Ayrık Yapay Arı Kolonisi Algoritması ile Protein Yapısı Tahmini

Turgay BATBAT[1], Celal ÖZTÜRK[2]

[1]Erciyes Üniversitesi, MF, Biyomedikal Mühendisliği
[2]Erciyes Üniversitesi, MF, Bilgisayar Mühendisliği
turgaybatbat@erciyes.edu.tr, celal@erciyes.edu.tr
(Geliş/Received: 21.04.2016; Kabul/Accepted: 24.09.2016)
DOI: 10.17671/btd.97757

Özet— Proteinlerin yerine getirdiği görevler üçüncül yapıları olarak adlandırılan doğal durumlarına doğrudan bağlıdır. Günümüzde birçok proteinin birincil yapı olarak adlandırılan dizilimleri bilinmekle beraber işlevlerini belirleyen üçüncül yapıları bilinmemektedir. Ele alınan dizide bulunan amino asitlerin üçüncül yapı oluşumuna yol açan özelliklerinin kullanılarak üçüncül yapılarının tahmin edilmesi proteinlerin gerçek doğasının tam anlaşılmasının önünü açacaktır. Proteinlerin üçüncül yapıları ile ilgili yapılan çalışmalarda en düşük serbest enerji düzeyinin elde edilmesi eğilimi tespit edilmiştir. Serbest enerji düzeylerini hesaplamada, sudan kaçma özelliği yoğun olarak kullanılmaktadır. Bu özelliğe göre sınıflandırılan amino asitleri temel alan HP modeli basitliği ve gerçekçiliği ile dikkat çekmektedir. Çalışmamızda HP modelinin doğadaki protein davranışı incelenerek geliştirilmiş bir türü olan Üç Boyutlu Hidrofobik Polar Yan Zincir modeli kullanılmıştır. Proteinlerin üçüncül yapı modellerinin bulunmasında arıların besin arama davranışlarını modelleyen Yapay Arı Kolonisi Algoritması temelli oluşturulan ayrık bir model önerilmiştir. Benzer çalışmalarda kullanılan başarımların test edildiği örnek dizilerde daha az maliyetle iyi sonuçlar alınmıştır.

Anahtar Kelimeler— Biyoenformatik, Protein Yapı Tahmini, Yapay Arı Koloni Algoritması, Ayrık Optimizasyon

Protein Structure Prediction with Discrete Artificial Bee Colony Algorithm

Abstract— Tasks that proteins perform are directly related to native conditions named their tertiary structure. Although arrays of numerous proteins' primary structure are known, tertiary structures those explain their functions are not well defined. Studies on translation of proteins to their tertiary structure found tendency of provision of minimal free energy level. HP model, used in calculation of free energy levels and based on the amino acids grouped by their hydrophobic characteristics, is standing out with its simplicity and authenticity. In this paper, three-dimensional hydrophobic side chain model that was developed by investigation of the protein behavior of HP model in nature was used. A new discrete model based on Artificial Bee Colony Algorithm that is simulating the foraging behaviors of bees is proposed to form the protein tertiary structures. The success of proposed model is tested on widely used benchmark protein arrays and compared with the results of similar methods from literature.

Keywords— Bioinformatics, Protein Structure Prediction, Artificial Bee Colony Algorithm, Discrete Optimization

1. GİRİŞ (INTRODUCTION)

Proteinlerin doğal yapılarına katlanmaları işlemi çok kısa sürede aşamalı olarak gerçekleşmektedir. Gereken faydanın sağlanması açısından asıl ilgilenilen süreç değil proteinlerin son halinin bilinmesidir. Deneysel olarak proteinlerin son hallerini gözlemek gelişmiş laboratuvarlar ve gözlenecek proteinin yapısını bozmadan saflaştırılmasını gerektirmektedir ki bu çok zor ve uzun bir süreçtir [1]. Bu nedenle, bilinen kurallar kullanılarak sonucun bilgisayarlar tarafından tahmin edilmesi günümüzde araştırmacıların en çok yoğunlaştıkları konulardandır.

Kesin sonuca ulaşmanın maliyetli olduğu durumlarda bilgisayar bilimcilerin en çok başvurdukları yöntem, birçok zor problemi başarıyla çözen Yapay Zekâ teknikleridir. Proteinlerin doğal yapısının en düşük serbest enerjili durumu olduğu bilgisi, problemin optimizasyon problemi olarak ele alınabileceği gerçeğini ortaya koymuştur. Bu tip çalışmalarda Anfisen teorisini kullanarak en düşük serbest enerjili yapının bulunması çok yaygındır. Moleküller üzerine çalışan bilim adamlarının ortaya koyduğu termodinamik yasalarına dayanan hesaplamalar ve yine proteinlerin gerçek yapıları temel alınarak oluşturulan gösterimler hesaplama maliyetlerini artırmaktadır. Bu nedenle, protein yapı taşı

olan amino asitleri H (sudan kaçınan) ve P (kutuplu) olarak iki sınıfa bölen ve kafes şeklinde bir yapının köşelerine yerleştiren HP modeli önem kazanmaktadır. Bu model hem proteinlerin doğasına uyar hem de anlaşılması ve uygulanması basit bir yöntemdir [2, 3].

Bu çalışmada HP modelin daha karmaşık ve daha ileri bir türü olan Üç Boyutlu Hidrofobik Polar Yan Zincir (3DHP-SC) modeli kullanılmıştır ve protein modellerinin oluşturulması için arıların besin arama davranışlarını modelleyen Yapay Arı Kolonisi Algoritması temelli yeni bir ayrık model önerilmiştir. Önerilen algoritmanın performansı literatürde mevcut yöntemlerin başarımları ile test edilmiştir ve örnek dizilerde daha az maliyetle iyi sonuçlar kaydedilmiştir. Makalenin organizasyonu şu şekildir; 2. bölümde benzer konularda yapılmış literatürde yayınlar verilmiştir. 3. bölümde makalenin teorik alt yapısında önemli bir yer kaplayan proteinler yapı ve model bakımından incelenmiştir. 4. bölümde çalışmalarda kullanılan Yapay Arı Kolonisi Algoritması (ABC) ve önerilen ayrık modeli anlatılmıştır. 5. bölümde de kullanılan 3DHP-SC modeli, ele alınan veri seti ve deneysel çalışmalar sonuçlarla birlikte verilmiştir.

2. LİTERATÜR ÖZETİ (LITERATURE REVIEW)

Protein yapı tahmininde Karami ve arkadaşları Bal Arısı, Yapay Arı Kolonisi, Parçacık Sürü Optimizasyon ve Klonal Seçilim algoritmalarına ve hibritlerine değinmişlerdir [4]. Benzer bir çalışma Ye tarafından yüksek lisans tezi olarak yapılmıştır. Bu çalışmada Protein Yapı Tahmin sürecinde evrimsel algoritma ile yapılan çalışmalarda karşılaşılan güçlüklere değinilmiştir [5]. Paluszewski, protein yapıları için kesin sonucu bilinmeyen durumlarda yöntemlerin değerlendirilmesi amacıyla geliştirilmiş kriterler üzerinde durmuştur[6]. Sistemin oluşturulmasında sonsuz çözüm içeren uzayın ortaya çıkaracağı zorluklarla ilgilenebilmek için ayrıklaştırma işlemi üzerinde durulmuştur. Zhang ve arkadaşları, Genetik Algoritmanın temel operatörlerinin başarımını artırmak için Tabu Arama algoritmasını temel alarak geliştirmeler yapmışlardır. Geliştirilen yeni yöntem özel olarak oluşturulmuş dört Fibonacci dizisine ve üç gerçek protein dizisine uygulanmış ve aynı diziler üzerinde yapılan Tabu Arama ve Evrimsel Algoritma türevleri ile karşılaştırılmıştır [7]. Genetik Algoritma ve Tabu Arama algoritmasından oluşturulmuş farklı bir yapı da iki boyutlu olarak Jiang ve arkadaşları tarafından HP modeli kullanılarak Protein Yapı Tahmini problemine uygulanmıştır. Elde edilen sonuçlar Genetik Algoritma, Evrimsel Monte Carlo ve Monte Carlo yöntemleri ile karşılaştırılmıştır [8].

AB kafessiz modelinin kullanıldığı bir çalışmada Chen ve arkadaşları tarafından önerilen Parçacık Sürü Optimizasyon algoritması üzerinde rastgele hareketlerin Levy-stable dağılımı ile elde edilmesi sağlanarak altı Fibonacci dizisi ve dört gerçek protein için sonuçlar elde edilmiştir. Bu sonuçlar PSO ile karşılaştırılmıştır [9]. Garza-Fabre ve arkadaşları, Protein Yapı Tahmin probleminde HP modeli için kullanılan enerji fonksiyonları ve gösterimlerinin etkili olabilmesi için yeni önerilerde bulunarak bu önerileri test etmişlerdir [10]. Lin ve Su, 3DHP modelinin kullanıldığı ve çözümlerin Parçacık Sürü Optimizasyon ve Genetik Algoritma kullanılarak oluşturulan hibrit bir yapıyla elde edildiği bir çalışma yapmışlardır. Çözümler, örnek diziler üzerinde algoritma 50000 adım koşularak elde edilmiştir [11].

İki boyutta koordinatlar açısından benzer ifade şeklini kullanan bir çalışma Chira tarafından yapılmıştır [12]. Çalışmada parmak izi ayrıklaştırma olarak adlandırılan katsayılarla tüm etkileşimlerin hesaba katıldığı bir HP modeli enerji fonksiyonu olarak kullanılmıştır. Çözüm için Evrimsel Algoritma temeli üzerine yerel tepe tırmanma yönteminin eklenmesi ile oluşturulan bir yöntem kullanmıştır (ELF). Elde edilen değerler evrimsel algoritma türevleri ile karşılaştırılmıştır. Johnson ve Katikireddy protein yapı tahmini için HP modeli üzerinde Genetik Algoritma kullanan diğer araştırmacılardır [13]. Unger ve Moult tarafından yapılan çalışma ve Patton ve arkadaşları tarafından yapılan çalışma ile sonuçları karşılaştırılmıştır.

Zydallis ve arkadaşları, Harvard Üniversitesi araştırmacıları tarafından geliştirilen ve molekül bazlı hesaplamalar için genel olarak kullanılan CHARMm enerji fonksiyonunu ikiye ayırmışlar gerçek protein yapısı üzerinden kendi geliştirdikleri Çok Amaç Fonksiyonlu Hızlı Dağınık Genetik Algoritma (MO fmGA) ile dihedral açıları buldurmuşlardır [14]. Oluşturulan yöntemler Met-Enkhephelin ve Polyalanine proteinleri üzerinde çalıştırılmış ve yarışmacı kalıp türüne göre referans alınan fmGA çalışması ile karşılaştırılmıştır. Bahamish, Abdullah ve Salam yaptıkları çalışmada, Protein Yapı Tahmini probleminde ABC algoritmasını kullanmışlardır. Çalışmada proteinin gösterimi torsiyon açıları kullanılan tam atom modelidir. Torsiyon açılarındaki değişimlerle protein yapıları elde edilir. Yayında SMMP moleküler dinamik paketi içindeki ECEPP/2 güç alanları denklemi enerji fonksiyonu olarak kullanılmıştır. Başlangıç çözümleri oluşturulduğunda SMMP paketi içinde gelen otomatik yöntemlerle çözümler başlangıçta iyileştirilmişlerdir [15]. Lin ve Su yaptıkları çalışmada iki ve üç boyutlu HP kafes modeli üzerinde ABC algoritmasının problem için düzenlenmiş bir tipi olan Düzenlenmiş Yapay Arı Kolonisi (MABC) algoritmasını kullanmışlardır [16].

Yakın zamanda yapılan çalışmalarda 3DHP kafes modeli gösterimiyle Markov Monte Carlo ve Karınca Koloni otimizasyon (ACO) [17], Ateş Böceği (FA) [18] ve Genetik Algoritma (GA) [19] bazlı yaklaşımların önerildiği görülmektedir. [20]'de ise bu çalışmadakine benzer 3DHP-SC gösterimi kullanılmıştır. Ancak, enerji hesaplamaları sadece HH bağ sayısı üzerinden yapılmıştır. Problemi farklı ve daha özel yönleri ile ele alan ek kısıtları ortadan kaldırma ve uzayı yeniden biçimlendirme, yeni bir gruplama yöntemi belirleme ve özel kuantum cihazlarında protein yapısı çözüm işlemine

ön bilgi oluşturma çalışmalarına [21–23]'de rastlanmaktadır.

3. PROTEIN YAPILARI ve MODELLERİ (PROTEIN STRUCTURES AND MODELS)

3.1. Proteinler (Proteins)

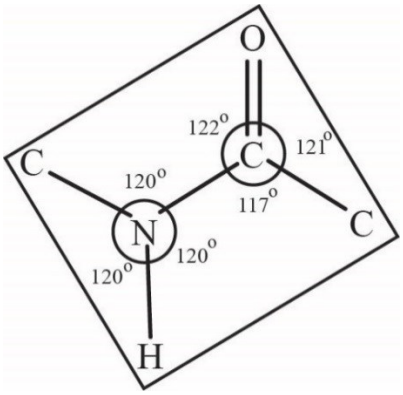
Proteinler canlılarda en fazla bulunan makro moleküllerdir. Proteinlerin birincil yapısı DNA tarafından denetlenir. Proteinler arası fark amino asit dizilimlerindeki farktan ileri gelir ve amino asitlerin farklı dizilimleriyle sayılamayacak kadar farklı protein elde edilebilir.

3.2. Protein Sentezi (Protein Synthesis)

DNA'dan RNA aracılığıyla alınan kodlar uygun amino asitleri seçer ve bu amino asitler arasında peptit bağları kurularak protein sentezlenir. Amino asitlerin seçimi için DNA'daki üç baza karşılık gelen kodlar kullanılır. Bu yapıya kodon ismi verilmektedir. Protein sentezine genellikle translasyon veya tercüme adı verilir. Protein sentezi esnasında canlının genetik kodunda bulunan dört baz, yirmi amino aside tercüme edilir. Sentez bitişiyle tüm elemanlar serbest bırakılır ve proteinler görevlerini yerine getirebilecek yapıya kavuştuktan sonra görev yerlerine giderler. Protein sentezi, hücrede çok hızlı gerçekleştirilen bir faaliyettir. Bu duruma rağmen protein sentezinin farklı adımları belirli bir sıra dâhilinde izlediği bilinmektedir [24].

3.3. Protein Yapıları (Protein Structures)

Protein yapıları, proteinlerin görevlerini ifa edecek üç boyutlu şekli almalarını açıklayan bir terimdir. Proteinlerde peptit bağ atomları Şekil 1'de görüldüğü gibi bağın düzlemine sabitlenmişlerdir. Proteinlerin esneklikleri, omurgadaki C_α atomuyla olan bağlardan gelir. Genel olarak protein molekülündeki her amino asit omurga içindeki hareketliliği ifade eden ϕ ve ψ açıları ile ifade edilirler. Ancak bu gösterimde yan zincir bilgisi eksiktir ve yan zincirlerin serbestlik dereceleri daha yüksek olabilmektedir. Bu iki açı ile birlikte, cis – trans durumunu ifade eden bir açı ve yan zincirler için χ_i olarak adlandırılan sekiz açı kullanılmaktadır. Polipetit zincirinin protein yapısını kazanması işleminde Gibbs'in minimum serbest enerji aranma fikri kabul görmektedir [12]. Protein yapıları genel olarak dört başlık altında incelenmektedir [1].



Şekil 1. Rijit peptit düzlemi
(Rigit peptide plane)

a) Protein birincil yapıları (Protein primary structures)

Protein birincil yapısı, proteinin kimyasal içeriğini gösterirken yerine getireceği fonksiyon hakkında bilgi vermemektedir. Amino asitler herhangi bir katlanma göstermeden sadece dizilim sıraları ile ifade edilirler. Protein birincil yapıları kendinden sonraki yapıların temelini oluşturur. Birincil yapıdaki amino asitlerin fiziksel özellikleri protein sentezinden hemen sonra üç boyutlu olarak protein zincirinin şekillenmesine neden olur. Birincil yapılar sonraki yapı türleri için kaynak noktası olduğundan sonraki yapıların tahmininde birincil yapıların kullanıldığı birçok yöntem bulunmaktadır. Gerçek birincil yapıya uygun bir örnek Şekil 2'de görülmektedir.

A	D	Q	A	Q	R	A	L	M	K	L	A	D	A

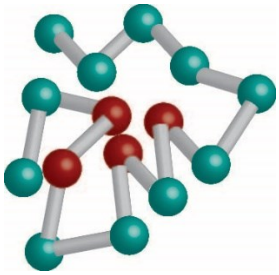
Şekil 2. Protein birincil yapı örneği
(Protein primary structure example)

b) Protein ikincil yapısı (Protein secondary structure)

Proteinlerin ikincil yapıları amino asitlerin etkileşerek oluşturdukları lokal yapılardır. Lokal olarak ele alındıklarından, ikincil yapılar tersiyer yapı tahmininden çözümü çok daha kolay bir problemdir. İkincil yapılar daha sonra etkileşerek daha üst yapıları oluştururlar. Bir grup ikincil yapının birlikte ele alınması ile oluşan yapıya motif, daha genel olarak ele alındığında karşılaşılan yapıya kat ismi verilmektedir.

c) Protein üçüncül yapıları (Protein tertiary structure)

Proteinlerin görevlerini açıklayan ve doğada tekil olarak bulundukları üç boyutlu yapıdır. Proteinleri tersiyer yapılarında kararlı halde tutan etkileşimler; hidrojen bağları, disülfit bağları, iyonik etkileşimler, Van der Waal's kuvvetleri, kutupsuz yan zincir etkileşimleridir. Canlı organizmada görev yapan proteinler Şekil 3'te örnek yapısı verilen üçüncül yapıları ile bulunurlar.



Şekil 3. Top çubuk yapısı ile protein tersiyer yapısı (Protein tertiary structure with ball stick model)

d) Protein kuaterner yapıları (Protein quaternary structure)

Proteinler tek polipeptit zincirinden oluşabileceği gibi bazı durumlarda iki veya daha fazla polipeptit zincirinden meydana gelebilirler. Bir araya gelerek protein oluşturan bu polipeptit zincirleri monomer olarak adlandırılırlar. İşte bu birden fazla tip proteinlerde monomorler bir araya gelerek yeni bir yapı oluştururlar. Oluşan bu yapı kuaterner yapı olarak isimlendirilir.

3.4. Protein Yapı Tahmini (Protein Structure Prediction)

Protein yapı tahmini, sadece amino asit dizisi bilinen polipeptitlerin işlevselliğini sağlayan doğal yapılarını tahmin etme işlemidir. Protein yapısını etkileyen yegâne bilgi amino asit dizisidir. Bu bilgi ışığında protein dizisini alarak tersiyer yapısını çıktı olarak veren algoritmalar geliştirilmektedir. Tahmin işleminde temel alınan kabul Anfisen'in termodinamik hipotezidir [25]. Protein doğal yapısını kazanırken temel değişikleri C_α etrafındaki bağlarda gerçekleşir. Bu işlem gerçekleşirken oluşan olaylar bilinmemektedir. Bu nedenle farklı yöntemler kullanılarak enerji sahası araştırılmakta ve bulunan sonuçlarda en düşük serbest enerjili katlanma seçilmektedir [26].

Proteinler için mümkün katlanma sayısının astronomik boyutlara ulaşmasından dolayı rastgele arama ile doğru yapının bulunması imkânsızdır. Bu duruma Levinthal Paradoksu denir [27]. Proteinler katlanırken büyük enerji engellerini aşmayacak şekilde ilerledikleri kabulü büyük enerji değişimleri olmadan sürecin modellenmesi genel kabulünü getirmiştir.

Proteinlerin doğal yapılarını nasıl aldıklarından ziyade ihtiyaç duyulan bilgi yapının kendisidir. Çözümde karşılaşılan ilk problem, kuantum mekaniği kurallarına göre katlandığı düşünülen proteinlerin enerji fonksiyonlarının belirlenmesinin şu an için imkânsız olmasıdır. Çalışmalarda bağ enerjisi ve elektrostatik kuvvetler gibi birçok farklı faktörün belirli katsayılarla toplanarak enerji fonksiyonlarını elde etme eğilimi yüksektir. En çok kullanılan yöntem ise amino asitlerin suyla etkileşimini ele alan yöntemlerdir.

Çözümü engelleyen önemli durumlardan bir diğeri ise çözüm uzayının kaba bir şekilde incelenmesi mümkün olmayan büyüklükte olmasıdır [28]. Bu nedenle çalışmalarda arama bölgesini en aza indirgenmesi amacı güdülmektedir. Bu yüzden çözüm uzayının sadece iyi çözümlerin bulunma ihtimalinin fazla olduğu kısımlarının aranması gerekmektedir. Enerji hesabında tam doğru sonucun bulunması yerine göreceli olarak hesaplama yeterli görülmektedir. Bu nedenle sabit değerlerin işleme katılmasına gerek yoktur ve katsayılarla enerji terimlerinin ağırlıklandırıldığı güç alanları kullanım için oldukça uygundur. Güç alanları birçok parametreye bağlı olduğu ve kesin enerji değerleri hesaplanmadığı için çalışmalarda aynı parametreler kullanılmadan karşılaştırmalar yapılamaz ki bu nedenle skorlar yerine tahminleri karşılaştırmak daha uygundur [29].

3.5. Protein Modelleri ve Hesaplama Yöntemleri(Protein Models and Calculation Techniques)

Yapılan çalışmalar protein yapı tahmin problemlerinde detaylı gösterimlerin karmaşık olan problemi daha içinden çıkılmaz hale getirdiği sonucunu ortaya koymuştur. Protein gösteriminde amaç mümkün olan en basit şekilde doğadaki proteinlere en yakın gösterimi elde etmektir. Sık kullanılan modeller iki kategoriye ayrılmaktadır; bunlar kafes modelleri ve kafessiz modellerdir. Kafes modellerinde bir ızgara bulunur ve yapısal elemanlar ızgaranın köşelerine yerleştirilirler. Kafessiz modellerde yapısal elemanlar sürekli uzayda konumlanırlar. İki grup içinde en çok kullanılan modeller HP kafes ve AB kafessiz modelleridir [10].

a) HP Kafes Modeli (HP Lattice Model)

HP kafes modeli basitliği ve anlaşılırlığının yanı sıra oluşturduğu modellerin kolay çözümlenebilmesi açısından çalışmalarda en çok tercih edilen modeldir. İki ve üç boyutlu olarak birçok farklı çeşidi geliştirilmiştir. En basit şekli iki boyutlu şeklidir. Genel olarak kare kafes yapısı tercih edilmektedir. HP modelinde, yirmi amino asit hidrofobik (H) ve kutuplu (P, hidrofilik) olarak iki sınıfa ayrılır. Bu ayrım su ile etkileşime girme eğilimleri üzerine belirlenmektedir. HP ile protein yapı tahmini, iki boyutlu kare ve üç boyutlu küp kafes için NP-Zor problem olarak sınıflandırılmıştır [10].

b) AB Kafessiz Modeli (AB Off-lattice Model)

AB kafessiz modeli, HP ile temel kurallar açısından ciddi benzerlikler gösterir. Amino asitler bu modelde de hidrofobik (A) ve kutuplu (B) olmak üzere iki sınıfa ayrılır. Amino asitler sürekli uzayda olduğu için konum hesaplarında bazı farklılıklar görülmektedir. Bu modelde monomerlerin birim uzunluktaki bağlarla polimer oluşturduğu kabulü yapılır. Enerji fonksiyonu Eşitlik (1)'de görülmektedir [9].

$$E = \sum_{i=2}^{n-1} \frac{1}{4}(1 - cos\theta_i) + \sum_{i=1}^{n-2} \sum_{j=i+2}^{n} 4[r_{ij}^{-12} - C(\xi_i, \xi_j) r_{ij}^{-6}] \quad (1)$$

Burada θ_i $(0 \leq \theta_i \leq \pi)$, iki ardıl bağ vektörü arasındaki açı; r_{ij}, i ve j kökleri arası uzaklık $(i < j)$; $C(\xi_i, \xi_j)$ ele alınan çiftlerin türüne göre değer alan sabittir (AA, BB ve AB için sırasıyla $+1, +\frac{1}{2}, \frac{1'}{2}$ dir). r_{ij} hesaplanırken üç boyutlu düzlemde θ bağ açıları ve β burulma açıları kullanılır. Üzerinde değişiklik yapılarak en iyi sonuç bulunmaya çalışılan parametreler bu açılardır. Her n amino asitli model $(n-2)$ bağ açısı ve $(n-3)$ burulma açısı ile ifade edilir. Konumların Kartezyen koordinatlarına dönüşümü Eşitlik (2)'deki gibidir.

$$pos(i) = \begin{cases} (0,0,0), i = 1 \\ (0,1,0), i = 2 \\ (cos(\theta_1), sin(\theta_1), 0), i = 3 \\ \begin{pmatrix} pos(i-1)_x + cos(\theta_{i-2}) cos(\beta_{i-3}), \\ pos(i-1)_y + sin(\theta_{i-2}) cos(\beta_{i-3}), \\ pos(i-1)_z + sin(\beta_{i-3}) \end{pmatrix}, 4 \leq_i \leq n \end{cases} \quad (2)$$

4. YAPAY ARI KOLONİSİ ALGORİTMASI (ARTIFICIAL BEE COLONY ALGORITHM)

4.1. Temel Yapay Arı Kolonisi Algoritması (Basic Artificial Bee Colony Algorithm)

Arıların doğadaki yiyecek arama davranışı doğadaki zeki süreçlerden birisidir. Bu süreç modellenerek çözümü klasik yöntemlerle zor olan problemlerin çözülebileceği Karaboğa tarafından ortaya atılmış ve süreci modelleyen Yapay Arı Kolonisi (Artificial Bee Colony) algoritması geliştirilmiştir [30]. Modelde yiyecek kaynaklarının yerleri optimizasyon probleminin olası çözümlerine, kaynakların nektar miktarları da bu çözümlerin probleme uygunluk değerlerine karşılık gelmektedir. Yapay Arı Kolonisi algoritmasında, ilk olarak kâşif arılar birer kaynak bulur ve görevli arı olurlar. Buldukları kaynaktan aldıkları nektarla kovana dönerler. Kovanda bekleyen gözcü arılar tüm kaynaklarla ilgili bilgileri aldıktan sonra bu bilgileri değerlendirip seçtikleri bir kaynağa yönelirler. Nektarı tükenen kaynaktaki görevli arı kâşif arı olup yeni bir kaynağı rastgele aramaya başlar [31].

Başlangıç çözümleri Eşitlik (3)'teki gibi kaynağı etkileyen parametrelerin alt ve üst sınırları arasında rastgele bir değerle ifade edilir.

$$x_{ij} = x_j^{min} + rand(0,1)(x_j^{max} - x_j^{min}) \quad (3)$$

Formülde i, $i = 1, \ldots, SN$ olacak şekilde kaynağı ve o kaynağa yönelen görevli arıyı ifade eder. j, $j = 1 \ldots D$ olacak şekilde tüm kaynakların parametrelerini gösteren indislerdir. SN, toplam kaynak sayısını; D, toplam parametre sayısını ifade eder. x_j^{max} ve x_j^{min}, j parametresinin alt ve üst sınırlarıdır. Ayrıca her kaynak için geliştirilememe sayısı olarak $failure_i$ sayacı başlangıçta sıfırlanır [32].

Görevli ve gözcü arılar nektar alınan kaynağın etrafında arama yaparak o kaynakların nektar miktarını kendi bulundukları kaynakla karşılaştırırlar. Etraftaki kaynakları araştırma işleminin matematiksel ifadesi Eşitlik (4)'te verilmiştir.

$$v_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{kj}) \quad (4)$$

$$v_{ij} = \begin{cases} x_j^{min} & , \quad v_{ij} < x_j^{min} \\ v_{ij} & , \quad x_j^{min} \leq v_{ij} \leq x_j^{max} \\ x_j^{max} & , \quad v_{ij} > x_j^{max} \end{cases} \quad (5)$$

$\vec{x}_i$ ile belirtilen her kaynaktan rasgele seçilen bir parametrenin değeri değiştirilerek $\vec{v}_i$ oluşturulur. Bu değiştirme işleminde rastgele seçilen k. çözüm kullanılmaktadır. Bu şekilde kaynaklar birbirine yaklaştıkça adım büyüklüğü küçülerek daha yoğun bir arama gerçekleşecektir. ϕ_{ij}, $[-1,1]$ aralığında rastgele bir sayıdır. Sınırları aşan çözümler sınırlara Eşitlik (5)'deki gibi sabitlenir. Elde edilen kaynakların nektar değerleri hesaplanarak Eşitlik (6) ile hesaplanarak Yapay Arı Kolonisi algoritmasında kullanılır. Elde edilen uygunluk değerleri ile açgözlü bir seçilim yapılır ve eğer eldeki çözüm yeni çözümle değiştirilirse $failure_i$ değeri sıfırlanır yoksa bu değer bir artırılır. Gözcü arıların kaynağı seçme olasılıkları Eşitlik (7)'de verildiği gibi hesaplanır.

$$fitness_i = \begin{cases} \frac{1}{1+f_i} & , \quad f_i \geq 0 \\ 1 + abs(f_i) & , \quad f_i < 0 \end{cases} \quad (6)$$

$$p_i = \frac{fitness_i}{\sum_{j=1}^{SN} fitness_j} \quad (7)$$

Tüm görevli ve gözcü arılar bir arama süreci çevrimini tamamladıktan sonra geliştirilemeyen kaynaklar kontrol edilir. Bu işlem için kullanıcı tanımlı limit isimli parametreyi aşan $failure_i$ değerine sahip kaynak bırakılır ve görevli arı kâşif arı olarak rastgele yeni bir kaynağa yönelir.

ABC algoritması nümerik problemler için geliştirilmiş olmakla beraber literatürde ayrık problem uygulamalarına da rastlanmaktadır. Ayrık optimizasyon, çözümlerin farklı ayrık değerlerden oluştuğu optimizasyon problemleridir [33]. Sürekli optimizasyon problemleri gibi her çözüm bir amaç değeri ile ifade edilmektedir. Protein yapı tahmini problemi de kafes modelleri kullanıldığında ayrık bir problem halini alır. Bu şekilde mümkün çözüm sayısı sonlu sahaya alınmış olur. Ancak, problemin anlaşılırlığı ve çözümlenebilirliği daha da karmaşık bir hal almış olabilir ve çakışma problemleri ile karşılaşma ihtimali artabilir.

ABC temelli yeni yaklaşımlarda ayrık optimizasyon problemler için uyarlamalar çok önemli bir yer tutmaktadırlar. Çözüm aranan bilindik ayrık optimizasyon problemlerinden İş Yeri Akış Çizelgeleme Problemi ve farklı tiplerdeki veri setlerinde dinamik kümeleme işlemleri için oluşturulan modeller önerilmiştir [32–37]. Ayrıca yayında ele alınan problem gibi yeni nesil problemler olarak nitelendirilebilecek Bulut Servis Kompozisyon Problemi ve DNA Çoklu Dizi Sıralama Problemi üzerine oluşturulan modeller de geliştirilmiştir [38, 39].

4.2. Önerilen Ayrık Yapay Arı Kolonisi Algoritması (Proposed Discrete Artificial Bee Colony Algorithm)

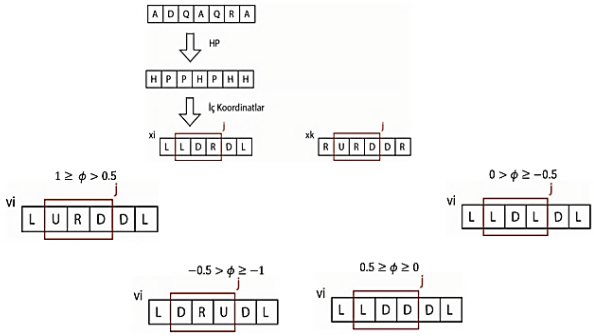
Bu çalışmada önerilen modelde ABC algoritmasının temel yaklaşımlarına tamamen sadık kalarak ayrık bir problem olan 3DHP-SC Modeliyle Protein Yapı Tahmini üzerinde çalışılmıştır [40]. Önerilen algoritmanın kaba kodu Şekil 4'te verilmiştir. Klasik HP modelinde her bir düğüm bir amino asidi gösterirken, 3DHP-SC modelinde her amino asit bir yan zincir ve bir omurga olmak üzere iki düğümle tanımlanmaktadır. Doğadaki moleküllerde hiçbir atom başka bir atomla aynı konumda bulunamaz. Bahsi geçen fiziksel kısıt nedeniyle elde edilen yapıda hiçbir yapı elemanının diğer elemanla çakışmaması gerekir. Ancak çakışmaları tamamen engellemek hem hesaplama maliyetini artırmakta hem de optimizasyon algoritmalarının kısa zamanda gelişimini durdurmasına neden olmaktadır. Bu nedenle verilen 3DHP-SC puan hesaplama sisteminde çakışmalar cezalandırılmaktadır.

Elde edilen bir yapının 3DHP-SC puanının hesaplanabilmesi için elemanların koordinatlarının, iç koordinat sisteminden Kartezyen koordinat sistemine dönüştürülmesi gerekmektedir. Dönüşüm işleminin yapılabilmesi için her zaman ilk omurga Kartezyen koordinat sisteminin orijininde, ilk yan zincir ise her zaman (0, -1, 0) noktasında alınır. Geri kalan dönüşümler ilk elemanların konumlarından hareketle yapılır [41]. Algoritmanın çalışması esnasında iç koordinatlar arasında önceki elemanla çakışmasına neden olacak koordinat engellenmektedir.

1- Başlangıç kaynaklarını belirle (3)
2- İşçi arı fazı: işçi arıları kaynaklara gönder ve komşu kaynakları keşfet (4,5)
3- Gözcü arı fazı: gözcü arıları kaynak miktarına göre belirlenen olasılıklara (7) bağlı olarak seçilen kaynaklara gönder ve komşu kaynakları keşfet (4, 5)
4- Kâşif arı fazı: terk edilen kaynaklar yerine yeni rastgele kaynak oluştur (3)
5- En iyi kaynağı tut
6- Durdurma kriteri sağlanmadıysa 2, sağlandıysa 7. adıma git
7- En iyi sonucu al ve bitir

Şekil 4. Önerilen ABC algoritma türevi kaba kodu (Psuedo code of proposed ABC algotihm)

Değerlerin ayrık oluşu ve aralarında herhangi bir sıralama ilişkisi bulunmaması sebebiyle komşu çözüm üretme işlemi için daha önce yapılan ayrık problem çözümleri göz önünde bulundurularak farklı bir strateji üretilmiştir. Bu stratejiye göre Yapay Arı Kolonisi algoritmasının kullandığı [-1,1] aralığındaki rastgele eleman, önerilen yöntemde seçilen aralıktaki elemanların yeni üretilen çözüme aktarılırken geçireceği işlemi belirlemektedir. Aynı aralıkta oluşturulan rastgele değer 0,5'ten büyükse alınan elemanlar çözümde aynı konuma işlem yapılmadan eklenmektedir. Rastgele değer -0,5'ten küçükse alınan değerler ters sıra ile ilgilenilen çözüme eklenerek yeni çözüm üretilmektedir. Üretilen rastgele değer (0, 0,5] aralığında ise, yeni çözüm üretmekte yararlanılan çözümden rastgele seçilen bir parametre ilgilenilen aralıktaki bir çözümle değiştirilir. Diğer durumlarda ise ilgilenilen aralıkta seçilen bir değer rastgele bir değerle değiştirilir. Böylece çözümlerin birbirine benzemesi belirli bir süre engellenir. Yapay Arı Kolonisi algoritmasının uygulanış örneği Şekil 5'te görülmektedir.



Şekil 5. Önerilen yeni çözüm üretme mekanizması (Proposed new candidiate generation mechanism)

5. DENEYSEL ÇALIŞMALAR (EXPERIMENTAL STUDIES)

5.1. Kullanılan Yöntemler (Used Techniques)

Protein tersiyer yapı çalışmalarında HP tipi etkileşim temel oluşturmaktadır. Proteinlerin doğasıyla ilgili çalışmalar ilerledikçe yeni düzenlemeler ile HP modelinin daha etkin türevleri geliştirilmektedir. Bahsi geçen hidrofobiklik özelliği amino asitlerin yan zincirine ait bir özelliktir. Şu halde daha etkili ve gerçekçi bir çözüm için amino asitleri modelde omurga ve yan zincir olarak iki boncukla göstermek daha gerçekçidir [41]. Yan zincir ait olduğu omurgaya birim uzaklıkta bulunmak zorundadır. Yan zincirler işleme katıldığı için başlangıç amino asidi için yapılan (0,0,0) noktasında bulunma kabulü omurga için korunarak ilk amino asidin yan zinciri her zaman (0,-1,0) noktasında olarak kabul edilir.

3DHP-SC modelinin enerji hesabında sadece H tipi amino asitlerin etkileşimlerinin değil diğer yapı elemanlarının etkileşimlerinin de ele alınması ile daha gerçekçi bir yapı elde edilmesi amaçlanmıştır. Gerçek proteinler H tipi amino asitleri merkeze alarak katlandıklarında bu

merkezin etrafını diğer yapı elemanları ile sarma eğilimi gösterirler. Böylece sıkı bir yapı oluşur ve görev yapabilir duruma geçerler. Elde edilen modelin daha sıkı olması için enerji fonksiyonumuza bu oluşumla görevli iki çarpan daha eklenmiştir [41].

$$score = [H - (NC \times CezaPuanı)] \times RadiusG_H \times RadiusG_P \quad (8)$$

$$H = \epsilon_{HH} \sum_{i=1,j>i}^{n} \delta_{r_{ij}^{HH}} + \epsilon_{BB} \sum_{i=1,j>i+1}^{n} \delta_{r_{ij}^{BB}} + \epsilon_{BH} \sum_{i=1,j\neq i}^{n} \delta_{r_{ij}^{BH}} + \epsilon_{BP} \sum_{i=1,j\neq i}^{n} \delta_{r_{ij}^{BP}} + \epsilon_{HP} \sum_{i=1,j>i}^{n} \delta_{r_{ij}^{HP}} + \epsilon_{PP} \sum_{i=1,j>i}^{n} \delta_{r_{ij}^{PP}} \quad (9)$$

$$RG_{aa} = \sqrt{\frac{\sum_{i=1}^{N_{aa}}[(x_i - \bar{X})^2 + (y_i - \bar{Y}) + (z_i - \bar{Z})]}{N_{aa}}} \quad (10)$$

$$RadiusG_H = maxRG_H - RG_H \quad (11)$$

$$RadiusG_P = \begin{cases} 1 & , \quad (RG_P - RG_H \geq 0) \\ \dfrac{1}{1 - (RG_P - RG_H)} & , \quad diğer\ durumlarda \end{cases} \quad (12)$$

Formüller incelendiğinde Eşitlik (10) için RG_{aa}, aa kısmına gelecek ilgilenilen amino asit için hesaplanan bir değerdir. N_{aa}, ele alınan amino asitle aynı sınıfta olan amino asitlerin sayısıdır. Eşitliklerdeki $RadiusG_P$, P tipi amino asitler için dönüş yarıçapı; $RadiusG_H$, H tipi amino asitler için dönüş yarıçapıdır. $maxRG_H$, zincir katlanmadan düz haldeyken RG_H'ın hesaplanmasıyla bulunan değerdir. Eşitlik (9) için ϵ_{XY}, XY tipi etkileşimin katsayısıdır. Her biri için bu değer önceden belirlenir. $\delta_{r_{ij}^{XY}}$, X tipinden bir yapı elemana birim uzaklıkta Y tipinden kaç eleman olduğunu göstermektedir. Eşitliklerde verilen NC, çakışma sayısıdır.

5.2. Kullanılan Parametreler (Used Parameters)

Çalışmada HP modelinde sıklıkla kullanılan, H ve P olmak üzere suyla etkileşim açısından iki elemanlı olarak oluşturulmuş referans diziler üzerinde çalışılmıştır. Bu referans diziler, çalışmaların çok farklı dizilimler üzerinde dizilimden bağımsız olarak elde edilecek başarıyı ölçecek şekilde HP modeli ile yapılan ilk çalışmalarda belirlenmiştir [2]. Kullanılan referans diziler Tablo 1'de verilmiştir.

Tablo 1. HP modeli için belirlenmiş Unger273d.* test dizileri [42] (Unger273d.* benchmark sequences for HP model)

İsim	Uzunluk	HP Zinciri
Unger273d.1	27	$(PH)^3H^2P^2(HP)^2P^{10}H^2P$
Unger273d.2	27	$PH^2P^{10}H^2P^2H^2P^2HP^2HPH$
Unger273d.3	27	$H^4P^5HP^5H^3P^8H$
Unger273d.4	27	$H^3P^2H^4P^3(HP)^2PH^2P^2HP^3H^2$
Unger273d.5	27	$H^4P^4HPH^2P^3H^2P^{10}$
Unger273d.6	27	$HP^6HPH^3P^2H^2P^3HP^4HPH$
Unger273d.7	27	$HP^2HPH^2P^3HP^5HPH^2(PH)^3H$
Unger273d.8	27	$HP^{11}(HP)^2P^7HPH^2$
Unger273d.9	27	$P^7H^3P^3HPH^2P^3HP^2HP^3$
Unger273d.10	27	$P^5H(HP)^5(PHH)^2PHP^3$

Yapay Arı Kolonisi algoritmasının program kodu MATLAB programlama ortamında oluşturulmuş ve i7 işlemcili 4GB RAM ve 64 bit işletim sistemine sahip masaüstü bilgisayarda koşulmuştur. Yapay Arı Kolonisi algoritması için kullanılan parametreler; limit değeri 200, koloni büyüklüğü (NP) değeri 50, jenerasyon sayısı 2000 şeklindedir. Kaynağın kalitesini belirleyen parametrelerin (D) sayısı, protein dizisinde bulunan amino asit sayısının iki katından iki eksik olarak belirlenmiştir. Bunun nedeni her amino asidin iki yapı elemanıyla (omurga ve yan zincir) belirtilmesi ve ilk iki yapı elemanının konumlarının sabit olarak kabul edilmesidir. Protein modeli olarak kullanılan 3DHP-SC modelinin parametreleri olarak ise H enerjisini hesaplamak için ϵ_{HH} 10; ϵ_{HP} ve ϵ_{BH} -3; $\epsilon_{PP}, \epsilon_{BP}$ ve ϵ_{BB} 1 olarak alınmıştır. Ayrıca ceza değeri 10 alınarak sonuçların çakışmasız olması sağlanmaya çalışılmıştır. Algoritma 10 kez koşulmuş ve çakışmasız en yüksek değerler elde edilmiştir. Alınan sonuçlar moleküllerin görselleştirilmesi için kullanılan en yaygın model olan top çubuk modeli ile verilmiştir.

5.3. Elde Edilen Sonuçlar (Obtained Results)

Protein Yapı Tahmin sürecinde birçok farklı model ve her bir model için farklı hesaplama yöntemleri kullanılmaktadır. Asıl amaç, yüksek skor elde etmek yerine proteinlerde katlanma sonucu artma eğiliminde olan etkileşimleri son yapıda en fazla sayıya ulaştırarak proteinin organizmadaki gerçek şekline en uygun yapıyı elde etmektir. Bu nedenle literatürde, elde edilen sonuçlar etkileşimler açısından değerlendirilmektedir. Moleküllerin doğal yapısından gözlemlenerek eklenen diğer etkileşimlerin sonuca direk etkisinin olmaması ve verilen katsayıların tartışmalı olması nedeniyle çözümün kalitesi sadece H tipi amino asitlerin etkileşim sayısı ile belirtilmektedir.

Yapay Arı Kolonisi algoritması ile 3DHP-SC modeli kullanılarak Protein Yapı Tahmin Problemi için üretilen çözümler daha önce yapılan Paralel Genetik Algoritma, Hiyerarşik Paralel Genetik Algoritma ve Yapay Arı Kolonisinin farklı bir gerçekleştiriminden elde edilen sonuçlar ile karşılaştırılmıştır. Sadece bu üç çalışmanın

seçilme nedeni, ilgili çalışmalarda 3DHP-SC modelinin kullanılmış olmasıdır [29, 41, 43].Karşılaştırma yapılan Paralel Genetik Algoritma (PGA), Yapay Arı Kolonisi, Hiyerarşik Paralel Genetik Algoritma (HPGA), Yapay Arı Kolonisi (Sequential ABC - sABC) ve paralelleştirilmiş şekilleri (MS-ABC, HH-ABC_1, HH-ABC_2) çalışmalarının daha sağlıklı değerlendirilebilmesi için uygulamada kullanılan parametreler ön bilgi olarak verilmektedir. HPGA ve PGA kullanılan çalışmalarda başlangıç popülasyonunun yarısı çakışmasız çözümlerden oluşmakta diğer yarısı ise rastgele üretilmektedir. PGA'da uygunluk hesaplama işleminde Eşitlik (13)'de görüldüğü gibi c_{mult} ile belirtilen kullanıcı tanımlı bir ölçekleme değeri alınmıştır. Parametre $1{,}2 \leq c \leq 2{,}0$ olacak şekilde belirlenmektedir. Ayrıca seçilim işleminde k-turnuva yöntemi kullanılmıştır. Tablo 2'de verilen parametrelerle 20 farklı koşma gerçekleştirilmiştir [29]. HPGA ile yapılan çalışmada, İkinci parametre göç oranı olarak adlandırılmaktadır. Bu parametre, her göç olayında diğer ana bilgisayara geçecek en iyi çözümlerin sayısını belirlemektedir. Tablo 3'te verilen parametreler kullanılarak 100 farklı koşma gerçekleştirilmiştir [41].

$$f' = c_{mult} \times f \qquad (13)$$

[30]'da ABC algoritmasının serial (sABC) ve 3 farklı paralel yöntemleri (MS-ABC, HH-ABC1, HH-ABC2) 3DHP-SC modeli ile benzer şekilde kullanılmıştır. Ancak ilgili çalışmada bu yöntemler diğer çalışmalarla karşılaştırılmamıştır. Gerçek değerlendirme kriteri yazarların önceki yayınlarında ve referans yayınlarda HH etkileşim sayısı olarak kabul görse de [30]'da uygunluk değerleri verildiği için aynı şartlarda uygunluk değerleri ile karşılaştırma yapılmıştır. Çalışma [43]'da koloni büyüklüğü sABC ve MS-ABC'de 250, HH-ABC_1'de 252, HH-ABC_2'de 1000 alınmış ve popülasyon 6000 olarak alınmış 100 defa koşularak sonuçlar elde edilmiştir, ilgili karşılaştırmalar skorların ortalama değerleri hesaplanarak yapılmıştır.

Tablo 2. PGA parametreleri
(PGA parameters)

Parametre İsmi	Değer
Jenerasyon sayısı	3000
Popülasyon büyüklüğü	472
Çakışmasız birey oranı	%3
Çaprazlama oranı	%80
Mutasyon oranı	%5
Turnuva boyutu	%3
c_{mult}	1,4

Tablo 3. HPGA parametreleri
(HPGA parameters)

Parametre İsmi	Değer
Jenerasyon sayısı	3000
Grup sayısı	4
Popülasyon büyüklüğü	500
Çaprazlama oranı	%80
Mutasyon oranı	%8
Turnuva boyutu	%3
Örnek seyreltme jenerasyonu	300
Turnuva için zayıf olunan jenerasyon sayısı	300
Göç boşluğu	120
Göç oranı	5

Koşmalar sonucu elde edilen en iyi HH etkileşim sayılarından oluşan sonuçlar HPGA ve PGA ile karşılaştırmalı olarak Tablo 4'te verilmiştir. Bulunan sonuçlara göre oluşan protein yapıları top çubuk modeliyle görselleştirilmiş ve Şekil 6'da verilmiştir. Elde edilen sonuçların süreç içerisindeki gelişimi Şekil 7'de verilen yakınsama grafiklerinden görülmektedir. Ayrıca ABC türevlerinin bulunduğu diğer çalışmada ortalamalar ele alındığı için koşmaların ortalamaları açısından karşılaştırma Tablo 5'te yapılmıştır. Önerilen ABC algoritması için seçilen parametreler incelendiğinde yaklaşık 1,000,000 değerlendirme yaptığı ve 28,320,000 değerlendirme yapan PGA'yı test dizilerinin yedisinde geçtiği ve üçünde aynı sonuçları verdiği görülmektedir. PGA üzerine yapılan geliştirmelerle değerlendirme sayısının arttığı ve çeşitli özelleştirmelerin yapıldığı HPGA ile hemen hemen aynı performansı gösterdiği görülmektedir. HPGA'nın parametreleri incelendiğinde 150,000,000 değerlendirme yaptığı görülmektedir ve bu yaklaşık 150 kat fazla maliyetli bir işleme neden olmaktadır. ABC'nin önceki gerçekleştirimi ilgili yayında diğer çalışmalarla karşılaştırılmamış ve sadece kararlılık açısından değerlendirilmiştir. Önceki ABC gerçekleştiriminin de 150,000,000 ve 600,000,000 değerlendirme yaptığı parametrelerden görülmektedir.

Tablo 4. Önerilen model ile PGA ve HPGA yöntemlerinin HH etkileşim sonuçları (HH interaction results of proposed ABC, PGA and HPGA)

Protein Modeli	Sonuçlar		
	PGA	HPGA	ABC
Unger273d.1	10	12	13
Unger273d.2	12	13	13
Unger273d.3	11	13	13
Unger273d.4	18	22	21
Unger273d.5	11	13	13
Unger273d.6	13	14	14
Unger273d.7	16	16	16
Unger273d.8	6	6	6
Unger273d.9	9	10	10
Unger273d.10	14	14	14

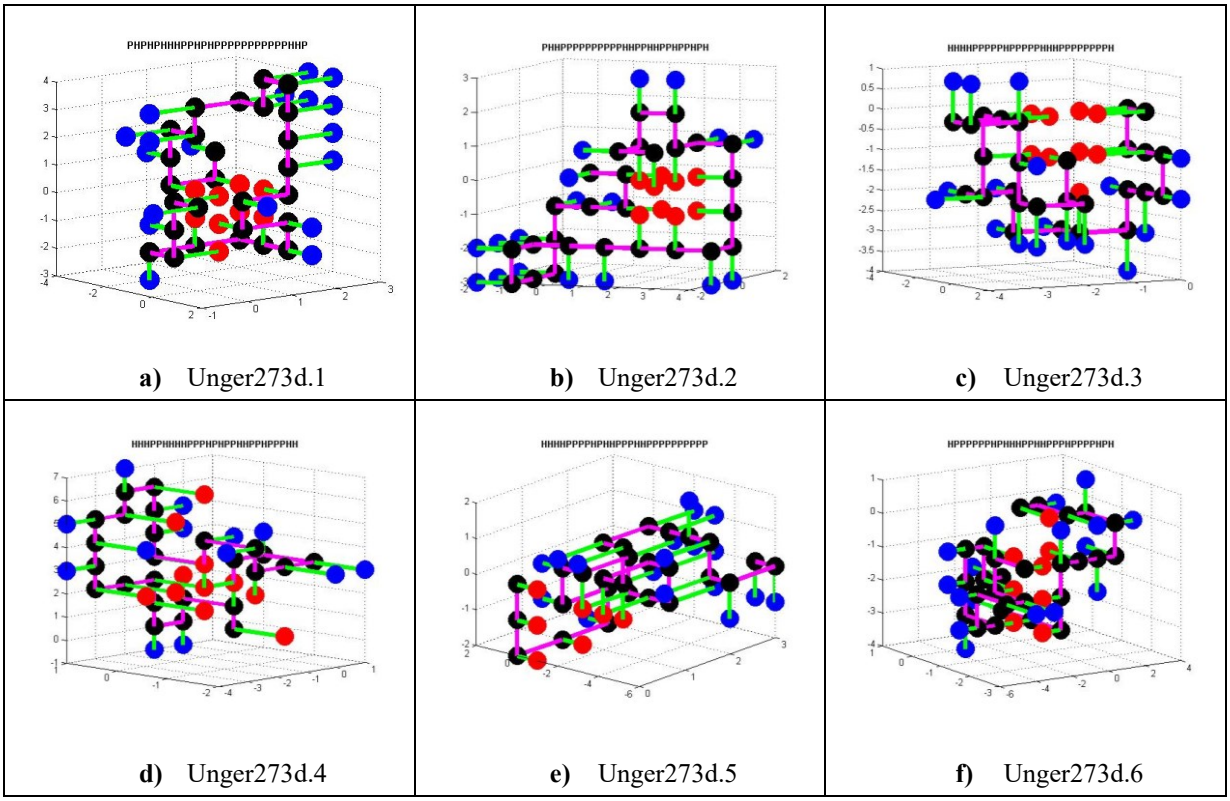
Tablo 5. Önerilen ABC modelinin ve çalışma [30]'dan alınan ortalama skor değerleri (The mean scores of proposed ABC and of the algorithms from [30])

Protein Modeli	sABC	MS-ABC	HH-ABC1	HH-ABC2	ABC
Unger 273d.1	523,27	524,15	581,83	517,21	954,98
Unger 273d.2	529,48	530,51	673,95	588,74	1114,25
Unger 273d.3	621,11	620,05	772,36	711,37	1108,19
Unger 273d.4	938,36	940,91	1002,59	989,55	1108,95

Şekil 6'da verilen top çubuk modelinde kırmızı ile H tipi yan zincirler, mavi ile P tipi yan zincirler ve kullandığımız model omurga ve yan zincirleri ayrı ele aldığı için siyahla amino asit omurgaları gösterilmiştir. Proteinlerde daha önce değinildiği gibi H tipi amino asit yan zincirleri çekirdek ve P tipi yan zincirler kabuk olacak şekilde bir yapı oluşturma eğilimi bulunmaktadır. Bu eğilim benzetimimizin çıkış noktalarından olduğu için elde edilen şekillerin beklentileri karşıladığı açık bir şekilde görülmektedir. Ayrıca Şekil 7'de verilen yakınsama grafikleri performans açısından gelişimi net olarak ortaya koymaktadır. Çözümün bulunma sürecinde 3DHP-SC skoru kullanıldığı için yakınsama grafiklerinde de bu değerin kullanılması daha uygundur. Yakınsama grafikleri 3DHP-SC skorları üzerinden verilmiştir.

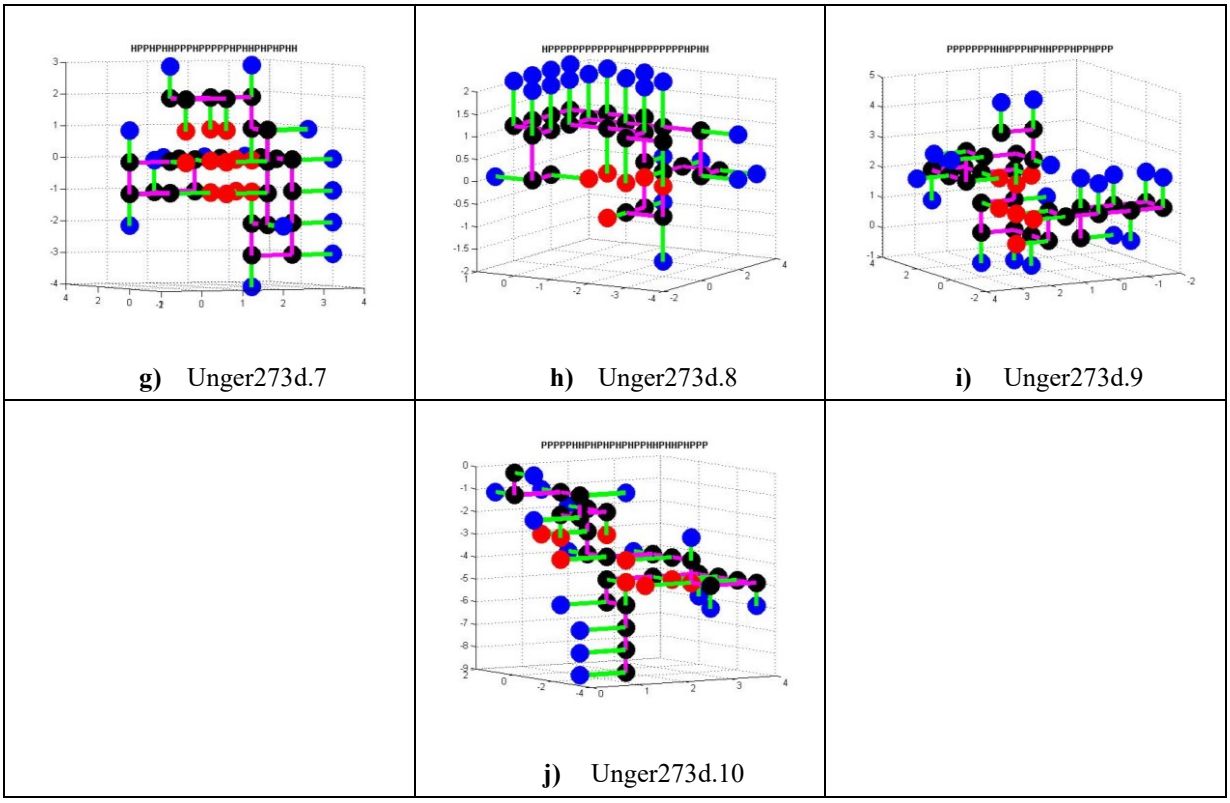
6. SONUÇ (CONCLUSION)

HP model üzerine geliştirilmiş olan Üç Boyutlu Hidrofobik Polar Yan Zincir modeli kullanılarak protein yapı tahmini için Yapay Arı Kolonisi Algoritması temelli oluşturulan ayrık bir model önerilmiştir. Önerilen ABC ile üretilen çözümler literatürde yer alan Paralel Genetik Algoritma, Hiyerarşik Paralel Genetik Algoritma ve Yapay Arı Kolonisi Algoritmasının farklı bir gerçekleştirimi kullanılarak elde edilen çözümler ile karşılaştırılmıştır. Yapılan karşılaştırmada performans olarak çok daha güçlü bir donanım üzerinde gerçekleştirilmiş olmasına rağmen çalışmamızın algoritmanın felsefesine sadık kalarak az maliyetle çok daha güçlü bir sistem ortaya koyduğu açıkça görülmektedir. Elde edilen sonuçlar çalışmanın amaçlarından olan deneysel çalışmaya ön bilgi oluşturmak, protein şekil bozukluklarını tahmin etmek, tıp ve biyoloji alanında yeni bilgiler edinmek ve biyolojik yapıları taklit etmek konularında araştırmacılara yol gösterecek seviyededir.



a) Unger273d.1 **b)** Unger273d.2 **c)** Unger273d.3

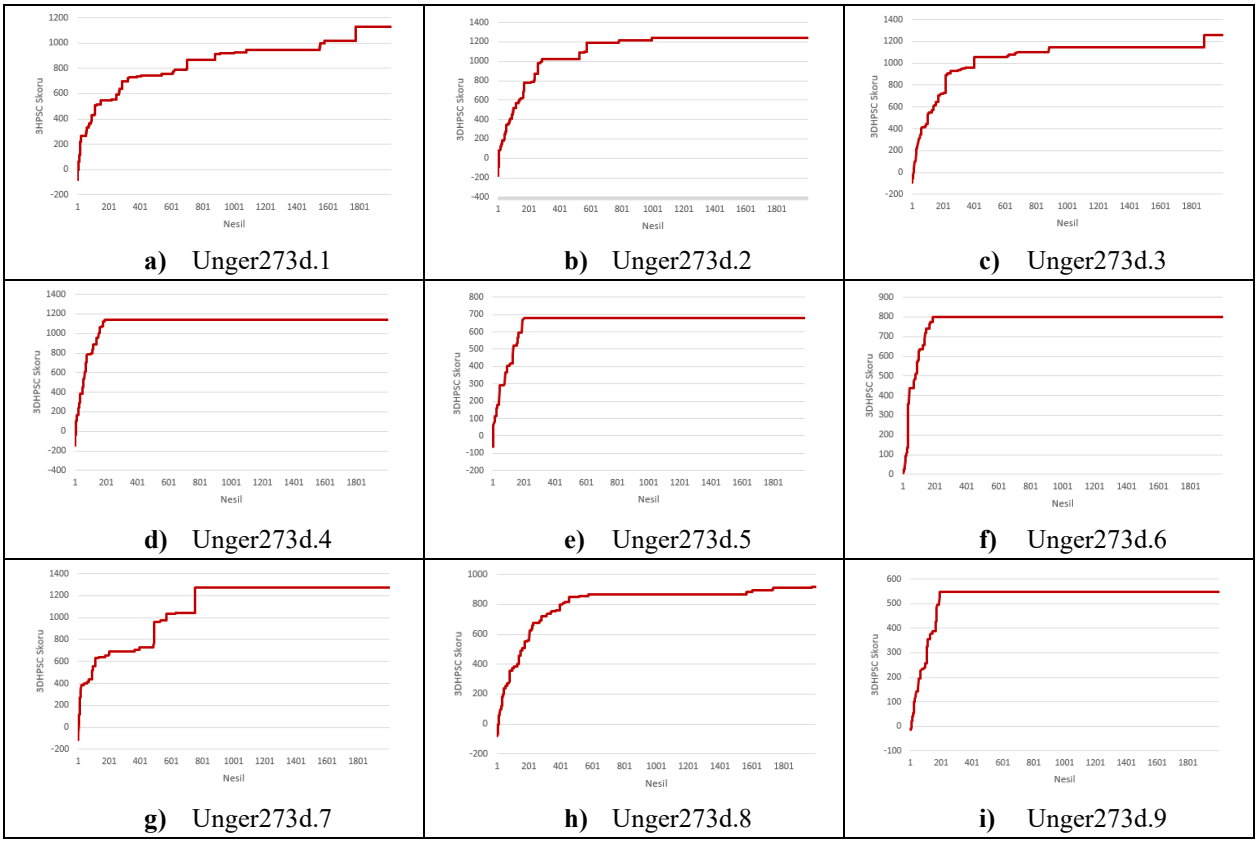
d) Unger273d.4 **e)** Unger273d.5 **f)** Unger273d.6



g) Unger273d.7 **h)** Unger273d.8 **i)** Unger273d.9

j) Unger273d.10

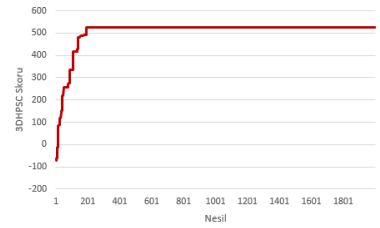
Şekil 6. Unger273d.* dizileri için ABC algoritması ile bulunan en yüksek skorlu yapıların top çubuk modelleri (Best ABC results as ball stick models for Unger273d.* sequence)



a) Unger273d.1 **b)** Unger273d.2 **c)** Unger273d.3

d) Unger273d.4 **e)** Unger273d.5 **f)** Unger273d.6

g) Unger273d.7 **h)** Unger273d.8 **i)** Unger273d.9



j) Unger273d.10

Şekil 7. Unger273d.* dizileri için ABC algoritmasının yakınsama grafikleri (ABC algorith convergency graphics for Unger273d.* sequences)

KAYNAKLAR (REFERENCES)

[1] M. Zvelebil and J. Baum, *Understanding bioinformatics*. New York: Garland Science, 2008.

[2] R. Unger and J. Moult, "Genetic algorithms for protein folding simulations.," *J. Mol. Biol.*, vol. 231, pp. 75–81, 1993.

[3] K. A. Dill, "Theory for the folding and stability of globular proteins.," *Biochemistry*, vol. 24, no. 6, pp. 1501–1509, 1985.

[4] Y. Karami, M. Fathy, H. Khakzad, H. Shirazi, and S. Arab, "Protein structure prediction using bio-inspired algorithm: A review," *AISP 2012 - 16th CSI Int. Symp. Artif. Intell. Signal Process.*, no. 978, pp. 201–206, 2012.

[5] J. Ye, "A review of artificial intelligence techniques applied to protein structure prediction," Simon Fraser University, 2007.

[6] M. Paluszewski, "Algorithms for Protein Structure Prediction," Københavns Universitet, Det Natur-og Biovidenskabelige Fakultet, Biologisk Institut, 2008.

[7] X. Zhang, T. Wang, H. Luo, J. Y. Yang, Y. Deng, J. Tang, and M. Q. Yang, "3D protein structure prediction with genetic tabu search algorithm," *BMC Syst. Biol.*, vol. 4 Suppl 1, no. Suppl 1, p. S6, 2010.

[8] T. Jiang, Q. Cui, G. Shi, and S. Ma, "Protein folding simulations of the hydrophobic–hydrophilic model by combining tabu search with genetic algorithms," *J. Chem. Phys.*, vol. 119, no. 8, p. 4592, 2003.

[9] X. Chen, M. Lv, L. Zhao, and X. Zhang, "An Improved Particle Swarm Optimization for Protein Folding Prediction," *Int. J. Inf. Eng. Electron. Bus.*, vol. 3, no. 1, pp. 1–8, Feb. 2011.

[10] M. Garza-Fabre, E. Rodriguez-Tello, and G. Toscano-Pulido, "Comparative Analysis of Different Evaluation Functions for Protein Structure Prediction Under the HP Model," *J. Comput. Sci. Technol.*, vol. 28, no. 5, pp. 868–889, 2013.

[11] C.-J. Lin and S.-C. Su, "Protein 3 D HP model folding simulation using a hybrid of genetic algorithm and particle swarm optimization," *Int. J. Fuzzy Syst.*, vol. 13, no. 2, pp. 140–147, 2011.

[12] C. Chira, "A Hybrid Evolutionary Approach to Protein Structure Prediction with Lattice Models," *Proteins*, pp. 2300–2306, 2011.

[13] C. M. Johnson and A. Katikireddy, "A Genetic Algorithm with Backtracking for Protein structure Prediction," *GECCO'06,July*, vol. 8, p. -12, 2006.

[14] R. O. R. O. Day, J. B. Zydallis, G. B. G. B. Lamont, and R. Pachter, "Solving the protein structure prediction problem through a multiobjective genetic algorithm," *Nanotechnology*, vol. 2, pp. 32–35, 2002.

[15] H. A. A. H. A. A. Bahamish, R. Abdullah, and R. A. A. Salam, "Protein Tertiary Structure Prediction Using Artificial Bee Colony Algorithm," *2009 Third Asia Int. Conf. Model. Simul.*, 2009.

[16] C.-J. C. Lin and S. S.-C. Su, "Using an efficient artificial bee colony algorithm for protein structure prediction on lattice models," *Int. J. Innov. Comput. Inf. Control*, vol. 8, no. 3, pp. 2049–2064, 2012.

[17] A. G. Citrolo and G. Mauri, "A Hybrid Monte Carlo Ant Colony Optimization Approach for Protein Structure Prediction in the HP Model," *Electron. Proc. Theor. Comput. Sci.*, vol. 130, pp. 61–69, 2013.

[18] B. Maher, A. Albrecht, M. Loomes, X.-S. Yang, and K. Steinhöfel, "A Firefly-Inspired Method for Protein Structure Prediction in Lattice Models," *Biomolecules*, vol. 4, no. 1, pp. 56–75, 2014.

[19] F. L. Custódio, H. J. C. Barbosa, and L. E. Dardenne, "A multiple minima genetic algorithm for protein structure prediction," *Appl. Soft Comput. J.*, vol. 15, pp. 88–99, 2014.

[20] L. F. Nunes, L. C. Galvão, H. S. Lopes, P. Moscato, and R. Berretta, "An integer programming model for protein structure prediction using the 3D-HP side chain model," *Discret. Appl. Math.*, vol. 198, pp. 206–214, 2016.

[21] M. Garza-Fabre, E. Rodriguez-Tello, and G. Toscano-Pulido, "Constraint-handling through multi-objective optimization: The hydrophobic-polar model for protein structure prediction," *Comput. Oper. Res.*, vol. 53, pp. 128–153, 2015.

[22] J. G. P. Jin and H. Xu, "3D protein structure prediction of influenza A virus based on optimization genetic algorithm," vol. 27, no. 3, pp. 735–739, 2014.

[23] A. Perdomo-Ortiz, N. Dickson, M. Drew-Brook, G. Rose, and A. Aspuru-Guzik, "Finding low-energy conformations of lattice protein models by quantum annealing," *Sci. Rep.*, vol. 2, pp. 1–7, 2012.

[24] A. Lesk, *Introduction to Bioinformatics*. Oxford University Press, 2014.

[25] C. B. Anfinsen, "Principles that govern the folding of protein chains," *Science*, vol. 181, no. 96, pp. 223–230, 1973.

[26] K. A. Dill, "Dominant forces in protein folding.," *Biochemistry*, vol. 29, no. 31, pp. 7133–7155, 1990.

[27] B. Honig, "Protein folding: from the levinthal paradox to structure prediction.," *J. Mol. Biol.*, vol. 293, no. 2, pp. 283–293, 1999.

[28] R. König and T. Dandekar, "Refined Genetic Algorithm Simulations to Model Proteins," *J. Mol. Model.*, vol. 5, no. 12, pp. 317–324, 1999.

[29] C. M. V. Benitez and H. S. Lopes, "A parallel genetic algorithm for protein folding prediction using the 3D-HP Side Chain model," in *2009 IEEE Congress on Evolutionary Computation*, 2009, pp. 1297–1304.

[30] D. Karaboga, "An idea based on Honey Bee Swarm for Numerical Optimization," *Tech. Rep. TR06, Erciyes Univ.*, no. TR06, p. 10, 2005.

[31] C. Ozturk, E. Hancer, and D. Karaboga, "Improved clustering criterion for image clustering with artificial bee colony algorithm," *Pattern Anal. Appl.*, vol. 18, no. 3, pp. 587–599, 2015.

[32] C. Ozturk, E. Hancer, and D. Karaboga, "Dynamic clustering with improved binary artificial bee colony algorithm," *Appl. Soft Comput.*, vol. 28, pp. 69–80, 2015.

[33] C. Ozturk, E. Hancer, and D. Karaboga, "A novel binary artificial bee colony algorithm based on genetic operators," *Inf. Sci. (Ny).*, vol. 297, pp. 154–170, 2015.

[34] D. Ye and Z. Chen, "A new approach to minimum attribute reduction based on discrete artificial bee colony," *Soft Comput.*, pp. 1893–1903, 2014.

[35] I. Ribas, R. Companys, and X. Tort-Martorell, "An efficient Discrete Artificial Bee Colony algorithm for the blocking flow shop problem with total flowtime minimization," *Expert Syst. Appl.*, vol. 42, no. 15–16, pp. 37–41, 2015.

[36] Y.-Y. Han, D. Gong, and X. Sun, "A discrete artificial bee colony algorithm incorporating differential evolution for the flow-shop scheduling problem with blocking," *Eng. Optim.*, vol. 47, no. 7, pp. 927–946, 2015.

[37] Z. Cui and X. Gu, "An improved discrete artificial bee colony algorithm to minimize the makespan on hybrid flow shop problems," *Neurocomputing*, vol. 148, pp. 248–259, 2015.

[38] D. Karaboga and S. Aslan, "A discrete artificial bee colony algorithm for detecting transcription factor binding sites in DNA sequences.," *Genet. Mol. Res.*, vol. 15, no. 2, pp. 1–11, 2016.

[39] Y. Huo, Y. Zhuang, J. Gu, S. Ni, and Y. Xue, "Discrete gbest-guided artificial bee colony algorithm for cloud service composition," *Appl. Intell.*, pp. 661–678, 2014.

[40] T. Batbat, "Protein Yapısının Yapay Arı Kolonisi Algoritması Ile Tahmini," Erciyes Üniversitesi, 2014.

[41] C. M. V. Benitez and H. S. Lopes, "Hierarchical Parallel Genetic Algorithm applied to the three-dimensional HP Side-chain Protein Folding Problem," in *2010 IEEE International Conference on Systems, Man and Cybernetics*, 2010, pp. 2669–2676.

[42] R. Unger and J. Moult, "Genetic Algorithm for 3D Protein Folding Simulations," in *Proceedings of the 5th International Conference on Genetic Algorithms*, 1993, pp. 581–588.

[43] C. Vargas Benítez, H. S. Lopes, C. M. V. Benítez, H. S. Lopes, C. Vargas Benítez, and H. S. Lopes, "Parallel Artificial Bee Colony Algorithm Approaches for Protein Structure Prediction Using the 3DHP-SC Model," in *Intelligent Distributed Computing IV SE - 27*, vol. 315, Springer, 2010, pp. 255–264.