Trptophan(W)	-0.9
Tyrosine(Y)	-1.3
Serine(S)	-0.8
Threonine(T)	-0.7
Proline(P)	-1.6
Glycine(G)	-0.4
Alanine(A)	1.8
Methionine(M)	1.9
Cysteine(C)	2.5
Phenylalanine(A)	2.8
Leucine(L)	3.8
Valine(V)	4.2
Isoleucine(I)	4.5

.

4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA

4.1. Elde edilen alfa-1 zincirlerinin PyCharm programında analizi

Projede vücudumuzda en bol bulunan Tip 1 kollajenin alfa-1 zinciri kullanılarak analiz yapılmıştır. Öncelikle PyCharm programında analiz yapmadan kollajen kaynaklı hastalıklar araştırılmıştır. UniProt veri tabanında Tip 1 kollajen'in alfa-1 zincirinde kollajen kaynaklı hastalıkların mutasyona uğramış aminoasitleri hakkında bilgi verilmektedir. İncelenen çalışmalarda Osteogenez imperfekta tip II hastalığının, Tip 1 kollajen alfa-1 zincirinde daha fazla mutasyona neden olduğu görülmüştür. Bunun üzerine Uniprot veri tabanında OI Tip II hastalığı üzerine makaleler incelenmiştir. Bodian ve arkadaşlarının yaptığı çalışma da 63 farklı kişide tespit edilen mutasyonlar kullanılmıştır (Bodian et al., 2009). Bu çalışmadaki mutasyona uğramış bütün aminoasitler tek bir kollajen alfa-1 zincirinde birleştirilmiştir. Elde edilen mutasyona uğramış zincir ile veri tabanından FASTA formatında elde edilen kollajen proteinin alfa-1 zinciri PyCharm programında tanımlanmıştır. Her iki zincirdeki aminoasitlerin sayısını hesaplayan Pyhton kodu yazılmıştır. En çok mutasyona uğrayan amino asidi tespit etmek için karşılaştırma yapılmıştır. Farklı kişilerden elde edilen mutasyonlarda en çok tekrarlanan aminoasit değişimleri tespit edilmiş, bu mutasyonların alfa-I zincirinde meydana getirdiği değişimler moleküler boyutta incelenmiştir. Fiziksel özellikler, moleküler ağırlık, hidropati indeks, izoelektrik noktası ve pH gibi kriterlere bakılmış ve bu kriterlerdeki değişimlerin moleküller arası etkileşimlere olan katkısı ve hastalık oluşum mekanizması üzerine etkisi yorumlanmıştır.

4.2. Alfa-I zincirinde birleştirilen mutasyonlardan elde edilen zincir ile normal alfa-I zincirinin amino asit dizilimlerinin karşılaştırılması

Farklı kişilerden elde edilen ve kollajen alfa-I zinciri üzerinde birleştirilen mutasyonlu zincir ile normal kollajen alfa-I zincirindeki aminoasit sayıları PyCharm programında Python programlama diliyle yazılan kodlarla hesaplanmıştır. Karşılaştırma sonucunda elde edilen aminoasit sayıları Tablo 4.1 'de gösterilmiştir.

Çizelge 4.1 Aminoasit sayılarının karşılaştırılması

Toplam aminoasit sayıları	**COL-A1-ZİNCİRİ**	**COL-A1-OL2-TÜM-MUTASYONLARI İÇEREN ZİNCİR**
Toplam glisin amino asit sayısı	391	366
Toplam alanin amino asit sayısı	139	139
Toplam arjinin amino asit sayısı	71	77
Toplam asparajin amino asit sayısı	28	28
Toplam aspartik asit amino asit sayısı	66	66
Toplam sistein amino asit sayısı	18	20
Toplam glutamik asit amino asit sayısı	75	74
Toplam glutamin amino asit sayısı	49	49
Toplam histidin amino asit sayısı	9	10
Toplam izolösin amino asit sayısı	24	24
Toplam lösin amino asit sayısı	48	48
Toplam lizin amino asit sayısı	57	57
Toplam metiyonin amino asit sayısı	13	13
Toplam fenilalanin amino asit sayısı	27	27
Toplam prolin amino asit sayısı	278	277
Toplam serin amino asit sayısı	60	65
Toplam treonin amino asit sayısı	45	47
Toplam triptofan amino asit sayısı	6	6
Toplam tirozin amino asit sayısı	13	13
Toplam valin amino asit sayısı	47	52

Tablo 4.1 incelendiğinde en fazla değişimin kollajenin yapısında en fazla bulunan ve kollajen sarmal yapısının stabilitesini sağlayan (Nassa et al., 2012) glisin(G) aminoasit sayısında olduğu görülmüştür. Glisinin en bariz özelliği molekül kütlesi ve hacmi en küçük amino asit olmasıdır. Glisin sayısındaki değişimin fazla olması stabiliteyi etkileyeceğinden Osteogenez İmperfekta hastalığında oluşan kırılgan kemik yapısının glisin yerine farklı aminoasitlerin gelmesi olarak düşünülmektedir.

4.3. Tekrarlanan mutasyonların incelenmesi

İncelenen çalışmada ki mutasyonlara bakıldığında en çok tekrar eden mutasyonlar şu şekilde tespit edilmiştir: Bunlar glisin(G) yerine aspartik asit(D), glisin(G) yerine arjinin(R), glisin(G) yerine serine(S) mutasyonlarıdır. Bu mutasyonlar arasından bazıları seçilerek ayrı ayrı yeni alfa-I zincirleri oluşturuldu. 1464 aminoasitten

oluşan alfa-I zincirinde, farklı bireylerden elde edilmiş olan 353. sıradaki glisinin aspartik asit'e, 614. sıradaki glisinin arjinine, 884. sıradaki glisinin serine olan mutasyonları seçildi. Mutasyon içeren alfa-I zincirleri ile normal alfa-I zinciri PyCharm programında FASTA formatında tanımlanmış Python programlama diliyle yazılan kodlarla alfa-I zincirlerinin ortalama izoelektrik (pı) değerleri ve toplam molekül ağırlıkları hesaplandı. Hesaplama sonucu Tablo 4.2 'de gösterilmektedir.

Çizelge 4.2 Tekrarlanan mutasyonların karşılaştırılması

Kollajen alfa-1 zinciri	**COL1A1**	**COL1A1-353-G-D**	**COL1A1-614-G-R**	**COL1A1-884-G-S**
Ortalama pı Değerleri	6.114	6.111	6.117	6.113
Toplam molekül ağırlığı	165300	165181	165222	165153

Tablo 4.2 incelendiğinde mutasyonlar sonucu alfa-I zincirlerinin ortalama izoelektrik değerlerinde değişim olmuş ve toplam molekül ağırlığı azalmıştır. Alfa-I zincirinde meydana gelen tek bir mutasyon sonucunda değişen değerlere bakıldığında bu değişimlerin Osteogenez imperfekta hastalığına etkisi olduğu söylenebilir.

4.4. Glisin ile glisin yerine ikame edilen aminoasitlerin özelliklerinin karşılaştırılması

Mutasyonlar sonucunda glisin yerine gelen aminoasitlerin değişen özelliklerini incelemek amacıyla Nassa ve arkadaşlarının çalışmalarından esinlenilerek aminoasitlerin fiziksel özelliklerini içeren bir tablo oluşturuldu. Tablo 4.3'de amino asitlerin bazı özellikleri gösterilmiştir.

Çizelge 4.3 Aminoasitlerin özellikleri

Amino asitler	**Molekül Ağırlığı**	**pı (İzoelektrik Noktası)**	**Hidropati İndeksi**	**Yük Durumu**	**Asidik ve Bazik özellikleri**
Glisin(G)	75.07	5,97	- 0.4	apolar(yüksüz)	nötr
Alanin(A)	89.10	6,01	1.8	apolar(yüksüz)	nötr
Arjinin(R)	174.20	10,76	-4.5	pozitif yüklü	kuvvetli bazik
Asparajin(N)	132.12	5,41	-3.5	polar(yüklü)	nötr
Aspartik asit(D)	133.11	2,77	-3.5	negatif yüklü	asidik
Sistein(C)	121.16	5,07	2.5	apolar(yüksüz)	nötr
Glutamik asit(E)	147.13	3,22	-3.5	negatif yüklü	asdik
Glutamin(Q)	146.15	5,65	-3.5	polar(yüklü)	nötr
Histidin(H)	155.16	7,59	-3.2	pozitif yüklü	hafif bazik
İzolösin(I)	131.18	6,02	4.5	apolar(yüksüz)	nötr
Lösin(L)	131.18	5,98	3.8	apolar(yüksüz)	nötr
Lizin(K)	146.19	9,74	-3.9	pozitif yüklü	bazik
Metiyonin(M)	149.21	5,74	1.9	apolar(yüksüz)	nötr
Fenilalanin(F)	165.19	5,48	2.8	apolar(yüksüz)	nötr
Prolin(P)	115.13	6,48	1.6	apolar(yüksüz)	nötr
Serin(S)	105.09	5,68	-0.8	polar(yüklü)	nötr
Treonin(T)	119.12	5,87	-0.7	polar(yüklü)	nötr
Triptofan(W)	204.23	5,89	-0.9	apolar(yüksüz)	nötr
Tirozin(Y)	181.19	5,66	-1.3	polar(yüklü)	nötr
Valin(V)	117.15	5,97	4.2	apolar(yüksüz)	nötr

İncelenen çalışmadaki tekrarlanan mutasyonlar temel alınarak Tablo 4.3 incelendi. Glisin yerine gelen arjinin, aspartik asit ve serin aminoasitlerinin özelliklerine bakıldı. Karşılaştırma sonucunda ise molekül kütlesinde artış, izoelektrik noktasında değişim, hidropati indeksinde küçülme, yük durumunda ve asidik-bazik özelliklerinde bir dizi değişim gözlendi. Hidropati indeksinde azalma hidrofilik özelliğin arttığını göstermektedir. Yüksüz olan glisin yerine değişen aminoasitlerin (aspartik asit (D), arjinin (R), serin (S)) ise yüklü olduğu görülmektedir. Bu değişimler alfa-I zincir yapısını etkilediğinden yapıdaki bozulmanın Osteogenez imperfekta hastalığıyla ilişkisi olduğu düşünülmektedir.

Genom dizileme projelerinin temel amaçlarından biri, hastalığın genetiğinin daha iyi anlaşılmasını sağlamak ve hastalıkla ilişkili genlerin tanımlanmasına yardımcı olmaktır (Green, 1997; Hogenesch et al., 2001; Lander et al., 2001). Çalışmamıza

benzer olarak yapılmış Flier vd. yaptığı çalışma incelendiğinde, Tip I prokollajen genlerinde ki mutasyonların osteoartrit ve osteoporoza, Tip III prokollajen genlerinde ki mutasyonların ise aort anevrizmasına neden olduğu görülmüştür. Fakat test ettikleri hipotezin, dokuların parçalandığı bazı yaygın hastalıklar için geçerli olduğuna artık ikna olmuş olsalar da, osteoporoz, osteoartrit veya aort anevrizması olan milyonlarca insanın ne kadarının kollajen genlerinde mutasyonlara sahip olduğu net değildir (Prockop, 1992). Bu nedenle, teknolojideki ilerlemelerle birlikte yakında bu tür genetik ve dizilimsel analizlerin tanı ve tedavi noktasında kullanılabilirliği beklemek makul görünmektedir.

5. SONUÇLAR VE ÖNERİLER

5.1 Sonuçlar

Bu çalışmada yapılan hesaplamalar ve incelenen özellikler ile kollajen alfa-I zincirinde meydana gelen tek bir amino asitin mutasyonu sonucunda değişen özellikler gözlemlenmiştir. Bu özellikler tüm alfa-I zincirinin stabilitesini, yük durumunu, hidrofilikliğini, molekül kütlesini, izoelektrik noktasını, asidik-bazik özelliklerini etkilemektedir. Tekrarlanan mutasyonlar incelendiğinde glisin yerine gelen amino asitlerin, yüklü olduğu, hidropati indeksinde azalma olduğu bununda hidrofilik özelliği arttırdığı, molekül kütlesinde artış olduğu, izoelektrik noktası ve asidik bazik özelliklerinin değiştiği gözlenmiştir. Ayrıca tüm mutasyonlara bakıldığında en çok değişimin glisin amino asitinde meydana gelmesi anlamlıdır. Osteogenez imperfekta'lı hastalarda görülen en belirgin özellik olan kırılgan kemik yapısının sebebinin kollajen yapısındaki stabilite ve mukavemeti sağlayan glisin varlığının mutasyonlarla yerini farklı amino asitlere bırakması olduğu anlaşılmaktadır. Kollajen sarmal yapısının stabilitesini sağlayan glisinin en bariz özelliği molekül kütlesi ve dolayısıyla hacmi en küçük amino asit olmasıdır. Bu mutasyonlar Osteogenez imperfekta tip II hastalığında görülen mutasyonlar olduğundan, bu değişimlerin cam kemik hastalığı olarak da bilinen Osteogenez imperfekta hastalığına sebep olduğu düşünülmektedir.

5.2 Öneriler

Literatürdeki kollajenin moleküler düzeydeki çalışmaları incelendiğinde deneysel çalışmaların yeterince bulunduğu halde teorik hesaplamalı çalışmaların son derece kısıtlı ve yetersiz olduğu görülmektedir. Oysa, bilgisayar teknolojilerinin ve programlama alanının son yıllardaki hızlı gelişimi ile birlikte, deneysel çalışmaların yanı sıra biyoinformatik çalışmalar yapılarak geniş bir aileye sahip olan kollajen molekülünün detaylı analizine katkı sağlamak mümkün olmaktadır. Ayrıca kollajen yapısındaki bozulmalar farklı hastalıklara ve işlev bozukluklarına yol açmaktadır. Osteogenez imperfekta bu hastalıklardan sadece bir tanesidir. Bu çalışmada sadece hastalığın en şiddetli ve ölümcül olan versiyonu Osteogenez imperfekta tip II üzerinde durulmuştur. Bu hastalık farklı tiplerde görülebilmektedir. Osteogenez imperfekta hastalığının diğer tipleri ve kollajen kaynaklı farklı hastalıklar içinde mutasyonlar tespit edilerek farklı biyoinformatik analiz çalışmaları yapılabilir.

KAYNAKLAR

Arjadi, R., Nauta, M., Chowdhary, N., & Bockting, C. (2015). A systematic review of online interventions for mental health in low and middle income countries: a neglected field. *Global Mental Health, 2*.

Arnold, W. V., Fertala, A., Sieron, A. L., Hattori, H., Mechling, D., Bächinger, H.-P., & Prockop, D. J. (1998). Recombinant procollagen II: Deletion of D period segments identifies sequences that are required for helix stabilization and generates a temperature-sensitive N-proteinase cleavage site. *Journal of Biological Chemistry, 273*(48), 31822-31828.

Aryal, S. (August 9, 2018). Amino Acids- Properties, Structure, Classification and Functions.

Bahadıroğlu, S., Büyükberber, S. G., Gürbüz, S., Nakkaş, E. Ç., & Şahin, A. ORGANİZMAMIZ İÇİN NİASİN.

Biro, J. (2006). Amino acid size, charge, hydropathy indices and matrices for protein structure analysis. *Theoretical Biology and Medical Modelling, 3*(1), 15.

Bodian, D. L., Chan, T.-F., Poon, A., Schwarze, U., Yang, K., Byers, P. H., . . . Klein, T. E. (2009). Mutation and polymorphism spectrum in osteogenesis imperfecta type II: implications for genotype–phenotype relationships. *Human molecular genetics, 18*(3), 463-471.

Bodian, D. L., Madhan, B., Brodsky, B., & Klein, T. E. (2008). Predicting the clinical lethality of osteogenesis imperfecta from collagen glycine mutations. *Biochemistry, 47*(19), 5424-5432.

Bodian, D. L., Radmer, R. J., Holbert, S., & Klein, T. E. (2011). Molecular dynamics simulations of the full triple helical region of collagen type I provide an atomic scale view of the protein's regional heterogeneity. In *Biocomputing 2011* (pp. 193-204): World Scientific.

Capriotti, E., Calabrese, R., & Casadio, R. (2006). Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics, 22*(22), 2729-2734.

Cen, L., Liu, W., Cui, L., Zhang, W., & Cao, Y. (2008). Collagen tissue engineering: development of novel biomaterials and applications. *Pediatric research, 63*(5), 492-496.

Chang, S.-W., Shefelbine, S. J., & Buehler, M. J. (2012). Structural and mechanical differences between collagen homo-and heterotrimers: relevance for the molecular origin of brittle bone disease. *Biophysical journal, 102*(3), 640-648.

Chen, Z., Pang, Y., Liu, X., Wang, X., Deng, Z., & Sun, X. (2001). *The Metabolic and Molecular*. Paper presented at the Bases of Inherited Disease; McGraw-Hill, Inc., Health Professions Division.

Cotton, R. G. (2000). Progress of the HUGO mutation database initiative: a brief introduction to the human mutation MDI special issue. *Human Mutation, 15*(1), 4-6.

Eyre, D. R., Paz, M. A., & Gallop, P. M. (1984). Cross-linking in collagen and elastin. *Annual review of biochemistry, 53*(1), 717-748.

Eyre, D. R., Weis, M. A., & Wu, J.-J. (2008). Advances in collagen cross-link analysis. *Methods, 45*(1), 65-74.

Fforde, A. (2019). *From plan to market: The economic transition in Vietnam*: Routledge.

Gargaud, M., Amils, R., & Cleaves, H. J. (2011). *Encyclopedia of astrobiology* (Vol. 1): Springer Science & Business Media.

Gautieri, A., Uzel, S., Vesentini, S., Redaelli, A., & Buehler, M. J. (2009). Molecular and mesoscale mechanisms of osteogenesis imperfecta disease in collagen fibrils. *Biophysical journal, 97*(3), 857-865.

Gelse, K., Pöschl, E., & Aigner, T. (2003). Collagens—structure, function, and biosynthesis. *Advanced drug delivery reviews, 55*(12), 1531-1546.

Green, E. D. (1997). The human genome project and its impact on the study of human disease. *The genetic basis of human cancer*, 33-64.

Hofmann, H., Fietzek, P., & Kühn, K. (1980). Comparative analysis of the sequences of the three collagen chains α1 (I), α2 and α1 (III): Functional and genetic aspects. *Journal of molecular biology, 141*(3), 293-314.

Hogenesch, J. B., Ching, K. A., Batalov, S., Su, A. I., Walker, J. R., Zhou, Y., . . . Cooke, M. P. (2001). A comparison of the Celera and Ensembl predicted gene sets reveals little overlap in novel genes. *Cell, 106*(4), 413-415.

Hyry, M. LYSYL HYDROXYLASES 1 AND 2.

Idrees, M., Mohammad, A. R., Karodia, N., & Rahman, A. (2020). Multimodal role of amino acids in microbial control and drug development. *Antibiotics, 9*(6), 330.

Ileana, R., & Francois, K. (2014). Biopolymers for application in photonics. *NBI-technologies*(4).

Lampe, A., & Bushby, K. (2005). Collagen VI related muscle disorders. *Journal of medical genetics, 42*(9), 673-685.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., . . . Fitzhugh, W. (2001). Erratum: Initial sequencing and analysis of the human genome: International Human Genome Sequencing Consortium (Nature (2001) 409 (860-921)). *Nature, 412*(6846), 565-566.

Miller, M. P., & Kumar, S. (2001). Understanding human disease mutations through the use of interspecific genetic variation. *Human molecular genetics, 10*(21), 2319-2328.

Mitaku, S., Hirokawa, T., & Tsuji, T. (2002). Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane–water interfaces. *Bioinformatics, 18*(4), 608-616.

Myllyharju, J., & Kivirikko, K. I. (2001). Collagens and collagen-related diseases. *Annals of medicine, 33*(1), 7-21.

Nassa, M., Anand, P., Jain, A., Chhabra, A., Jaiswal, A., Malhotra, U., & Rani, V. (2012). Analysis of human collagen sequences. *Bioinformation, 8*(1), 26.

Nimni, M. E. (1973). Metabolic pathways and control mechanisms involved in the biosynthesis and turnover of collagen in normal and pathological connective tissues. *Journal of Oral Pathology & Medicine, 2*(4), 175-202.

Prockop, D. J. (1992). Mutations in collagen genes as a cause of connective-tissue diseases. *New England Journal of Medicine, 326*(8), 540-546.

Prockop, D. J., & Kivirikko, K. I. (1995). Collagens: molecular biology, diseases, and potentials for therapy. *Annual review of biochemistry, 64*(1), 403-434.

Rauch, F., & Glorieux, F. H. (2004). Osteogenesis imperfecta. *The Lancet, 363*(9418), 1377-1385.

Stefl, S., Nishi, H., Petukh, M., Panchenko, A. R., & Alexov, E. (2013). Molecular mechanisms of disease-causing missense mutations. *Journal of molecular biology, 425*(21), 3919-3936.

Steplewski, A., Majsterek, I., McAdams, E., Rucker, E., Brittingham, R. J., Ito, H., . . . Fertala, A. (2004). Thermostability gradient in the collagen triple helix reveals its multi-domain structure. *Journal of molecular biology, 338*(5), 989-998.

van Dijk, F. S., Nikkels, P. G., den Hollander, N. S., Nesbitt, I. M., van Rijn, R. R., Cobben, J. M., & Pals, G. (2011). Lethal/severe osteogenesis imperfecta in a large family: a novel homozygous LEPRE1 mutation and bone histological findings. *Pediatric and Developmental Pathology, 14*(3), 228-234.

Wagner, I., & Musso, H. (1983). New naturally occurring amino acids. *Angewandte Chemie International Edition in English, 22*(11), 816-828.