



Derleme Makale / Review Article

ALPHAFOLD: DERİN ÖĞRENME VE SİNİR AĞLARI YOLUYLA PROTEİN KATLAMASINDA DEVRİM YARATMAK

ALPHAFOLD: REVOLUTIONIZING PROTEIN FOLDING THROUGH DEEP LEARNING AND NEURAL NETWORKS

Burcu TEKİN¹

Rafiq GURBANOV²

<https://doi.org/10.55071/ticaretfbid.1323165>

Sorumlu Yazar / Corresponding Author
rafig.gurbanov@bilecik.edu.tr

Geliş Tarihi / Received
30.10.2023

Kabul Tarihi / Accepted
07.12.2023

Öz

AlphaFold, bir protein dizisinin üç boyutlu yapısını tahmin etmek için derin sinir ağlarını ve gelişmiş makine öğrenimi tekniklerini kullanan, DeepMind ekibi tarafından geliştirilmiş bir protein katlama tahmin aracıdır. Protein katlanmasının tahmini, hesaplamalı biyolojide uzun süredir devam eden bir sorun olmuştur ve doğru protein yapısı tahmin yöntemlerinin geliştirilmesi, bilim camiasının büyük ilgisini çekmiştir. AlphaFold, önce bir proteinin yerel yapısını tahmin ettiği ve ardından genel yapıyı bir araya getirdiği iki aşamalı bir yaklaşım kullanır. AlphaFold, iki yılda bir yapılan CASP (Yapı Tahmininin Kritik Değerlendirmesi) deneylerinde diğer son teknoloji yöntemleri geride bırakarak çok çeşitli proteinlerin yapısını tahmin etmede kayda değer bir başarı elde etmiştir. AlphaFold'un tahminlerinin doğruluğu, protein işlevini ve hastalık mekanizmalarını, ilaç keşfini ve sentetik biyolojiyi anlamak için önemli etkilere sahiptir. Bu derlemede, AlphaFold'un geliştirilmesine, temel metodolojisine ve CASP deneylerindeki performansına genel bir bakış sunulmaktadır. Ek olarak, AlphaFold'un protein mühendisliği, ilaç keşfi ve yapısal biyolojideki potansiyel uygulamaları da tartışılmaktadır.

Anahtar Kelimeler: AlphaFold, makine öğrenmesi, protein katlanması, protein yapı tahmini.

Abstract

AlphaFold is a protein folding prediction tool developed by the DeepMind team, which leverages deep neural networks and advanced machine learning techniques to predict the three-dimensional structure of a protein sequence. The prediction of protein folding has been a long-standing challenge in computational biology, and the development of accurate protein structure prediction methods has been of great interest to the scientific community. AlphaFold employs a two-stage approach, in which it first predicts the local structure of a protein and then assembles the global configuration. AlphaFold has achieved remarkable success in predicting the structure of a wide range of proteins, outperforming other state-of-the-art methods in the biennial CASP (Critical Assessment of Structure Prediction) experiments. The accuracy of AlphaFold's predictions has significant implications for understanding protein function and disease mechanisms, drug discovery, and synthetic biology. This review provides an overview of AlphaFold's development, basic methodology, and performance in CASP experiments. Moreover, potential applications of AlphaFold in protein engineering, drug discovery, and structural biology are also discussed.

Keywords: AlphaFold, machine learning, protein structure prediction, protein folding.

¹Bilecik Şeyh Edebali Üniversitesi, Lisansüstü Eğitim Enstitüsü, Biyoteknoloji Doktora Programı, Bilecik, Türkiye. burcutekinofficial@gmail.com. Orcid.org/0000-0003-4177-2245.

²Bilecik Şeyh Edebali Üniversitesi, Mühendislik Fakültesi, Biyomühendislik Anabilim Dalı, Bilecik, Türkiye. Merkezi Araştırma Laboratuvarı Uygulama ve Araştırma Merkezi, Bilecik Şeyh Edebali Üniversitesi, Bilecik, Türkiye. rafig.gurbanov@bilecik.edu.tr, Orcid.org/0000-0002-5293-6447.

1.GİRİŞ

Doğal olarak evrim geçirmiş proteinler, ince ayarlı üç boyutlu yapıları sayesinde şaşırtıcı çeşitlilikteki moleküler işlevleri sergilerler. Proteinlerin üç boyutlu yapıları, genetik olarak kodlanmış amino asit dizileri tarafından belirlenir (Kuhlman & Bradley, 2019). Amino asit dizisi ile protein yapısı arasındaki ilişkinin kapsamlı bir şekilde anlaşılması, genom dizisi verilerinden fonksiyon tahmini yapmak ve belirli yöntemlerle amino asit dizilerinin tasarımı yoluyla yeni protein işlevlerinin rasyonel mühendisliğini mümkün kılarak yeni olanaklar sağlamaktadır.

Ancak, bir proteinin nasıl katlandığının belirlenmesi, moleküler biyolojinin en karmaşık sorunlarından biridir (Dill & MacCallum, 2012) ve bu sürecin anlaşılması, tıp, biyoteknoloji ve ilaç keşfi alanları için oldukça önemlidir. Son yıllarda proteinlerin üç boyutlu yapılarının tahmini ve tasarımı alanlarında önemli ilerlemeler kaydedilmiştir. Birçok alanda kullanılan yapay zeka ve makine öğrenmesi algoritmaları, temel amino asit dizilerinden protein yapılarının tahmin edilmesinde de kullanılmaya başlanmış ve bu alanda önemli gelişmeler yaşanmıştır (Pakhrin ve ark., 2021). Bununla birlikte yıllar içerisinde protein yapılarının tahmin edilmesinde bulunan teknolojik yenilikler de yapay zeka tabanlı çalışmalara hız kazandırmıştır. Bu yenilikler arasında proteinlerin enerji fonksiyonlarının anlaşılması (Faure ve ark., 2022; Gasic ve ark., 2021), Protein Data Bankası (PDB) (<https://www.rcsb.org/>) (Burley ve ark., 2022) gibi veri tabanlarında deneysel olarak yapısı belirlenmiş protein sayısındaki artış, protein modelleme algoritmalarının (Li ve ark., 2022) çoğalması ve algoritmaları oluşturacak/çalıştıracak bilgisayar tabanlı güçlerin (CPU ve GPU) artması sıralanabilir. Dahası yeni nesil sekanslama teknikleri ile birlikte artan genom bilgisi ve metagenomik çalışmalar da amino asit dizilerinden proteinlerin yapılarının tahmin edilmesinde ilk kaynağı oluşturularak alandaki gelişmelere hız kazandırmıştır (Chen ve ark., 2016; Kuhlman & Bradley, 2019).

Bu çalışmada, proteinlerin katlanması, amino asit dizilerinden protein yapılarının tahmin edilmesi ve amino asit dizisinden protein yapısını tahmin etmek için derin öğrenme tabanlı bir sinir ağı yaklaşımı kullanan AlphaFold teknolojisinin çalışma prensibi ve pratik uygulamaları geniş bir kapsamda derlenmiştir.

2. PROTEİN KATLANMASI

Proteinlerin üç boyutlu yapılarına katlanma süreci, moleküler biyoloji alanındaki en karmaşık problemlerden biridir ve bu sürecin anlaşılması, hayatın yapı taşı olan proteinlerin işlevlerini daha iyi anlamamıza yardımcı olacaktır. Proteinler, kimyasal reaksiyonları katalizlemeden hücreler arasında sinyal iletimine kadar birçok biyolojik işlevi yerine getiren önemli moleküllerdir. Bir proteinin işlevi, onu oluşturan amino asit dizisi tarafından belirlenen benzersiz üç boyutlu yapısına bağlıdır.

Proteinler olmaları gereken üç boyutlu konformasyona spontane olarak katlanabilme yeteneğine sahiptirler. Yeniden katlanma deneyleri, bir proteinin katlanmış şeklini (doğal formu) belirlemek için gerekli bilginin tamamen onun lineer amino asit dizisinde bulunduğunu ortaya koymuştur (Anfinsen ve ark., 1961; Lumry & Eyring, 1954). Anfinsen'in termodinamik hipotezine göre proteinlerin katlanma bilgisi polipeptit zincirinin enerji düzeyine göre ayarlanmaktadır. Deneye göre bir proteinin doğal (katlanmış) formunun sahip olduğu termodinamik enerji kararlılığı en düşük serbest enerji düzeyidir (Anfinsen & Scheraga, 1975; Anfinsen, 1973).

Protein katlama işlemi, hidrofobik etkileşimler, elektrostatik etkileşimler, hidrojen bağı ve van der Waals kuvvetleri dahil olmak üzere fiziksel ve kimyasal kuvvetlerin karmaşık bir etkileşimi tarafından yürütülür (Bolen & Baskakov, 2001). Bu kuvvetler, protein zincirini, proteinin en

kararlı ve biyolojik olarak aktif durumunu temsil eden en düşük enerji konformasyonuna yönlendirir.

Suda çözünür, küresel proteinlerin katlanması destekleyen itici gücün, hidrofobik amino asit yan zincirlerinin sudan uzağa gömülmesi olduğu düşünülmektedir; katlanma, esnek bir polipeptit zincirinin tanımlanmış bir üç boyutlu konformasyona doğru evrilmesine eşlik eden konfigürasyonel entropinin kaybıyla karşı karşıyadır (Dill, 1990). Protein çekirdeğindeki polar olmayan yan zincirlerin sıkı bir şekilde paketlenmesi, çekici van der Waals etkileşimlerini geliştirir ve entropik olarak elverişsiz iç boşlukları ortadan kaldırır (Burley & Petsko, 1988). Bunun yanında, katlanma sırasında polar grupların proteinin iç kısımlarında gömülü kalmasından dolayı su ile etkileşimde bir açık meydana gelir. Bu açık protein içi hidrojen bağların ve tuz köprülerinin oluşumları ile telafi edilir. Bu etkileşimler proteinin kararlılığından çok proteinin özgülüğünün belirlenmesine katkı sağlar (Xu ve ark., 2020).

Biyolojide protein katlanmasının merkezi önemine rağmen, bir proteinin doğal katlanmış yapısını tahmin etmek temel bir zorluk olmaya devam etmektedir. Bunun nedeni kısmen, bir proteinin katlanabileceği olası yolların sayısının, belirli bir protein için astronomik düzeyde (Levinthal Paradoksu) büyük olmasıdır (Levinthal, 1969; Zwanzig ve ark., 1992). Ayrıca, katlama işlemi tam olarak anlaşılabilir değildir. Şaperon proteinleri, translyasyon sonrası modifikasyonlar, pH ve sıcaklık gibi çevresel koşullar dahil olmak üzere birçok faktör proteinin katlanmasını etkileyebilir (Salahuddin ve ark., 2016).

Son birkaç on yılda, protein yapısını amino asit dizisinden tahmin etmek için hesaplama yöntemlerinin geliştirilmesinde önemli ilerlemeler kaydedilmiştir. Bu yöntemler, deneysel verilerle tutarlı bir dizi olası protein yapısı oluşturmak için ampirik modellerin, moleküler dinamik simülasyonlarının ve makine öğrenimi algoritmalarının bir kombinasyonuna dayanır (Pearce & Zhang, 2021). Bununla birlikte, en gelişmiş yöntemlerin bile, özellikle daha büyük ve daha karmaşık proteinler için doğrulukları hala sınırlıdır. Protein yapılarını tahmin etme becerileri geliştirmek, hastalık mekanizmalarını, ilaç keşfini ve biyoteknolojiyi anlamak için önemlidir. Araştırmacılar, katlanma sürecini daha iyi anlayarak ve daha doğru hesaplama yöntemleri geliştirerek, proteinlerin işlevine ilişkin yeni içgörülerin kilidini açabilir ve çok çeşitli hastalıklar için yeni tedaviler geliştirebilirler.

3. PROTEİN YAPI TAHMİNİ

Canlı organizmalarda en yaygın olarak bulunan proteinler sahip oldukları spesifik özellikleri ve karmaşık yapıları nedeniyle akademik çalışmalarda büyük ilgi görmektedir. Proteinlerin yapısını ve işlevini ortaya çıkarmada çeşitli ileri deneysel teknikler kullanılmaktadır. Deneysel olarak yapısı belirlenen protein verileri barındıran Uniprot (Evrensel Protein Veri Bankası) ("UniProt: the universal protein knowledgebase in 2021," 2021; "UniProt: the Universal Protein knowledgebase in 2023," 2023) ve PDB'de (Burley ve ark., 2022) veri sayısı her yıl katlanarak artmaktadır. Ancak gelişen gen sekanslama yöntemleri ile proteinlere ait amino asit dizi verileri daha büyük bir hızla artmaktadır. Deneysel olarak proteinlerin yapılarını ortaya çıkarmada kullanılan X-ışını kristalografisi ve NMR (Nükleer manyetik rezonans) gibi yöntemler zaman, kaynak ve iş gücü tükettiğinden artan günümüzde protein sekanslarının yapılarını belirlemede yetersiz kalmaktadır (Marx, 2022).

Mart 2023 (<https://www.rcsb.org/>) itibarıyla PDB'de deneysel olarak tanımlanmış 200.000'i aşkın protein yapısı olmasına rağmen; bu rakam Uniprot'ta (<https://www.uniprot.org/>) bulunan protein genlerine ait verinin çok küçük bir kısmını karşılamaktadır. Büyüyen ihtiyaçla birlikte, proteinlerin üç boyutlu yapısını tahmin edebilen biyogisayar tabanlı yöntemlerde artmaktadır. Bu yöntemler

çalışma prensiplerine göre farklılık göstermektedir. Protein yapı tahmininde kullanılan bilgisayar tabanlı metotlar iki ana başlık altında toplanabilir (Kuhlman & Bradley, 2019). Bunlardan ilki PDB’de bulunan protein yapılarını kullanan şablon tabanlı (template-based) metottur. Kısaca özetlenirse, şablon tabanlı protein yapı tahmininde yapısı hedeflenen proteine homolog olan protein yapıları PDB üzerinde taranır ve bulunan homolog yapılara bağlı olarak protein yapı tahmini oluşturulur (de Almeida Paiva ve ark., 2022). Homoloji modelleme, karşılaştırmalı modelleme ve bilinen protein yapıları ile hizalamaya dayanan (threading) metotlar, şablon tabanlı protein yapı tahmininde sıklıkla kullanılmaktadır (Arnold ve ark., 2006; Waterhouse ve ark., 2018).

Şablon tabanlı yöntemler , homolog protein yapılarının bulunma durumuna dayanırken, şablon tabanlı olmayan (template-free) yöntemler, bilinen herhangi bir protein yapısına ihtiyaç duymadan hedeflenen proteinin üç boyutlu yapısını tahmin etmeyi amaçlar (Dhingra ve ark., 2020). Bu yöntem genellikle uygun şablonlar (homolog protein yapıları) bulunmadığında veya yapısı hedeflenen protein bilinen yapılardan oldukça farklı olduğunda oldukça yararlıdır.

Yapısı bilinen bir proteine ait veri şablon olarak kullanılmadığında hedeflenen proteinin yapısını tahmin etmek için elektrostatik, van der Waals kuvvetleri ve hidrojen bağı gibi temel ilkeler kullanılır (Dhingra ve ark., 2020). Bu yöntemler belirli bir protein dizisinin en düşük enerji konformasyonunu belirlemek için enerji minimizasyonu algoritmalarını kullanmaktadır. Önceki bölümde de bahsedildiği gibi proteinin en düşük enerjili konformasyonunu bulmak için denenmesi gereken geniş konformasyonel olasılıklar bulunur. Ek olarak yapısı hedeflenen proteinin en düşük enerjiye sahip konformasyonunu araştırmada moleküler dinamik simülasyonlar (Jiang & Wu, 2014) ve Monte Carlo (Ovchinnikov ve ark., 2016) yöntemleri de yaygın olarak kullanılmaktadır. Bir proteinin yapısını tahmin etmede kullanılan bir diğer yöntem ise, yapısı belirli olan proteinlerin yapılarının istatistiksel analizini kullanmaktır. Bu yöntemin temelinde yatan bilgi, protein ikincil yapı elemanları ve protein kıvrımları gibi korunmuş yapısal özelliklerine dayanmaktadır (Marks ve ark., 2012; Yang ve ark., 2020).

Son yıllarda, şablon tabanlı olmayan yöntemler arasında makine öğrenimi destekli uygulamalar popülerlik kazanmıştır. Makina öğreniminde kullanılan yapay zeka (artificial intelligence/AI) verilerdeki kalıpları öğrenmek ve bu verilere dayalı tahminler yapmada çeşitli algoritmalar kullanır. Yapay zeka, protein dizileri ve üç boyutlu yapılardaki kalıpları öğrenerek hedeflenen proteinin yapısını tahmin edebilir. Makina öğrenimi algoritmalarına sinir ağları, karar ağaçları, rasgele ormanlar ve vektör makinaları örnek verilebilir (AlQuraishi, 2021; Jisna & Jayaraj, 2021). Bu algoritmalar çeşitli veri setleri ile eğitilebilir algoritmalarlardır. Amino asit dizileri, protein üç boyutlu yapıları, NMR veya X-ışını kristalografisi verileri gibi çeşitli deneysel veriler bu algoritmalar için eğitici kaynaklardır (Perrakis & Sixma, 2021).

Makine öğrenmesine dayalı protein yapı tahmini için yaygın olarak kullanılan yöntemlerden biri ters katlanmadır (inverse folding) (Hsu ve ark., 2022). Ters katlanma yönteminde hedef istenen bir yapıya katlanacak protein dizisini tahmin etmektir. Bu yaklaşım varyasyonel otokodlayıcı veya üretken çekişmeli ağ gibi modelleri kullanır (Ongie ve ark., 2020). Modeller belirli bir yapıya katlanması muhtemel protein dizilerinin dağılımını öğrenir ve bu dağılımı kullanarak hedef yapıya benzer yeni diziler oluşturur.

Başka bir makina öğrenmesi temelli yaklaşım, derin öğrenme modeli ile etkileşim tahminidir. Bu yöntem, bir proteinin üç boyutlu yapısında bulunan amino asitlerin kendi arasında oluşturabilecekleri etkileşimleri tahmin eder (Li ve ark., 2023). Evrişimli sinir ağları veya tekrarlayan sinir ağları gibi derin öğrenme modelleri, protein dizilerindeki amino asitlerin etkileşimlerini tahmin etmede kullanılmaktadır (Jisna & Jayaraj, 2021; Pakhrin ve ark., 2021). Bu modeller, sonuçlanması muhtemel amino asit etkileşimlerini öğrenerek hedeflenen proteinin yapı

tahminini oluşturabilir. Dahası, sinir ağları ile kalıntılar arası mesafe haritaları oluşturulabilir, proteindeki burulma açıları ve protein zincirindeki dihedral açılar tahmin edilebilmektedir (Gogoi ve ark., 2023; Shen & Bax, 2013). Yukarıda açıklanan tüm bu makine öğrenmesi algoritmaları ile tahmin edilen değerler daha sonra bir proteinin üç boyutlu yapısını ortaya çıkarmada kullanılmaktadır.

3.1. Protein Yapısı Tahmin Deneylerinin Kritik Değerlendirmesi

Protein Yapısı Tahmin Deneylerinin Kritik Değerlendirmesi (CASP) protein yapısının modellenmesinde en iyi uygulamaları belirlemek için iki yılda bir tekralanan uluslararası yarışmadır (Kryshtafovych ve ark., 2019). CASP deneyleri ilk olarak 1994 yılında John Moult tarafından başlatılmıştır (Marx, 2022). Yarışma kapsamında katılımcılara hedef proteinlerin amino asit dizileri verilir ve hedef proteinlere ait üç boyutlu modellerin oluşturulması istenir. Bağımsız değerlendiriciler farklı uygulamaların tahmin ettiği protein yapılarını karşılaştırır. Katılımcıların hedef proteinlerin deneysel yapılarına erişimleri yoktur ve değerlendirmeyi yapan görevliler sunumu yapılan yöntemlerin kimliğini bilmezler. CASP'ta değerlendirilen kriterler arasında, protein yapı modellerine ek olarak, deneysel yapıya yakın uygulamaların iyileştirilmesi, genel bir saptaması, yapı modelinin ve her bir kalıntının doğruluğunun değerlendirilerek bunun yanında protein oligomerlerinin yapısının modellenmesi, çeşitli seyrek veri türlerini kullanarak modelleri geliştirme yeteneği ve işlevi belirlenen proteinin yapısal özelliklerinin doğruluğu gösterilebilir.

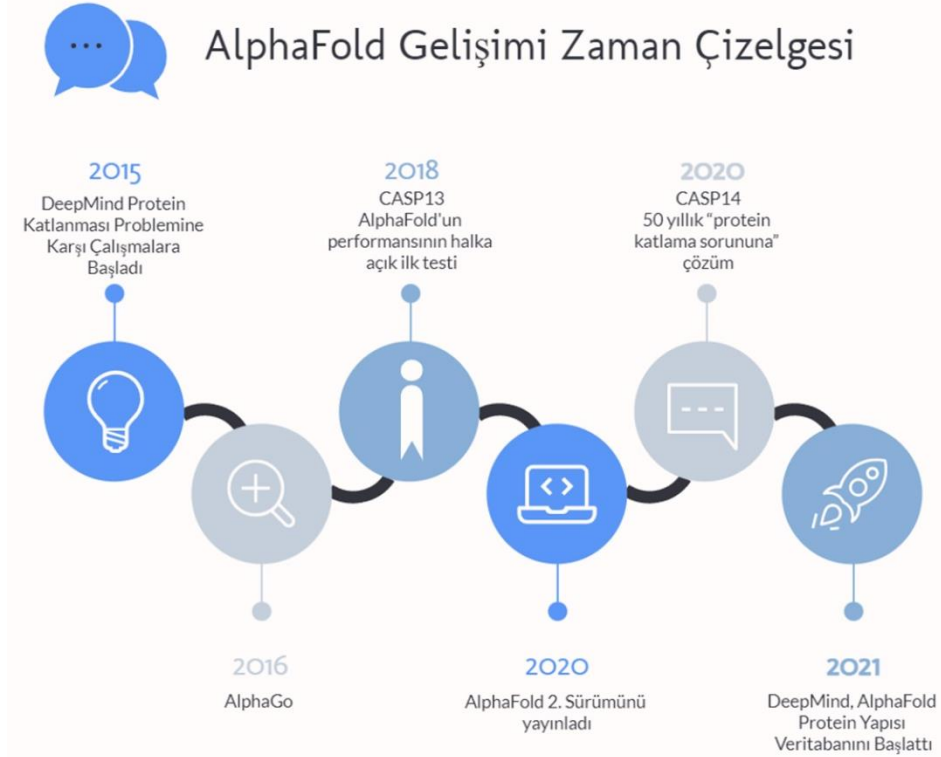
CASP deneylerinin en önemli amacı protein yapı tahmininde şablon temelli olmayan uygulamaların gelişimini desteklemektir. CASP deneylerinin en son versiyonu CASP15 (Cheng ve ark., 2023) 2022 yılında yapılmıştır ve önceki yıllara ait verilerde dahil olmak üzere yarışmaya katılan uygulamalar Protein Yapı Tahmin Merkezi'nin sitesinde (www.predictioncenter.org) herkese açık şekilde yayınlanmaktadır.

4. ALPHAFOLD'UN TARİHSEL GELİŞİMİ

AlphaFold, DeepMind tarafından geliştirilmiş, proteinlerin üç boyutlu yapısını dikkate değer bir doğrulukla tahmin edebilen, derin öğrenme tabanlı bir algoritmadır. AlphaFold'un gelişim sürecinin başlaması 2015 tarihine kadar dayanır (Hey ve ark., 2020) (Şekil 1). Google'un sahibi olduğu DeepMind bu tarihte protein katlanması problemi üzerine makine öğrenmesi algoritmaları kullanarak çalışmaya başlamıştır. Bunu takip eden 2016 yılında yine DeepMind'in geliştirdiği bir makine öğrenmesi algoritması olan AlphaGo bilgisayar sistemlerinin kazanması en zor olan Go oyununu dünya şampiyonuna karşı kazanaak etkinliğini kanıtlamıştır (Li & Du, 2018). DeepMind'in elde ettiği bu başarı aynı algoritmanın protein katlanması problemi içinde kullanılabileceğini düşündürmüştür. Dolayısıyla, AlphaGo'daki aynı algoritma geliştirilerek protein katlanması problemi üzerine çalışan AlphaFold geliştirilmiştir.

AlphaFold'un tanıtılması ve gelişimi 2018 yılında DeepMind'in CASP13 yarışmasına katılacağını duyurması ile başlamıştır. CASP13 deneyinde AlphaFold dikkat çekici bir performans elde ederek proteinlerin üç boyutlu yapılarını yüksek bir doğrulukla tahmin etmiştir (AlQuraishi, 2019). AlphaFold'un CASP13'te elde ettiği RMSD (kök-ortalama-kare sapması) değerleri protein domainlerin için 1.6 angstrom ve tüm bir protein için 2.1 angstrom olarak hesaplanmıştır. RMSD değeri tahmin edilen protein yapısının, aynı proteinin deneysel olarak belirlenen yapısına olan benzerliğinin bir ölçütüdür. Bu değer küçük olması algoritmanın protein tahminindeki başarısını sergilemektedir (Kufareva & Abagyan, 2012).

CASP13'te elde ettiği başarının hemen ardından DeepMind AlphaFold ile ilgili makalesini *Nature* dergisinde yayınlarak algoritmanın mimarisini ve performansını açıkladı (Senior ve ark., 2020). Böylece AlphaFold'un proteinlerin üç boyutlu yapısını X-ışını kristalografisi gibi deneysel teknikler ile karşılaştırılabilir bir doğrulukta tahmin edebildiğini akademik camiaya duyurdu. İlerleyen yıllarda AlphaFold'un algoritması dahada geliştirilerek bir üst versiyon olan AlphaFold2 sürümü yayınlandı. CASP14 (2020) deneyinde test edilen AlphaFold2 önceki versiyonunun çok üzerinde bir başarı elde ederek CASP14'ü büyük bir farkla kazandı (Kryshtafovych, ve ark., 2021). AlphaFold2 önceki versiyonundan üç kat daha doğrulukta 1 angstrom RMSD değerinden daha küçük bir hata ile hedeflenen potenin yapısını tahmin etmeyi başardı. Bu büyük başarının ardından AlphaFold2 CASP düzenleyicileri tarafından, 50 yıldır üzerine çalışılan protein katlanma sorununa bir çözüm olarak kabul edildi (Jumper ve ark., 2021).



Şekil 1. DeepMind AlphaFold'un Tarihsel Gelişimi

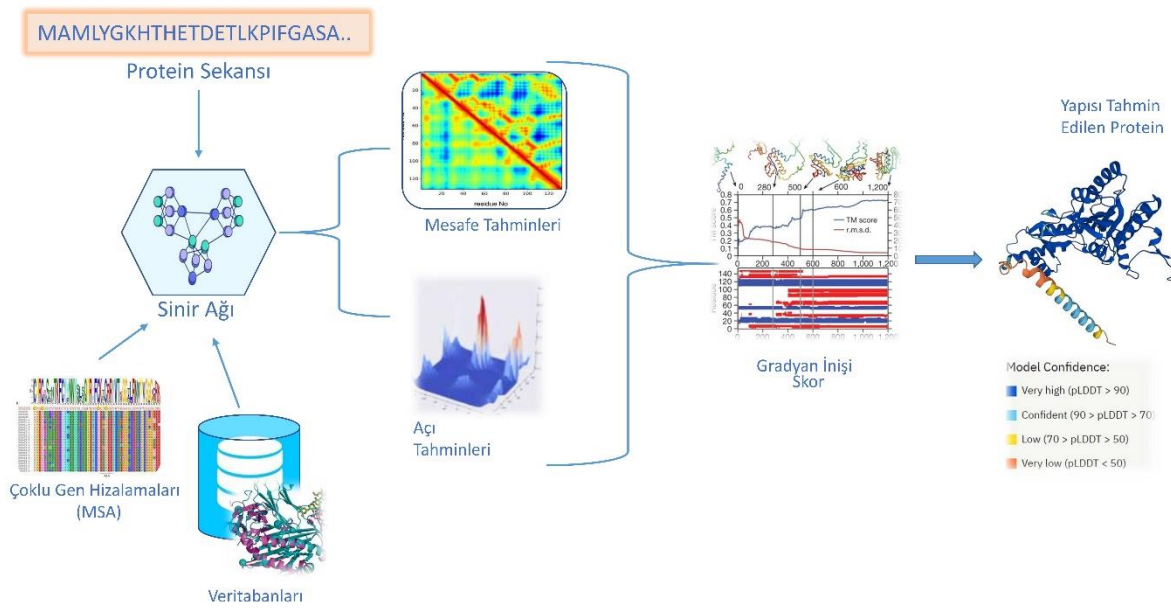
AlphaFold2'un CASP14'teki başarısının *Nature*'de (Jumper ve ark., 2021) yayınlanmasının hemen ardından, DeepMind ikinci bir makale ile neredeyse tüm insan proteomunu tahmin ettiğini gösterdi (Tunyasuvunakool ve ark., 2021). Bununla birlikte, DeepMind EMBL-EBI (Avrupa Moleküler Biyoloji Laboratuvarı-Avrupa Biyoinformatik Enstitüsü) ile yaptığı ortaklık sonucu AlphaFold Protein Yapısı Veritabanı'nı (www.alphafold.ebi.ac.uk) kullanıma açık olarak başlattı. Kullanıma açıldığı tarihten itibaren AlphaFold Protein Yapısı Veritabanı'nın (AlphaFold DB) sahip olduğu veri miktarı sürekli artmaktadır. Sadece insan proteomuna ait veriler değil, bunun yanında diğer bir çok organizmaya ait protein yapıları veritabanı üzerinde mevcuttur. Günümüzde, AlphaFold DB bilimsel çalışmalara katkı sağlamak ve bu çalışmaları hızlandırmak için 200 milyondan fazla proteinin tahmin edilmiş yapılarına açık erişim sağlamaktadır.

AlphaFold'un geliştirilmesi protein tahmini alanında büyük bir ilerleme olarak kabul edilebilir ve hesaplamalı biyoloji dışındaki alanlarda da önemli kullanımları olabilir. AlphaFold'un başarısı, derin öğrenme algoritmalarının gücünün ve onlarca yıllık araştırmalar sonucunda üretilen büyük miktarda verinin bir kanıtıdır. Derin öğrenme algoritmalarının protein yapı verileriyle entegrasyonu, proteinlerin üç boyutlu yapılarının eşi benzeri görülmemiş bir doğrulukla tahmin

edilmesini sağlamıştır. Bu gelişme, ilaç keşfi ve tasarımı için yeni olanaklar sunmaktadır. Dahası, AlphaFold, protein fonksiyonu ve dinamiğinin çalışılmasını hızlandırarak moleküler düzeyde yaşamın temel mekanizmalarına dair bilgiler sağlamaktadır.

4.1. AlphaFold'un Çalışma Prensibi

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 kullanır (Senior ve ark., 2020). Bu sayede protein dizilerindeki amino asit kalıntıları arasındaki karmaşık etkileşimleri modelleyebilir; bu bilgileri evrimsel ve biofiziksel olaylarla birleştirebilir. AlphaFold algoritmasının eğitilmesi için veriler esas olarak PDB'den gelir. AlphaFold'un ilk versiyonu genel olarak basit bir algoritma mimarisine sahiptir (Şekil 2). Tüm sistem, çoklu dizi hizalamaları oluşturmak için, hedef proteinin amino asit dizisine benzer dizileri çeşitli veritabanları üzerinden bulma işlemi ile başlar (Senior ve ark., 2020).



Şekil 2. AlphaFold Versiyon 1 Çalışma Prensibinin Şematik Gösterimi

Çoklu dizi hizalamalarının sistem için önemini kavramada bazı bilgilerin bilinmesi önemlidir. Bir proteinin üç boyutlu yapısı, belirli bir proteini kodlayan genetik diziden çok daha karardır. Bir proteini oluşturan spesifik amino asitlerdeki mutasyonlar doğada oldukça yaygındır, ancak yapıdaki değişiklikler çok daha nadirdir (Romero & Arnold, 2009). Canlı organizmaların çoğu ortak bir atadan türemiştir. Bu nedenle yapılarındaki tüm biyolojik işlevleri yerine getiren proteinler ortak ataya sahip canlılar arasında çok benzer şekillere sahip olacaklardır. Çünkü oksijen transportu gibi aynı temel biyolojik işlevleri nesiller boyunca evrimsel ekonomi ilkelerine uyumlu şekilde sürdürmeleri gerekmektedir. Eğer bir mutasyon, bir proteinin katlanmış halindeki şeklini önemli ölçüde etkiliyorsa, bu proteinin işlevini yerine getirme kabiliyeti kaybolacak ve bu durum, organizmanın hayatta kalma şansını azaltabilecektir. Bu nedenle, doğal seçim, organizmaların uygunluğuna etki ederek bu tür mutasyonlara sahip bireyleri popülasyondan ayıklayacak ve zararlı alellerin sonraki nesillere aktarılmasını engelleyecektir. Protein yapısını ciddi şekilde etkileyen mutasyonlar, protein fonksiyonunu bozarak doğal seçim tarafından ortadan kaldırıldığından, sadece protein yapısını önemli ölçüde etkilemeyen mutasyonlar hayatta kalabilir. Bu nedenle, protein kodlayan bölgelerde meydana gelen mutasyonların çoğunluğunun nötr veya neredeyse nötr olduğu, sadece küçük bir kısmının protein yapısı ve fonksiyonu üzerinde önemli etkilere sahip olduğu düşünülmektedir. Bu nedenle, ortak bir atadan türeyen çok sayıda türdeki belirli bir

proteinin dizilimlerinin karşılaştırılması, proteinin evrimsel tarihine ilişkin değerli bilgiler sağlar (Ponting & Russell, 2002). Bu evrimin bir sonucu olarak, ilgili protein dizilerinin güçlü yapısal benzerliklere sahip olması beklenmektedir. Çoklu dizi hizalamaları, bu benzerlikleri saptamak için güçlü bir araç olarak hizmet eder ve protein yapısı/işlevinde önemli roller oynayabilen korunmuş kalıntıların ve motiflerin tanımlanmasına olanak tanır (Thompson ve ark., 2001). Dolayısıyla bu hizalamalar, proteinlerin yapısal özelliklerini tahmin etmek için değerli bilgiler sağlayabilir; protein katlanması ve kararlılığını hesaplamak için geliştirilen modellere bir temel girdi olarak hizmet edebilir.

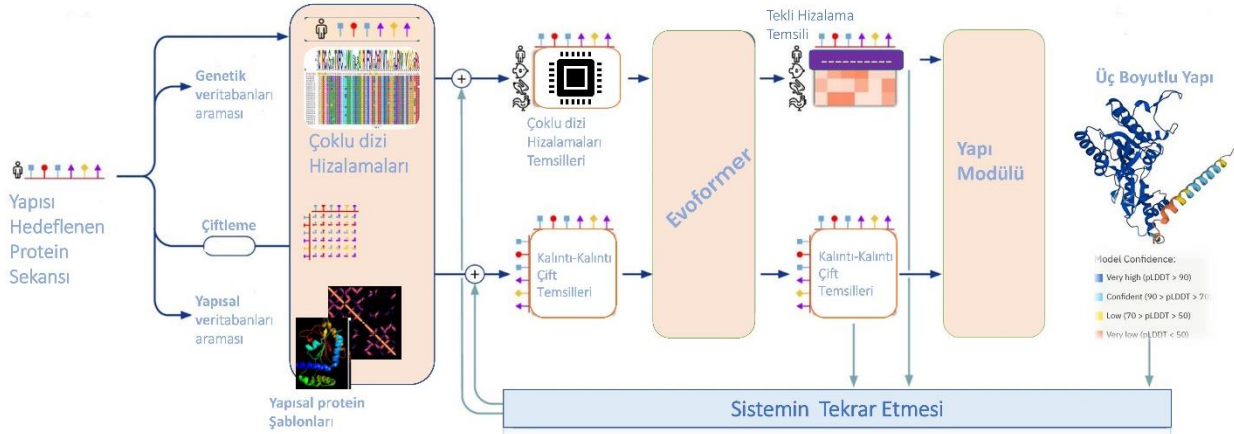
AlphaFold'un ilk sürümü, hedef amino asit dizisini kullanarak ilgili türlerdeki evrimsel eşleşmeleri bulmak için çeşitli veritabanlarını taramaktadır. Çoklu dizi hizalamalarından elde edilen bilgiler, ek girdi özellikleri ile işlenerek yeniden şekillendirilir ve bir iki boyutlu diziyeye dönüştürülür (Senior ve ark., 2020). Bu dizi, ResNet adı verilen ve görüntü sınıflandırması (He ve ark., 2016) gibi işlemler için kullanılan derin rezidüel konvolüsyonel sinir ağına beslenir. ResNet'in ana hedef çıktısı, öngörülme çalışılan proteinin distogramıdır. Distogram, proteinin üç boyutlu yapısında birbirleriyle temas halinde olma olasılığı olan her amino asit çiftinin birbirine olan uzaklığını tahmin eden bir olasılıksal ölçüdür (Senior ve ark., 2020). AlphaFold, hedef distogramı doğrudan tahmin etmek yerine, mesafelerin olası dağılımını tahmin eden bir yaklaşım sergiler. Bu yaklaşımda, iki boyutlu distogram matrisindeki her bir piksel için AlphaFold, 64 farklı mesafe aralığı üzerinden bir olasılık dağılımı tahmin etmektedir. Bu yöntem, modelin belirli kalıntılar arasındaki mesafe hakkında ne kadar emin olduğunu tahmin etmek için oldukça faydalıdır. Nihai olarak, tahmin edilen bu distogram dağılımı, proteinin üç boyutlu şeklini elde etmek için kullanılmaktadır.

Dahası, AlphaFold1, protein yapılarının keşfedilmesinde kullanılan bir yaklaşım olan gradyan inişi yöntemini kullanır. Bu yöntem, her bir ardışık kalıntı arasındaki phi ve psi açılarına dayalı olarak parametre edilmiş bir protein zincirinin fiziksel modelinin oluşturulması ile başlar. Ardından, bu omurga (protein backbone) geometrisi, bir sinir ağı tahmini ile birleştirilmiş bir dizi fiziksel simüle edilmiş kuvvetleri kullanan bir skorlama fonksiyonu ile iteratif olarak iyileştirilir. Bu yöntemin amacı, modelin tahminlerinin fiziksel gerçeklikle uyumlu olmasını sağlamaktır (Senior ve ark., 2020). Özetlenecek olursa, AlphaFold1, öncelikle ilgili türdeki organizmaların protein dizileri arasında çoklu dizi hizalaması yapar. Daha sonra, distogram dağılımını doğrudan tahmin etmek yerine, bir ortakalan sinir ağı kullanarak distogram dağılımını tahmin eder. Bu, iki kalıntı arasındaki mesafe hakkında modelin ne kadar emin olduğunu ölçer. Son olarak, tahmin edilen distogram dağılımı bir dizi fiziksel kuvvet ile birlikte kullanılarak proteinin omurga geometrisi tekrarlanan bir şekilde optimize edilir.

AlphaFold'un ilk versiyonundan farklı olarak, AlphaFold2 üç bileşenden oluşan farklı bir mimariye sahiptir (Şekil 3) (Jumper ve ark., 2021). İlk bileşen olan pre-processing stage (ön işlem aşaması), hedef protein dizisinin hakkında ek bilgi almak için harici veritabanlarını sorgular ve bu bilgiyi kullanarak çoklu dizi hizalaması ve çift temsillemesi oluşturur. İkinci bileşen olan evoformer, çoklu dizi hizalamasını ve çift temsillemesini güncelleyen ve ince etkileşimleri yakalayan 48 katmanlı bir dönüştürücü benzeri ağıdır. Üçüncü bileşen olan yapı modülü, evoformer tarafından oluşturulan soyut temsili alır ve proteinin somut üç boyutlu koordinatlarına dönüştürür.

AlphaFold2, ilk versiyonuyla benzer olarak; benzer evrimsel dizilimleri genetik veritabanında arayarak öncelikle çoklu dizi hizalaması oluşturur. Bunun yanı sıra, hedef protein gen dizilimi ile benzer dizi parçalarına sahip olan bilinen protein yapılarını, yani yapısal şablonları da arar. Eğitim süreci sırasında AlphaFold2, yapısal şablonları, bazen modelin çok zayıf veya hiç şablon almamasına neden olacak şekilde örnekler. Bu yaklaşım, ağın bir çıktı yapısı oluşturmak için basitçe iyi bir girdi şablonunu kopyalamaya güvenememesine neden olur; bunun yerine sistem dizi bilgisinden öğrenmeye zorlanır. AlphaFold2 daha sonra giriş dizisinden, çoklu dizi

hizalamasından ve şablon yapılarından gelen tüm bilgileri, sonrasında daha fazla ayrıntılandırma için kullanılan bir çift gösterime ve bir çoklu dizi gösterimine entegre eder. Verilerin fazlalığı nedeniyle, önışleme aşaması genellikle birkaç kez tekrarlanır ve farklı çift ve çoklu dizi temsilleri oluşturulur. Bu temsiller daha sonra yapı modülüne aktarılmadan önce birleştirilerek ortalaması alınır (Şekil 3).



Şekil 3. AlphaFold Versiyon 2 Network Yapısının Şematik Bir Gösterimi

AlphaFold2'nun ikinci bileşeni olan Evoformer (Jumper ve ark., 2021; Wang ve ark., 2022), çoklu dizi ve çift temsillerinden gelen bilgileri iteratif olarak geliştirmek için tasarlanmış 48 bloktan oluşan bir derin sinir ağıdır. Bu bloklar, her iki temsil türünden gelen girdi ve çıktıları işlemektedir. Evoformer, yapısal modülün aksine, tekrarlayan bir ağı değildir. Katmanları, girdi aşamasından gelen protein temsillerinin güncellenmesini optimize etmek için teknik aşamaları içerir. İki mod arasındaki bilgi alışverişi yoluyla, Evoformer, proteinlerdeki uzamsal ve evrimsel ilişkiler hakkında doğrudan veri elde etme olanağı sağlar.

AlphaFold2'nin son bölümü olan yapı modülü, Evoformer yığınının çıktısını alan ve onu üç boyutlu protein yapısına dönüştüren bir sinir ağıdır (Jumper ve ark., 2021). Yapı modülüne girdisi, Evoformer'dan elde edilen nihai çoklu dizi hizalama gösteriminin ilk satırı ve çift gösterimidir. Yapı modülü (Jumper ve ark., 2020), aynı ağırlıkları paylaşan 8 ağı katmanından oluşur ve proteinin üç boyutlu yapısını yinelemeli olarak iyileştirmek için tekrarlayan bir sinir ağı mimarisi kullanır. Ağı, küresel bir referans çerçevesine göre her kalıntı için bir dizi bağımsız dönüş ve öteleme olarak temsil edilen, proteinin somut bir üç boyutlu omurgası üzerinde çalışır. Omurga, tahmin edilen bir amino asit dizisine dayalı olarak inşa edilir ve ağı eğitim süreci sırasında yinelemeli olarak rafine edilir. Tekrarlayan ağı her yinelemesinde yapı modülü, omurga atomlarının konumlarını, yan zincir yönelimlerini ve kalıntılar için atomik konumları tahmin ederek proteinin üç boyutlu yapısını oluşturur (Jumper ve ark., 2020). Ağı çıktısı, atomik düzeyde doğru olan protein için tahmin edilen bir üç boyutlu yapıdır. Genel olarak, AlphaFold2'nin yapı modülü, protein yapılarını olağanüstü hız ve doğrulukla doğru bir şekilde tahmin ederek protein yapısı tahmini alanında devrim yaratan güçlü bir araçtır.

4.1.1. AlphaFold2 algoritma eğitimi

AlphaFold2, en önemlisi çerçeve hizalı nokta hatası (Fape) (Evans ve ark., 2021) kaybı olmak üzere dört farklı kayıp bileşeninin bir kombinasyonu kullanılarak eğitilmiştir. Fape, tahmin edilen yapının atom koordinatlarını temel gerçeklik yapısınıninkilerle karşılaştırmak için tasarlanmıştır, ancak her yerel kalıntı çerçevesi için farklı bir hizalama kullanır. Bu, yalnızca küresel koordinat doğruluğuna odaklanmak yerine her kalıntı için yerel olarak tutarlı yapılar elde etmenin önemini vurgulamaktadır. Bu yaklaşımın arkasındaki mantık, mükemmel atom koordinatlarını elde

etmenin zorlu bir görev olmasıdır ve başlangıçta doğru alt yapılara öncelik vermek ve ardından küresel yerleşimler hakkında bilgiler edinmek tercih edilir. Özünde Fape, yerel yapısal tutarlılığın önemini vurgulayarak AlphaFold2'nin doğru protein yapıları oluşturmasına yardımcı olabilecek değerli bir kayıp bileşenidir (Jumper ve ark., 2021).

Fape kaybının dışında, AlphaFold2 performansını iyileştirmek için bir dizi yardımcı kayıp kullanır. İlk olarak, son çift temsili, bir mesafe dağılımına doğrusal olarak yansıtılır ve bir çapraz entropi kaybı kullanılarak puanlanır. Evoformer, doğruluğunu artırmak için hedef distogram tarafından denetlenir. İkinci olarak, girdi çoklu dizi hizalamasına rasgele maskeleye uygulanır ve ağa, çıktı çoklu dizi hizalama temsillerini kullanarak bu maskelenmiş bölgeleri yeniden yapılandırma görevi verilir, bu da giriş çoklu dizi hizalama yığınının denetimsiz bir dil ön eğitimiyle sonuçlanır (Lin ve ark., 2023). Bu teknik, AlphaFold2'nin çoklu dizi hizalamalarındaki diziler arasındaki evrimsel ilişkileri öğrenmesine yardımcı olur. Son olarak, yapı modülünün çıkışındaki tek temsiller, pLDDT (per-residue confidence score/her bir amino asit kalite puanı) puanlarını tahmin etmek için kullanılır. AlphaFold2, her kalıntı konumu için modelleme kalitesini tahmin etmeye çalışır. Bu yardımcı kayıpların birincil amacı, modeli yerel tutarlılığa, evrimsel ilişkilere ve bireysel kalıntıların konumlarının doğruluğuna odaklanmaya yönlendirerek tahmin edilen yapıyı iyileştirmektir. Ek olarak yan zincir konumlarını daha iyi tahmin etmek ve örneğin iki atomun aynı yerde olacağı yapı ihlallerini önlemek için kullanılan iki küçük yardımcı kayıp vardır. Tüm bu farklı kayıpların temel amacı, modelin her bir ana bileşenine, eğitim sırasında her bir alt birimin amacına yönelik bir kılavuz görevi gören bireysel bir hedef eklemektir.

Protein yapısının omurgası genellikle bir kalıntı gazı olarak temsil edildiğinden, ağ tahminleri atom koordinatlarının çakışmalarına neden olabilir. Bu çakışmalar, iki kalıntının fiziksel olarak mümkün olanın ötesinde yakın olduğu tahmin edildiğinde ortaya çıkar ve tahmin edilen yapının doğruluğunu ve güvenilirliğini tehlikeye atabilir. Bu sorunu çözmek için, son tahminler üzerinde amber kuvvet alanı (Hornak ve ark., 2006) kullanarak kısıtlı enerji minimizasyonu uygulanır. Bu teknik, belirli atomların hareketlerini kısıtlayarak, serbest enerjinin en aza indirilmesi sürecini içerir ve atomların birbirlerine çok yakın hareket etmelerini önler. Böylece, yapıda kalan herhangi bir çakışma veya yapısal ihlaller giderilerek, daha doğru ve güvenilir bir tahmin ile protein yapısı elde edilir.

AlphaFold2, eğitim verilerini artırmak ve performansını yükseltmek için kendi kendine damıtma olarak bilinen bir teknik kullanmıştır. Başlangıçta AlphaFold2, PDB'deki tüm mevcut protein yapıları üzerinde eğitildi. Ardından, yapıları bilinmeyen yaklaşık 350.000 ek proteinin yapısı tahmin edildi ve bu tahmin edilen yapılar eğitim verilerine dahil edildi. Ancak sadece AlphaFold2'un yüksek hassasiyetle tahmin ettiği kalıntıların gradyan güncellemeleri kullanıldı ve modelin en güvenilir verilerden öğrenmesi sağlandı. Ek olarak AlphaFold2, modelin çıktılarını tekrar tekrar belirli modüllere geri beslemeyi içeren, geri dönüşüm adı verilen bir teknik kullandı. Bu süreç, modelin derinliğini artırdı ve tahmin edilen protein yapısının yinelemeli olarak iyileştirmesine izin vererek daha yüksek doğruluk ve kesinlik sağladı. Genel olarak, kendi kendine damıtma ve geri dönüşüm tekniklerinin kombinasyonu, AlphaFold2'un protein yapısı tahminlerinin doğruluğunu ve güvenilirliğini geliştirmede oldukça etkili oldu.

AlphaFold'un başarısı, tek bir çıkır açan fikre dayandırılmayacak kadar karmaşık bir sürecin sonucudur. Protein katlanmasının çözümü, birçok akıllıca tasarlanmış mekanizma ve yeni fikirlerin bir araya gelmesiyle oluşan, birçok tasarım döngüsünde sürekli olarak iyileştirilmiştir. AlphaFold ekibi, protein katlanması sorununu çözmek için sistematik bir yaklaşım benimsemiştir. Birkaç iterasyonu içeren bir geliştirme sürecine dahil olan sistem, her bir döngüde yeni fikirleri ve geliştirmeleri barındırır. Son sistem, birçok küçük geliştirmenin ve yenilikçi çözümlerin bir araya gelmesiyle elde edilmiştir. Elde edilen başarı, birden fazla sofistike teknik ve tasarım unsurunun entegrasyonuna atfedilebilir. Her biri, çoklu iterasyonlar boyunca dikkatlice rafine edilmiş ve

optimize edilmiştir. Sonuç olarak, AlphaFold, mühendislik tasarımında metodik bir yaklaşımın etkinliğini gösteren mükemmel bir örnektir.

4.2. Protein Katlanması Tümüyle Çözüldü mü?

AlphaFold, dünyanın en yetenekli yapay zeka araştırmacıları tarafından gerçekleştirilen muazzam bir mühendislik başarısıdır. Model, özellikle tek domainli proteinlerinin yapılarını inanılmaz bir doğrulukla tahmin edebilmektedir. Bu, yapay zekanın yeni ilaçlar tasarlama, hastalık yollarını anlama, virüslerle mücadele etme ve daha birçok alanda vazgeçilmez bir araç olarak kullanılacağı tamamen yeni bir bilgisayar biyolojisi çağının başlangıcına işaret etmektedir.

Ancak AlphaFold'un doğruluğu her zaman mükemmel değildir. Örneğin, biyolojide birçok işlem, birçok farklı proteinin bir araya gelerek kompleks moleküler makineler oluşturduğu büyük protein komplekslerine dayanmaktadır (Ziegler ve ark., 2021) ve bu şu anda AlphaFold'un yeteneklerinin dışındadır. AlphaFold, doğal olarak bulunan proteinlerin bir veri kümesinde eğitilmiştir. Ancak bu yeni araçlarla tasarlamak isteyebileceğimiz tamamen yeni *de novo* proteinlere yönelik tahminlerin ne kadar iyi olacağını göstermemektedir. Ek olarak, çoğu protein, hedeflerine bağlandıklarında şekil değiştirerek işlev görür ve bu moleküler dinamikler AlphaFold'un statik protein yapılarına sahip bakış açısıyla yakalanamaz (Nussinov ve ark., 2022). Gerçek bir proteinin bu farklı dinamik konformasyonlardaki esnekliğini modelleyebilmek, protein-protein etkileşimlerini incelemek için hayati önem taşıyacaktır. Ancak, AlphaFold'un, doğru türde eğitim verileri ve mühendislik çabaları sağlandığında bu yeni zorluklara da çözüm üretebileceği muhtemeldir (Goverde ve ark., 2022; Jendrusch ve ark., 2021).

5. ALPHAFOLD UYGULAMALARI

Önceki bölümlerde detaylandırıldığı gibi, DeepMind araştırma ekibi tarafından geliştirilen bir yapay zeka aracı olan AlphaFold, protein yapısı tahmini alanında devrim yaratmıştır. AlphaFold, proteinlerin üç boyutlu yapısını tahmin etmek için derin sinir ağlarını kullanarak, tahminleri genellikle deneysel yöntemlerle eşleşen ve hatta onu aşan, benzeri görülmemiş bir doğruluk elde etmiştir. Bu atılım, ilaç keşfi, hastalık araştırması, sentetik biyoloji, tarım (Miyazawa ve ark., 2022), yapısal biyoloji ve protein mühendisliği gibi alanlarda AlphaFold için bir dizi heyecan verici uygulamanın önünü açmıştır. Bu bölümde, AlphaFold'un bilimsel araştırmaları ilerletmek ve en zorlayıcı biyolojik ve tıbbi problemlerin bazılarının çözümü için potansiyel uygulamalarına değinilecektir.

5.1. İlaç Keşfi/Tasarımı ve Hastalık Mekanizmalarını Anlamak

AlphaFold, son zamanlarda ilaç keşfi ve tasarımı ile hastalık mekanizmalarını anlamak gibi önemli biyomedikal araştırmalarda kullanılmakta olan bir yapay zeka teknolojisidir. AlphaFold'un temel özelliği, proteinlerin üç boyutlu yapılarını öngörme kabiliyetidir. Protein yapıları, proteinlerin işlevlerini belirleyen en önemli faktörlerdir ve bu nedenle birçok hastalığın mekanizması, proteinlerin yapısındaki bozukluklar ile ilgilidir. Yeni bir ilaç tasarlamadan önce, hedef proteinin yapısının bilinmesi gereklidir. Çünkü ilaçlar, hedef proteinin yapısına bağlanarak işlevlerini gerçekleştirirler. AlphaFold'un protein yapılarını doğru bir şekilde öngörmesi, ilaç tasarımı sürecinde büyük bir zaman ve maliyet tasarrufu sağlar. Ayrıca, AlphaFold, birçok hastalığın moleküler mekanizmalarını anlamak için kullanılmaktadır. Örneğin, Alzheimer gibi nörodejeneratif hastalıklar, proteinlerin birbirleriyle etkileşimleri sonucu oluşan toksik agregatların birikmesiyle ilişkilidir (Liu ve ark., 2022). AlphaFold'un bu proteinlerin yapısını doğru bir şekilde öngörmesi, bu hastalıkların mekanizmalarını anlamak için kullanılan bir araç haline gelmiştir (Borkakoti & Thornton, 2023).

AlphaFold, Alzheimer ve Parkinson gibi hastalıklar için potansiyel ilaç hedeflerinin belirlenmesi de dahil olmak üzere birçok ilaç keşif projesinde halihazırda kullanılmaktadır. AlphaFold2, Alzheimer hastalığında etkili olan tau proteininin yapı tahmininde ve ayrıca bu proteine bağlanan serin proteaz inhibitörleri üzerinde yapılan çalışmalarda kullanılmıştır (Jiang ve ark., 2023; Savvitskiy ve ark., 2023). Yakın zamanda yapılan bir çalışmada ise AlphaFold, Alzheimer hastalığının tedavisinde kullanılabilir bir bitki olan *Artemisia argyi*'nin (Pelin otu) terapötik etkilerinin ortaya çıkarılmasında kullanılmıştır. Araştırmacılar bitkinin etkileyebileceği protein yapılarının tahmin edilmesinde AlphaFold'dan yararlanmışlardır (Wang & Huang, 2023). Dahası AlphaFold'un protein agregasyonunun çalışılmasında etkili bir araç olabileceği Pinheria ve arkadaşları tarafından ayrıntılı olarak derlenmiştir (Pinheiro ve ark., 2021). Derlemede bahsedildiği gibi, protein agregasyonu yaygın bir fenomendir ancak bu yapıların detayları iyi anlaşılabilmiştir. AlphaFold, globüler proteinler tasarlamak için kullanışlı olabilir, ancak agregasyon yapılarının tahmin edilmesi çok zorlu bir süreçtir. Birden fazla amiloid polimorfizmi ve heterojen yapıların olması, aynı zamanda bu teknolojinin temel aldığı evrimsel yaklaşımların anormal agregasyonlar için uygun olmaması bu zorluğu arttırmaktadır. Ancak, fonksiyonel amiloidler, AlphaFold'un amiloid yapısının tahmininde kullanımı için bir başlangıç noktası olabilir.

Bir diğer çalışmada ise, Toronto Üniversitesi ve Insilico Medicine'den bilim insanları, ilaç keşfinde tanımlama sürecini hızlandırmak için AlphaFold'un ilk başarılı uygulamasını duyurdular (Ren ve ark., 2023). AlphaFold, Pandaomics ve Chemistry42 adlı yapay zeka destekli ilaç keşfi platformlarıyla birlikte bir hedefe yönelik yeni bir ilaç molekülünü tespit etmede uygulanmıştır. Bu yaklaşım, hedef molekül seçiminden hit bileşiği tespitine kadar tüm süreçleri, zaman ve maliyet açısından olumlu bir şekilde etkilemiştir. Hepatoselüler karsinom tedavisi için ilgi çekici bir protein olan CDK20'ye yönelik küçük bir hit bileşiği, bu yaklaşım ile sadece 30 günde ve sadece yedi bileşik sentezlenerek tespit edilmiştir. Bu önemli çalışma, yapay zekanın ilaç keşif sürecini dönüştürme ve hayat kurtaran tedavileri hastalara geleneksel yöntemlerden yıllar önce getirme gücünü göstermektedir.

COVID-19 bağlamında, AlphaFold, birkaç önemli SARS-CoV-2 viral proteini yapılarının tahmininde kullanılmıştır (Kryshtafovych, Moulton, ve ark., 2021). Bu proteinler arasında spike protein, ana proteaz ve RNA bağımlı RNA polimeraz bulunur (Sadek ve ark., 2021). Bu tahminler, COVID-19'a karşı tedavi ve aşılarda geliştirilmesine olanak sağlamıştır. Örneğin, Pfizer-BioNTech COVID-19 aşısı, SARS-CoV-2'nin spike proteinine yönelik hedeflenmiştir ve tasarımı, AlphaFold'un spike proteininin yapısının tahmini ile iyileştirilmiştir (Higgins, 2021). Benzer şekilde, virüsün ana proteazına yönelik etki eden birkaç küçük molekül ilaç, AlphaFold tahminleri temelinde geliştirilmiştir. Ayrıca, AlphaFold, virüsün insan konak hücreleriyle etkileşimine ilişkin, yapısal temeli anlamak için kullanılmıştır (Meng ve ark., 2022). Örneğin, AlphaFold tahminleri, viral giriş için temel olan insan reseptörü ACE2 (Anjiyotensin dönüştürücü enzimi 2) ile SARS-CoV-2 spike protein kompleksinin yapısını ortaya çıkarmıştır. Bu bilgi, viral enfeksiyonun mekanizmasına dair önemli bilgiler sağlamış ve bu etkileşimi hedefleyen tedavilerin geliştirilmesine yardımcı olmaktadır (Yang ve ark., 2022). Dolayısıyla, AlphaFold, tedavilerin ve aşılarda hızlı tasarım ve geliştirilmesine olanak sağlayarak COVID-19 ile mücadelede kritik bir rol oynamıştır.

Sivrisinek kaynaklı hastalık, sıtma, yıkıcı küresel etkilere sahip hastalıklardan biridir. Bu nedenle sıtmayı önlemeyi hedefleyen aşı ve gelişmiş tedavilerin bulunması oldukça önemlidir. COVID-19'dan farklı olarak, sıtma paraziti (*Plasmodium falciparum*) için aşı geliştirmesi, parazitin sahip olduğu yüzlerce yüzey proteini ve değişken yapısı nedeniyle oldukça zorlu bir süreçtir. Bu aşamada, yapay zeka destekli AlphaFold algoritmasının, hedeflenen protein yapılarını kısa zamanda, yüksek doğrulukta tahmin etmesi sıtma parazitine karşı aşı çalışmalarını temel bilim

aşamasından, klinik öncesi ve klinik geliştirme aşamasına çıkararak sürece hız kazandırmıştır (Ko ve ark., 2022).

Antimikrobiyal direnç (Antimicrobial resistance/AMR), tüm ekonomik ve sosyal düzeyleri etkileyen ve 2050 yılına kadar yaklaşık 10 milyon ölüme neden olabilecek küresel bir sağlık sorunudur. AMR sadece hayvan ve insan sağlığını değil, aynı zamanda bitki patojenleri nedeniyle yıllık küresel tarımsal ürün kayıplarının %40'ına kadarına da sebep olmaktadır. İnsan, hayvan ve bitki AMR patojenlerinin etkin bir şekilde kontrol edilmesi için acil önlemler alınması elzem olup, bu çerçevede yeni antibiyotik keşfinin başarı oranlarının artırılması için akademik ve endüstriyel sektörler arasında işbirliği gerekmektedir (Krell & Matilla, 2022). Bakterilerde enzimler tarafından kontrol edilen direnç mekanizmalarını (bakteri membran değişimleri) bloke etmek antibiyotiklerin işlevsel olmasında önemlidir. Ancak, deneysel yöntemlerle bu enzimlerin bileşimini anlamak son derece zorlu bir görev olarak ortaya çıkmaktadır. AlphaFold gibi teknolojiler, bakteriyel enzim yapılarının çok daha hızlı çözümlerini sağlayan bir başlangıç modeli sunar. AlphaFold aracılığı ile protein yapısı elde edildiğinde, bu yapı bilgisi proteinin kimyasal reaksiyonları nasıl katalize ettiğini anlamada kullanılabilir. Bu bilgi, bakteriyel antibiyotik direnci durumunda yeni terapötiklerin üretimi için kullanılabilir (Krell & Matilla, 2022).

Sonuç olarak, AlphaFold, ilaç keşfi/tasarımı ve hastalık mekanizmalarını anlamak gibi önemli biyomedikal araştırmalarda kullanılan güçlü bir araçtır. Proteinlerin doğru bir şekilde öngörülen yapıları, hastalıkların moleküler mekanizmalarının anlaşılmasına ve yeni ilaçların tasarlanmasına olanak tanır.

5.2 Yapısal Biyoloji

AlphaFold, protein yapı tahmininden ilaç keşfine kadar değişen uygulamalarla yapısal biyoloji alanında paha biçilmez bir araç haline gelmiştir. Yapısal biyolojide, protein yapılarının kesin olarak belirlenmesi, işlevlerini ve diğer moleküllerle olan etkileşimlerini anlamak için çok önemlidir. AlphaFold'un yapısal biyolojideki en önemli uygulamalarından biri, protein yapı tahminidir. AlphaFold, proteinlerin üç boyutlu yapısını doğru bir şekilde tahmin ederek, protein işlevinin mekanizmaları hakkında bilgi sağlar ve bu işlevleri test etmek için deneylerin tasarımına yardımcı olabilir (Akdal ve ark., 2022). AlphaFold'un doğruluğu ve hızı, onu yapısal biyoloji topluluğunda değerli bir araç haline getirmiştir ve birçok araştırmacı AlphaFold'u daha önce çözülmemiş proteinlerin yapılarını tahmin etmek için kullanmaktadır (Sezgin & Tekin, 2023; Zhang & Chen, 2022).

AlphaFold'un yapısal biyolojideki bir diğer önemli uygulaması, protein etkileşimlerinin incelenmesidir. Proteinler birbirleriyle ve diğer moleküllerle karmaşık şekillerde etkileşime girer ve bu etkileşimleri anlamak, işlevlerini anlamak için gereklidir (Evans ve ark., 2021). AlphaFold, protein komplekslerinin yapılarını tahmin etmek için kullanılabilir ve tek tek proteinler arasındaki etkileşimlerin ayrıntılı bir görünümünü sağlar (Yu ve ark., 2023). Protein-protein etkileşimlerini incelemek ve bu etkileşimleri hedefleyen yeni ilaçlar geliştirmek için DeepMind, yeni bir sürüm olan AlphaFold-Multimer'i geliştirmiştir. AlphaFold-Multimer, tek bir zincirden oluşan proteinlerin yapısını yüksek bir doğrulukla öngörebilen orijinal AlphaFold modelinin aksine, birden fazla zincirden oluşan protein komplekslerinin yapılarını öngörebilmek için özel olarak eğitilmiş bir AlphaFold protein katlanma tahmin modelidir (Evans ve ark., 2021). AlphaFold-Multimer, protein komplekslerinin içindeki zincirler arasındaki arabirimleri de içeren kompleksin tüm yapısını tahmin etmek için eğitilmiştir. Bu nedenle, orijinal AlphaFold modeline kıyasla AlphaFold-Multimer, komplekslerin yapısını daha doğru bir şekilde tahmin edebilir. Bu, protein komplekslerinin yapısını doğru bir şekilde tahmin etmek, biyolojik süreçleri anlamak ve yeni terapötikler geliştirmek için önemli olduğundan, protein yapı tahmin alanında önemli bir ilerlemedir.

AlphaFold, protein yapısı tahmini ve protein etkileşimlerinin incelenmesindeki uygulamalarının yanı sıra, yapısal biyolojinin çeşitli başka alanlarında da kullanılmaktadır. Örneğin, deneysel olarak incelenmesi zor olduğu bilinen zar proteinlerinin yapılarını tahmin etmek için kullanılabilir. Membran proteinleri, hücre zarında gömülü olan ve moleküllerin zarın içinden geçmesi veya hücre sinyalleşmesi gibi çeşitli biyolojik süreçlerde önemli rol oynayan proteinlerin bir sınıfıdır (Tan ve ark., 2008). Ancak, membran proteinlerinin yapılarını tahmin etmek, membranın hidrofobik ortamı nedeniyle yapısal biyolojide zor bir sorun olmuştur (Hong ve ark., 2022). Başlangıçta suda çözünebilir proteinler üzerinde geliştirilmiş ve eğitilmiş olsa da, son çalışmalar AlphaFold'un yüksek bir doğrulukla membran proteinlerinin yapılarını da tahmin edebildiğini göstermiştir (Dobson ve ark., 2023; Pei & Cong, 2023). Ek olarak, AlphaFold, genetik varyasyonun hastalığındaki rolünü anlamak için önemli olan mutasyonların protein yapısı ve işlevi üzerindeki etkilerini tahmin etmek için de kullanılabilir (Buel & Walters, 2022).

Yapısal biyolojideki en önemli gelişmelerden bir diğeri de, yapay zeka tabanlı sistemlerin kompleks ve en büyük proteinlerinden (megadalton/MDa protein kompleksi) biri olan nükleer por kompleksinin (Nuclear pore complex/NPC) yapısının tamamlanmasında kullanılmasıdır (Fontana ve ark., 2022). NPC, ökaryotik hücrelerin nükleer zarındaki moleküler geçiş noktasıdır ve çekirdek ile sitozol arasında biyomoleküllerin taşınımını düzenler. Omurgalı hücrelerde NPC'lerin moleküler kütleleri yaklaşık 110 ila 125 MDa arasında ve çapları yaklaşık 120 nm'dir. NPC'ler birçok biyolojik süreçte rol oynamalarıyla ilişkilendirilmiştir ve işlev bozuklukları giderek artan sayıda ciddi hastalıklarla ilişkilendirilmektedir. Son yirmi yılda birçok grubun öncü çalışmalarına rağmen, NPC'lerin organizasyonu, dinamikleri ve karmaşıklığı hakkında tam bir anlayışa sahip olunamamıştır. AlphaFold deneysel verilerle birleştirilerek önceki çalışmalara göre daha ayrıntılı bir NPC yapısının aydınlatılması sağlanmıştır (Fontana ve ark., 2022). Bu başarı, NPC işlevi, biyogenez ve düzenleme açısından önemli sonuçları olan bir gelişme olup, NPC içindeki moleküler etkileşimler hakkında anlayışımızı genişletmek ve NPC'nin tam moleküler mimarisini anlamak için önemli bir adım atmamızı sağlar.

5.3 Protein Mühendisliği

Protein mühendisliği, stabilite, aktivite ve özgüllük gibi özelliklerini geliştirmek için proteinlerin modifikasyonunu ve manipülasyonunu içerir. Proteinlerin üç boyutlu yapısını doğru bir şekilde tahmin etmek, amino asit kalıntıları ve proteinin katlanma modeli arasındaki etkileşimleri hakkında bilgi sağladığı için protein mühendisliğinde kritik öneme sahiptir. Protein yapısını doğru bir şekilde tahmin etme yeteneği, enzim aktivitesi, substrat özgüllüğü ve bağlanma afinitesi gibi belirli işlevlere ve özelliklere sahip proteinlerin tasarlanmasını sağlayabilir (Freschlin ve ark., 2022).

AlphaFold, belirli işlevlere ve özelliklere sahip yeni proteinlerin tasarımı da dahil olmak üzere protein mühendisliğinde çeşitli alanlara uygulanabilir. Örneğin, AlphaFold, gelişmiş katalitik aktiviteye, substrat özgüllüğüne ve stabiliteye sahip proteinler tasarlamak için kullanılabilir. Belirli işlevlere sahip proteinler tasarlama yeteneği, ilaç hedeflerinin belirli yapı ve işlevlere sahip proteinler olduğu ilaç keşfinde özellikle önemlidir. AlphaFold'un protein yapılarını doğru bir şekilde tahmin etme yeteneği, potansiyel ilaç hedeflerinin belirlenmesine yardımcı olabilir ve özellikle bu proteinleri hedefleyen ilaçların tasarlanmasını sağlayarak daha etkili ve daha güvenli ilaçlar üretilebilir (Ren ve ark., 2023).

İlaç keşfine ek olarak AlphaFold, biyokataliz ve biyoremediasyon gibi endüstriyel uygulamalar için enzimlerin geliştirilmesinde kullanılabilir. Biyokatalizörler, kimyasal reaksiyonları katalize eden ve biyoyakıt ve ilaç üretimi gibi çok çeşitli endüstriyel işlemlerde kullanılan enzimlerdir. AlphaFold, biyokatalitik süreçleri optimize etmek için artan aktivite ve kararlılık gibi belirli özelliklere sahip enzimler tasarlamak için kullanılabilir (Mazurenko ve ark., 2019).

Son olarak, AlphaFold yeni malzemelerin geliştirilmesinde kullanılabilir. Proteinler, malzemeler için çok yönlü yapı taşlarıdır ve özellikleri, dayanıklılık, esneklik ve biyouyumluluk gibi benzersiz niteliklere sahip malzemeler oluşturmak için tasarlanabilir. AlphaFold'un protein yapılarını doğru bir şekilde tahmin etme yeteneği, çeşitli uygulamalar için yeni malzemelerin geliştirilmesini kolaylaştırarak, belirli özelliklere sahip proteinlerin tasarlanmasını sağlayabilir (Miserez ve ark., 2023).

Sonuç olarak, AlphaFold, protein yapısı tahmininde devrim yaratması ile birlikte, protein mühendisliğinde de çeşitli uygulamaların gelişmesine katkı sağlayabilir. Protein yapılarını doğru bir şekilde tahmin etme yeteneği, ilaç keşfi, biyokataliz, biyoremediasyon ve malzeme biliminde önemli etkileri olan belirli işlev ve özelliklere sahip proteinlerin tasarlanmasına olanak tanımaktadır. Teknoloji gelişmeye devam ettikçe, AlphaFold şüphesiz yeni proteinlerin ve malzemelerin geliştirilmesinde kritik bir rol oynayacaktır.

6. ALPHAFOLD'UN SINIRLAMALARI

AlphaFold, DeepMind adlı yapay zeka araştırma kuruluşu tarafından geliştirilen bir protein katlanma tahmini sistemidir. Sistem, amino asit dizilerine dayanarak proteinlerin üç boyutlu yapılarını tahmin etmek için derin sinir ağları kullanır. AlphaFold, protein yapılarını tahmin etmedeki etkileyici doğruluğu nedeniyle geniş çapta övülmüştür, ancak diğer herhangi bir teknoloji gibi sınırlamaları vardır (Perrakis & Sixma, 2021). AlphaFold'un bir sınırı, yüksek bir hesaplama gücü gerektirmesidir. Son sürümü olan AlphaFold 2, tahminler üretmek için büyük miktarda GPU belleği ve hesaplama kaynakları kullanır. Bu, yüksek performanslı bilgi işlem kaynaklarına erişimi olmayan araştırmacılar için sistemi etkili bir şekilde kullanmayı zorlaştırabilir. Bu sınırlama AlphaFold geliştiricileri tarafından Google Colabs ile ortak çalışılarak giderilmeye çalışılmaktadır (Varadi & Velankar, 2022). AlphaFold'un ilk ve ikinci versiyonlarının kaynak kodları github (<https://github.com/deepmind/alphafold>) üzerinden elde edilebileceği gibi; aynı zamanda Google Colab (AlphaFold2.ipynb) üzerinden hiçbir sistem özelliği gerekmeden kullanıma açılmıştır.

AlphaFold'un diğer bir sınırı, büyük miktarda eğitim verisine ihtiyaç duymasıdır. Sistem, yaklaşık 170.000 protein yapılarından oluşan bir veri kümesinde eğitilmiştir (Wodak ve ark., 2022). Bu, var olan protein sayısı göz önüne alındığında oldukça küçük bir sayıdır. Bu nedenle, AlphaFold'un tahminleri daha az doğru olabilecek birçok protein olabilir. AlphaFold'un ayrıca, iyi anlaşılmayan protein yapılarının tahmin edilmesinde sınırlamaları vardır. Örneğin, benzersiz özelliklere sahip olan veya nadir hastalıklarla ilişkili olan proteinlerin yapılarını tahmin etmekte zorluk çekebilir. AlphaFold çoğu hedefte diğer yöntemlerden daha iyi performans göstermiş olsa da özellikle çoklu alanlı veya beta-yapıları zengin protein hedeflerinde iyi performans göstermediği bazı durumlar vardır. Dahası, karmaşık biyolojik süreçlerle ilişkili olan proteinlerin yapılarını tahmin etmekte zorlanabilir, çünkü bu yapılar hesaplamalı bir modelde hesaba katılmaları zor faktörlerden etkilenebilir (Evans ve ark., 2021). Ek olarak, AlphaFold'un tahminleri etkileyici olsa da her zaman mükemmel değildir. Özellikle modellenmesi zor olan proteinler için hala yanlış tahminler yapabilir. Bu nedenle, araştırmacılar AlphaFold'un tahminlerini X-ışını kristalografisi veya kriyo-elektron mikroskopisi gibi deneysel yöntemlerle doğrulamak zorunda kalabilirler.

7. ALPHAFOLD'UN GELECEĞİ VE SONUÇ

Sonuç olarak AlphaFold, benzeri görülmemiş bir doğruluk ve hız sunarak protein yapısı tahmininde büyük bir atılımı temsil etmektedir. Sınırlamalarına rağmen, diğer tüm yöntemlerden büyük bir farkla daha iyi performans gösterdiği CASP14 deneyinde doğrulanmıştır. AlphaFold'un

başarısı, derin öğrenme yaklaşımlarının gücünü vurgular ve bunların yapısal biyolojideki karmaşık sorunları çözmek için değerli araçlar olabileceğini öne sürer.

İleriye bakıldığında, AlphaFold'daki gelişmeler muhtemelen uygulanabilirliğini küresel ve küçük transmembran proteinlerin ötesinde diğer protein türlerine genişletmeye odaklanacaktır. Bu, protein-protein etkileşimleri veya katlanma sonrası değişiklikler hakkında bilgi gibi modele ek özelliklerin dahil edilmesini içerebilir. Algoritma ve yapay zeka mekanizmaları, veri kaynakları ve ön işleme yöntemleri, doğrulama ve performans değerlendirmesi, genelleştirilebilirlik ve uygulanabilirlik gibi alanlarda yapılacak daha fazla araştırma ve geliştirme, AlphaFold'un yeteneklerini genişletmek ve bilimsel anlamda büyük ilerlemeler kaydetmek için önemlidir. AlphaFold'un daha genelleştirilebilir, yorumlanabilir ve güvenilir modeller sağlaması, proteinlerin yapısını ve işlevini anlama sürecinde önemli bir adım olabilir. Bu potansiyel, biyoloji ve tıp alanlarında büyük ilerlemeler sağlama açısından büyük bir umut kaynağıdır.

Yazarların Katkısı

Yazarların makaleye katkıları eşit orandadır.

Çıkar Çatışması Beyanı

Yazarlar arasında herhangi bir çıkar çatışması bulunmamaktadır.

Araştırma ve Yayın Etiği Beyanı

Yapılan çalışmada araştırma ve yayın etiğine uyulmuştur.

KAYNAKÇA

- Akdel, M., Pires, D. E., Pardo, E. P., Jänes, J., Zalevsky, A. O., Mészáros, B., Bryant, P., Good, L. L., Laskowski, R. A. & Pozzati, G. (2022). A structural biology community assessment of AlphaFold2 applications. *Nature Structural & Molecular Biology*, 1-12.
- AlQuraishi, M. (2019). AlphaFold at CASP13. *Bioinformatics*, 35(22), 4862-4865.
- AlQuraishi, M. (2021). Machine learning in protein structure prediction. *Current opinion in chemical biology*, 65, 1-8.
- Anfinsen, C. & Scheraga, H. (1975). Experimental and theoretical aspects of protein folding. *Advances in protein chemistry*, 29, 205-300.
- Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science*, 181(4096), 223-230.
- Anfinsen, C. B., Haber, E., Sela, M. & White Jr, F. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences*, 47(9), 1309-1314.
- Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*, 22(2), 195-201.
- Bolen, D. & Baskakov, I. V. (2001). The osmophobic effect: natural selection of a thermodynamic force in protein folding. *Journal of molecular biology*, 310(5), 955-963.

- Borkakoti, N. & Thornton, J. M. (2023). AlphaFold2 protein structure prediction: Implications for drug discovery. *Current Opinion in Structural Biology*, 78, 102526.
- Buel, G. R. & Walters, K. J. (2022). Can AlphaFold2 predict the impact of missense mutations on structure? *Nature Structural & Molecular Biology*, 29(1), 1-2.
- Burley, S. & Petsko, G. (1988). Weakly polar interactions in proteins. *Advances in protein chemistry*, 39, 125-189.
- Burley, S. K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G. V., Duarte, J. M., Dutta, S., Fayazi, M. & Feng, Z. (2022). RCSB Protein Data Bank: Celebrating 50 years of the PDB with new tools for understanding and visualizing biological macromolecules in 3D. *Protein Science*, 31(1), 187-208.
- Chen, I.-M. A., Markowitz, V. M., Chu, K., Palaniappan, K., Szeto, E., Pillay, M., Ratner, A., Huang, J., Andersen, E. & Huntemann, M. (2016). IMG/M: Integrated genome and metagenome comparative data analysis system. *Nucleic acids research*, 507-516.
- Cheng, J., Roy, R. S., Liu, J., Giri, N. & Guo, Z. (2023). Combining pairwise structural similarity and deep learning interface contact prediction to estimate protein complex model accuracy in CASP15. *bioRxiv*, 2023.2003.2008.531814.
- de Almeida Paiva, V., de Souza Gomes, I., Monteiro, C. R., Mendonça, M. V., Martins, P. M., Santana, C. A., Gonçalves-Almeida, V., Izidoro, S. C., de Melo-Minardi, R. C. & de Azevedo Silveira, S. (2022). Protein structural bioinformatics: An overview. *Computers in Biology and Medicine*, 105695.
- Dhingra, S., Sowdhamini, R., Cadet, F. & Offmann, B. (2020). A glance into the evolution of template-free protein structure prediction methodologies. *Biochimie*, 175, 85-92.
- Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry*, 29(31), 7133-7155.
- Dill, K. A. & MacCallum, J. L. (2012). The protein-folding problem, 50 years on. *Science*, 338(6110), 1042-1046.
- Dobson, L., Szekeres, L. I., Gerdán, C., Langó, T., Zeke, A. & Tusnády, G. E. (2023). TmAlphaFold database: membrane localization and evaluation of AlphaFold2 predicted alpha-helical transmembrane protein structures. *Nucleic acids research*, 51(D1), 517-522.
- Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Židek, A., Bates, R., Blackwell, S. & Yim, J. (2021). Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2021.2010.2004.463034.
- Faure, A. J., Domingo, J., Schmiedel, J. M., Hidalgo-Carcedo, C., Diss, G. & Lehner, B. (2022). Mapping the energetic and allosteric landscapes of protein binding domains. *Nature*, 604(7904), 175-183.
- Fontana, P., Dong, Y., Pi, X., Tong, A. B., Hecksel, C. W., Wang, L., Fu, T.-M., Bustamante, C. & Wu, H. (2022). Structure of cytoplasmic ring of nuclear pore complex by integrative cryo-EM and AlphaFold. *Science*, 376(6598), 1-29.

- Freschlin, C. R., Fahlberg, S. A. & Romero, P. A. (2022). Machine learning to navigate fitness landscapes for protein engineering. *Current Opinion in Biotechnology*, 75, 102713.
- Gasic, A. G., Sarkar, A. & Cheung, M. S. (2021). Understanding protein-complex assembly through grand canonical maximum entropy modeling. *Physical Review Research*, 3(3), 033220.
- Gogoi, C. R., Rahman, A., Saikia, B. & Baruah, A. (2023). Protein Dihedral Angle Prediction: The State of the Art. *ChemistrySelect*, 8(5), e202203427.
- Goverde, C., Wolf, B., Khakzad, H., Rosset, S. & Correia, B. E. (2022). De novo protein design by inversion of the AlphaFold structure prediction network. *bioRxiv*, 2022.2012.2013.520346.
- He, K., Zhang, X., Ren, S. & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778.
- Hey, T., Butler, K., Jackson, S. & Thiyagalingam, J. (2020). Machine learning and big scientific data. *Philosophical Transactions of the Royal Society A*, 378(2166), 20190054.
- Higgins, M. K. (2021). Can we AlphaFold our way out of the next pandemic? *Journal of molecular biology*, 433(20), 167093.
- Hong, H., Choi, H.-K. & Yoon, T.Y. (2022). Untangling the complexity of membrane protein folding. *Current Opinion in Structural Biology*, 72, 237-247.
- Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. & Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3), 712-725.
- Hsu, C., Verkuil, R., Liu, J., Lin, Z., Hie, B., Sercu, T., Lerer, A. & Rives, A. (2022). Learning inverse folding from millions of predicted structures. *International Conference on Machine Learning*, 1-25.
- Jendrusch, M., Korbel, J. O. & Sadiq, S. K. (2021). AlphaDesign: A de novo protein design framework based on AlphaFold. *bioRxiv*, 2021.2010.2011.463937.
- Jiang, F., & Wu, Y.-D. (2014). Folding of fourteen small proteins with a residue-specific force field and replica-exchange molecular dynamics. *Journal of the American Chemical Society*, 136(27), 9536-9539.
- Jiang, L., Chakraborty, P., Zhang, L., Wong, M., Hill, S. E., Webber, C. J., Libera, J., Blair, L. J., Wolozin, B. & Zweckstetter, M. (2023). Chaperoning of specific tau structure by immunophilin FKBP12 regulates the neuronal resilience to extracellular stress. *Science Advances*, 9(5), eadd9789.
- Jisna, V. & Jayaraj, P. (2021). Protein structure prediction: conventional and deep learning perspectives. *The Protein Journal*, 40(4), 522-544.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A. & Potapenko, A. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589.

- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Tunyasuvunakool, K., Ronneberger, O., Bates, R., Židek, A. & Bridgland, A. (2020). AlphaFold 2. *In Fourteenth Critical Assessment of Techniques for Protein Structure Prediction (Abstract Book)*.
- Ko, K.-T., Lennartz, F., Mekhaieel, D., Guloglu, B., Marini, A., Deuker, D. J., Long, C. A., Jore, M. M., Miura, K. & Biswas, S. (2022). Structure of the malaria vaccine candidate Pfs48/45 and its recognition by transmission blocking antibodies. *Nature Communications*, 13(1), 5603.
- Krell, T. & Matilla, M. A. (2022). Antimicrobial resistance: progress and challenges in antibiotic discovery and anti-infective therapy. *Microbial Biotechnology*, 15(1), 70-78.
- Kryshtafovych, A., Moulton, J., Billings, W. M., Della Corte, D., Fidelis, K., Kwon, S., Olechnovič, K., Seok, C., Venclovas, Č. & Won, J. (2021). Modeling SARS-CoV-2 proteins in the CASP-commons experiment. *Proteins: Structure, Function, and Bioinformatics*, 89(12), 1987-1996.
- Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K. & Moulton, J. (2019). Critical assessment of methods of protein structure prediction (CASP)—Round XIII. *Proteins: Structure, Function, and Bioinformatics*, 87(12), 1011-1020.
- Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K. & Moulton, J. (2021). Critical assessment of methods of protein structure prediction (CASP)—Round XIV. *Proteins: Structure, Function, and Bioinformatics*, 89(12), 1607-1617.
- Kufareva, I. & Abagyan, R. (2012). Methods of protein structure comparison. *Homology Modeling: Methods and Protocols*, 231-257.
- Kuhlman, B. & Bradley, P. (2019). Advances in protein structure prediction and design. *Nature Reviews Molecular Cell Biology*, 20(11), 681-697.
- Levinthal, C. (1969). How to fold graciously. *Mossbauer spectroscopy in biological systems*, 67, 22-24.
- Li, F. & Du, Y. (2018). From AlphaGo to power system AI: What engineers can learn from solving the most complex board game. *IEEE Power and Energy Magazine*, 16(2), 76-84.
- Li, S., Wu, S., Wang, L., Li, F., Jiang, H. & Bai, F. (2022). Recent advances in predicting protein-protein interactions with the aid of artificial intelligence algorithms. *Current Opinion in Structural Biology*, 73, 102344.
- Li, Y., Liu, Y. & Yu, D.-J. (2023). Machine learning for protein inter-residue interaction prediction. *Machine Learning in Bioinformatics of Protein Sequences: Algorithms, Databases and Resources for Modern Protein Bioinformatics*, 183-203. World Scientific.
- Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O. & Shmueli, Y. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637), 1123-1130.
- Liu, W., Wang, G., Wang, Z., Wang, G., Huang, J. & Liu, B. (2022). Repurposing small-molecule drugs for modulating toxic protein aggregates in neurodegenerative diseases. *Drug discovery today*, 27(7), 1994-2007.

- Lumry, R. & Eyring, H. (1954). Conformation changes of proteins. *The Journal of physical chemistry*, 58(2), 110-120.
- Marks, D. S., Hopf, T. A. & Sander, C. (2012). Protein structure prediction from sequence variation. *Nature biotechnology*, 30(11), 1072-1080.
- Marx, V. (2022). Method of the year: Protein structure prediction. *Nature methods*, 19(1), 5-10.
- Mazurenko, S., Prokop, Z. & Damborsky, J. (2019). Machine learning in enzyme engineering. *ACS Catalysis*, 10(2), 1210-1223.
- Meng, B., Abdullahi, A., Ferreira, I. A., Goonawardane, N., Saito, A., Kimura, I., Yamasoba, D., Gerber, P. P., Fatihi, S. & Rathore, S. (2022). Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity. *Nature*, 603(7902), 706-714.
- Miserez, A., Yu, J. & Mohammadi, P. (2023). Protein-based biological materials: Molecular design and artificial production. *Chemical Reviews*, 123(5), 2049-2111.
- Miyazawa, T., Hiratsuka, Y., Toda, M., Hatakeyama, N., Ozawa, H., Abe, C., Cheng, T.-Y., Matsushima, Y., Miyawaki, Y. & Ashida, K. (2022). Artificial intelligence in food science and nutrition: a narrative review. *Nutrition Reviews*, 80(12), 2288-2300.
- Nussinov, R., Zhang, M., Liu, Y. & Jang, H. (2022). AlphaFold, artificial intelligence (AI), and allostery. *The Journal of Physical Chemistry B*, 126(34), 6372-6383.
- Ongie, G., Jalal, A., Metzler, C. A., Baraniuk, R. G., Dimakis, A. G. & Willett, R. (2020). Deep learning techniques for inverse problems in imaging. *IEEE Journal on Selected Areas in Information Theory*, 1(1), 39-56.
- Ovchinnikov, S., Kim, D. E., Wang, R. Y. R., Liu, Y., DiMaio, F. & Baker, D. (2016). Improved de novo structure prediction in CASP 11 by incorporating coevolution information into Rosetta. *Proteins: Structure, Function, and Bioinformatics*, 84, 67-75.
- Pakhrin, S. C., Shrestha, B., Adhikari, B. & Kc, D. B. (2021). Deep learning-based advances in protein structure prediction. *International Journal of Molecular Sciences*, 22(11), 5553.
- Pearce, R. & Zhang, Y. (2021). Deep learning techniques have significantly impacted protein structure prediction and protein design. *Current Opinion in Structural Biology*, 68, 194-207.
- Pei, J. & Cong, Q. (2023). AFTM: a database of transmembrane regions in the human proteome predicted by AlphaFold. *Database*, 2023, baad008.
- Perrakis, A. & Sixma, T. K. (2021). AI revolutions in biology: The joys and perils of AlphaFold. *EMBO reports*, 22(11), e54046.
- Pinheiro, F., Santos, J. & Ventura, S. (2021). AlphaFold and the amyloid landscape. *Journal of Molecular Biology*, 433(20), 167059.
- Ponting, C. P. & Russell, R. R. (2002). The natural history of protein domains. *Annual Review of Biophysics and Biomolecular Structure*, 31(1), 45-71.

- Ren, F., Ding, X., Zheng, M., Korzinkin, M., Cai, X., Zhu, W., Mantsyzov, A., Aliper, A., Aladinskiy, V. & Cao, Z. (2023). AlphaFold accelerates artificial intelligence powered drug discovery: efficient discovery of a novel CDK20 small molecule inhibitor. *Chemical Science*, 14(6), 1443-1452.
- Romero, P. A. & Arnold, F. H. (2009). Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology*, 10(12), 866-876.
- Sadek, A., Zaha, D. & Ahmed, M. S. (2021). Structural insights of SARS-CoV-2 spike protein from Delta and Omicron variants. *bioRxiv*, 2021.2012. 2008.471777.
- Salahuddin, P., Siddiqi, M. K., Khan, S., Abdelhameed, A. S. & Khan, R. H. (2016). Mechanisms of protein misfolding: Novel therapeutic approaches to protein-misfolding diseases. *Journal of Molecular Structure*, 1123, 311-326.
- Savytskyi, O. V., Sirmans, T. N., Coban, M. A., Weber, C. A., Murray, M. E. & Caulfield, T. R. (2023). Computational modeling and molecular mapping of serine protease inhibitor family A5 (SERPINA5) structure, associated with tau expression and Alzheimer's disease. *Biophysical Journal*, 122(3), 471a.
- Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Židek, A., Nelson, A. W. & Bridgland, A. (2020). Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792), 706-710.
- Sezgin, E. & Tekin, B. (2023). Molecular evolution and population genetics of glutamate decarboxylase acid resistance pathway in lactic acid bacteria. *Frontiers in Genetics*, 1-14.
- Shen, Y. & Bax, A. (2013). Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *Journal of biomolecular NMR*, 56, 227-241.
- Tan, S., Tan, H. T. & Chung, M. C. (2008). Membrane proteins and membrane proteomics. *Proteomics*, 8(19), 3924-3932.
- Thompson, J. D., Plewniak, F., Ripp, R., Thierry, J.-C. & Poch, O. (2001). Towards a reliable objective function for multiple sequence alignments. *Journal of Molecular Biology*, 314(4), 937-951.
- Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Židek, A., Bridgland, A., Cowie, A., Meyer, C. & Laydon, A. (2021). Highly accurate protein structure prediction for the human proteome. *Nature*, 596(7873), 590-596.
- UniProt: the universal protein knowledgebase in 2021. (2021). *Nucleic acids research*, 49(D1), D480-D489.
- UniProt: the Universal Protein knowledgebase in 2023. (2023). *Nucleic acids research*, 51(D1), D523-D531.
- Varadi, M. & Velankar, S. (2022). The impact of AlphaFold Protein Structure Database on the fields of life sciences. *Proteomics*, 2200128.
- Wang, G., Wu, Z., Fang, X., Xiang, Y., Liu, Y., Yu, D. & Ma, Y. (2022). Efficient AlphaFold2 Training using Parallel Evoformer and Branch Parallelism. *arXiv preprint arXiv:2211.00235*.

- Wang, Y. & Huang, R. (2023). Identification of Artemisia Argyi (AA) Therapy in Alzheimer's Disease (AD) Using Network Pharmacology and Molecular Docking. *Advanced Biology*, 2200256.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C. & Bordoli, L. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research*, 46(W1), W296-W303.
- Wodak, S. J., Vajda, S., Lensink, M. F., Kozakov, D. & Bates, P. A. (2022). Critical Assessment of Methods for Predicting the 3D Structure of Proteins and Protein Complexes. *Annual Review of Biophysics*, 52.
- Xu, Z., Cen, Y.-K., Zou, S.-P., Xue, Y.-P. & Zheng, Y.-G. (2020). Recent advances in the improvement of enzyme thermostability by structure modification. *Critical reviews in biotechnology*, 40(1), 83-98.
- Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S. & Baker, D. (2020). Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences*, 117(3), 1496-1503.
- Yang, Q., Jian, X., Syed, A. A. S., Fahira, A., Zheng, C., Zhu, Z., Wang, K., Zhang, J., Wen, Y. & Li, Z. (2022). Structural comparison and drug screening of spike proteins of ten SARS-CoV-2 variants. *Research*.
- Yu, D., Chojnowski, G., Rosenthal, M. & Kosinski, J. (2023). AlphaPulldown—a python package for protein–protein interaction screens using AlphaFold-Multimer. *Bioinformatics*, 39(1), btac749.
- Zhang, J. & Chen, B. (2022). Fighting SARS-CoV-2 with structural biology methods. *Nature methods*, 19(4), 381-383.
- Ziegler, S. J., Mallinson, S. J., John, P. C. S. & Bomble, Y. J. (2021). Advances in integrative structural biology: Towards understanding protein complexes in their cellular context. *Computational and Structural Biotechnology Journal*, 19, 214-225.
- Zwanzig, R., Szabo, A. & Bagchi, B. (1992). Levinthal's paradox. *Proceedings of the National Academy of Sciences*, 89(1), 20-22.