

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

FİDAN GELİŞİMİ ALGORİTMASI İLE PROTEİN YAPI TAHMİNİ

SEDA ARSLAN TUNCER

Tez Yöneticisi:
Yrd.Doç.Dr.Ali KARCI

YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

ELAZIĞ, 2008

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

FİDAN GELİŞİM ALGORİTMASI İLE PROTEİN İKİNCİL YAPI TAHMİNİ

Seda ARSLAN TUNCER

YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

Bu tez, tarihinde aşağıda belirtilen jüri tarafından oybirliği /oyçokluğu ile başarılı / başarısız olarak değerlendirilmiştir.

Danışman: Yrd. Doç. Dr. Ali KARCI
Üye: Doç. Dr. İbrahim TÜRKOĞLU
Üye: Yrd. Doç. Dr. Mehmet KAYA

Bu tezin kabulü, Fen Bilimleri Enstitüsü Yönetim Kurulu'nun/...../..... tarih ve sayılı kararıyla onaylanmıştır.

TEŞEKKÜR

Tez çalışmalarım esnasında, ilgi ve yardımlarını esirgemeyen danışman hocam Sayın **Yrd.Doç.Dr.Ali KARCI** ya desteğini hissettiğim eşime teşekkürlerimi sunarım.

İÇİNDEKİLER

TEŞEKKÜR	
İÇİNDEKİLER.....	I
ŞEKİLLER LİSTESİ.....	II
TABLolar LİSTESİ.....	IV
ÖZET	V
ABSTRACT.....	VI
1. GİRİŞ	1
2. BİYOİNFORMATİK	3
2.1. Biyoinformatik Veri Tabanları.....	3
2.2. Biyolojik Bilgi Analizi.....	6
2.3. Biyoinformatik Çalışma Alanları.....	7
3. BİYOKİMYA ve PROTEİNLER	10
3.1. Proteinler.....	11
3.1.1. Proteinin Tanımı.....	12
3.1.2. Proteinlerin Yapılarındaki Bağlar.....	14
3.1.3. Protein Moleküllerinin Yapısı ve Konformasyonu.....	14
3.2. Aminoasitler	20
3.3 Proteinlerin Özellikleri.....	21
3.4. Proteinlerin Yapılarına Göre Sınıflandırılması.....	23
3.5 Proteinlerin Biyolojik Rollerine Göre Veya Fonksiyonel Olarak Sınıflandırılması.....	26
4. FİDAN GELİŞİM ALGORİTMASI.	28
4.1. Fidanların Ekimi.....	29
4.2. Fidanların Büyümesi	32
5. PROTEİN YAPI TAHMİNİ İÇİN UYGULANAN DİĞER YÖNTEMLER.....	35
6. GELİŞTİRİLEN YÖNTEM VE SONUÇLAR	37
7. SONUÇLAR	49

TABLÖLER LİSTESİ

Tablo 6.1. Amino asitlerin nümerik değerlere dönüşüm tablosu	38
Tablo 6.2. Açık sınıfları.....	38
Tablo 6.3. 1a9a proteini için başlangıç popülasyonu açık değerleri.....	39
Tablo 6.4. 1a89 proteini için başlangıç popülasyonu açık değerleri.....	39
Tablo 6.5. 1bbe proteini için başlangıç popülasyonu açık değerleri	39
Tablo 6.6. 1a9a proteininin eşleştirilmesi sonucunda oluşan yeni fidanlar	39
Tablo 6.7. 1a89 proteininin eşleştirilmesi sonucunda oluşan yeni fidanlar.....	40
Tablo 6.8. 1bbe proteininin eşleştirilmesi sonucunda oluşan yeni fidanlar.....	40
Tablo 6.9. Versiyon1, 1a9a proteini için test açık sınıfı ve en iyi fidan için açık sınıfı değerleri..	42
Tablo 6.10. Versiyon1, 1a9a proteini için test açık değerleri ve en iyi fidana ait açık değerleri...	42
Tablo 6.11. Versiyon1, 1a89 proteini için test açık sınıfı ve en iyi fidan için açık sınıfı değerleri..	43
Tablo 6.12. Versiyon1, 1a89 proteini için test açık değerleri ve en iyi fidana ait açık değerleri.	44
Tablo 6.13. Versiyon1, 1bbe proteini için test açık sınıfı ve en iyi fidana ait açık değerleri.....	44
Tablo 6.14. Versiyon1, 1bbe proteini için test açık değerleri ve en iyi fidana ait açık değerleri....	45
Tablo 6.15. Versiyon2, 1a9a proteini için test açık sınıfı ve en iyi fidana ait açık sınıfı değerleri.	45
Tablo 6.16. Versiyon2, 1a9a proteini için test açık değerleri ve en iyi fidana ait açık değerleri ...	46
Tablo 6.17. Versiyon2, 1a89 proteini için test açık sınıfı ve en iyi fidana ait açık sınıfı değerleri..	46
Tablo 6.18. Versiyon2, 1a89 proteini için test açık değerleri ve en iyi fidana ait açık değerleri	46
Tablo 6.19. Versiyon2, 1bbe proteini için test açık sınıfı ve en iyi fidana ait açık sınıfı değerleri..	47
Tablo 6.20. Versiyon2, 1bbe proteini için test açık değerleri ve en iyi fidana ait açık değerleri	47
Tablo 6.21. Versiyon1 ve Versiyon2 için performans değerleri	48
Tablo 7.1. Versiyon1 ve Versiyon2 için performans değerleri.....	50

ÖZET

Yüksek Lisans Tezi

FİDAN GELİŞİMİ ALGORİTMASI İLE PROTEİNLERİN İKİNCİL YAPI TAHMİNİ

Seda ARSLAN TUNCER

Fırat Üniversitesi

Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

2008, sayfa:52

NP türü problemleri çözmek amacıyla, özellikle doğadaki sistemlerden esinlenme konusunda yapılan çalışmalara son yıllarda ilgi gittikçe artmıştır. Genetik algoritmalar, karınca koloni algoritmaları, yapay sinir ağları, parçacık sürü optimizasyonu ve yapay bağışıklık sistemleri gibi teknikler yoğun olarak optimizasyon amaçlı kullanılmaktadır.

Bu tez çalışmasında, fidanların gelişiminden esinlenerek geliştirilen hesapsal bir yöntem olan Fidan Gelişimi Algoritması kullanılmıştır. Fidan Gelişimi Algoritması hem yerellik ve hem de küresellik özelliğine sahip yeni bir hesapsal yöntemdir.

Aminoasitler arasındaki etkileşimler analitik olarak bilinmediğinden, protein yapı tahmini yapmak için yaklaşımsal yöntemler kullanmak uygundur.

Bu tezde, proteinlerin peptit bağları arasındaki karbon atomları kullanılarak proteinlerin ikincil yapısının tahmini yapılmaya çalışılmıştır. Veriler Fidan Gelişimi Algoritması kullanılarak eğitime tabi tutulmuştur. Algoritmanın test işlemleri 1a9a, 1a89, 1b8e proteinleri için yapılmıştır ve proteinlerin açılımları tespit edilerek proteinlerin yapıları tahmin edilmiştir.

Anahtar kelimeler: Proteinler, Protein Yapı Tahmini, Fidan Gelişimi Algoritması, Karıştır-Geliştir.

ABSTRACT

Master Thesis

PROTEIN SECONDARY STRUCTURE ESTIMATION WITH TREE GROWING UP ALGORITHM

Seda ARSLAN TUNCER

Firat University

Natural and Applied Science Institute

Department of Computer Engineering

2008, page:52

There is interest in the methods were inspired by the natural processing systems at aim of solving NP type problems, recently. Genetic algorithms, ant colony algorithms, artificial neural Networks, particle swarm optimization, and artificial immune systems are used for optimization.

In this thesis, tree growing up algorithm which was developed with respect to inspiring by growing up of trees, was used. Tree growing up algorithm consists of local and global search together, and it is a new method.

Approximation method was used in this study, since the interactions amongs the amino acids are not known, analytically.

In this study, we used amino acid primary structure chains. The angles of centre carbons in amino acids sequence in two dimensions were used by grouping centre carbons three by three. In order to accomplish this study, we used the actual angle values obtained from data bank for training of tree growing up algorithm. The proteins 1a9a, 1a89, 1bbe were used for testing process and angles' classes were determined for estimation of secondary structure of proteins.

Keywords: Proteins, Protein Structure Estimation, Tree Growing up Algorithm, Mix-Improve.

1.GİRİŞ

İnsanoğlu tarihi boyunca sürekli gelişme, geliştirme çabası içinde olmuştur. Bu çabaların asıl hedefi insanın refah ve mutluluğunu artırmaktır. Dolayısıyla gelişme süreci insandan kaynaklanan ve sonunda yine insana ulaşan bir döngüdür. Bu süreç insan zekası, bilim ve teknoloji öğeleri arasındaki sebep-sonuç ilişkileri ile olur ve gelişir.

Geçtiğimiz yüzyıl içerisinde bilimin çok hızlı ilerlemesi insanların üzerine büyük bir bilgi yükü getirmiştir. Zaman içerisinde artan bilginin bir şekilde depolanması gerekmektedir. Ancak bilgilerin depolanmasının yanında bilgiye rahat ve hızlı erişim de önemlidir. Çünkü bilginin en iyi şekilde kullanılabilmesi etkin bir erişimi gerektirmektedir..

Biyoinformatik bilimi, biyolojik bilgilerin yaratılması ve saklanması için veritabanlarının oluşturulmasıdır. Bu konudaki çalışmaların çoğu biyolojik verilerin analizi ile ilgilidir. Artan sayıdaki projelerde biyolojik bilgilerin organizasyonu gerekmektedir. Bu alanda oluşturulan veritabanlarının büyük bir kısmını nükleik asitler oluşturmaktadır. Bu nedenle biyoloji ve biyokimya bilimi ile bilgisayar biliminin tekniklerini birleştirip, bu alanlardaki problemlere çözüm arayan bilim dalı, günümüzde çok popüler bir bilim dalı haline gelmiştir[5].

Proteinler, yapı ve fonksiyon maddeleri olarak, biyokimyasal açıdan çok önemli bir yere sahiptirler. Canlı organizmanın doku hücrelerinde proteinlerin yer almadığı biyokimyasal bir olaya rastlamak hemen hemen imkansızdır. Virüslerden başlayarak, en küçük canlıdan insana kadar her canlı için yapısal ve yaşamsal önem taşır. Yaşamla ilgili her metabolik tepkimede doğrudan veya dolaylı şekilde proteinlerin rolü vardır. Bu nedenle son yıllarda gelişen bilgisayar teknikleriyle beraber bu konuda çalışmalarda artmıştır. Birçok bilim adamı bu proteinler üzerinde çalışmaktadırlar. Bu çalışmalara rağmen hala proteinlerin keşfedilemeyen yönleri araştırmacıların ilgisini çekmektedir ve bu alanda çalışmalar her geçen gün artarak devam etmektedir[1].

NP türü problemler genellikle optimizasyon problemleridir. Optimizasyon problemlerini çözmek için problem uzayında en iyi çözüm aranır. En iyi çözüm için tüm algoritmalarda bir amaç fonksiyonu vardır. Tüm mühendislik dallarında optimizasyon problemleri bir amaç fonksiyonuna sahip doğadaki algoritmalar ile çözülür. Bu yaklaşım algoritmalarının mühendislik problemlerine uygulanmasının nedeni, problemin matematiksel modelinin kurulmasının zor veya çözüm zamanı maliyetinin çok yüksek olmasıdır. Doğadan esinlenerek yapılan çalışmalar son yıllarda çokça ilgi uyandırmıştır. Bunlardan Genetik algoritma, Karınca Koloni algoritması, Yapay Sinir Ağları, Yapay bağışıklık sistemi, Fidan gelişim algoritması örnek algoritmalarıdır.

Fidan gelişimi algoritması Günümüzde kullanılan yumuşak hesaplama tekniklerine alternatif bir yöntem olarak geliştirilmiştir. Fidanlar beraber aynı ortamda büyümektedirler. Bir ormanda bir fidandan sadece bir tane vardır. Birden fazla olamaz. Bu algoritmada fidanların büyüyeceği alana düzgün dikilmiş olduğu düşünülmüştür. Fidanlar büyürken dallanmaların birbirine çok yakın olmaması hesapsal yöntemde bir operatör olarak ele alınır. Fidanlar aşılama vermeye başladıktan sonra çift çenekli ağaçlarda çiftleşme ile yeni fidanların üremesi sağlanır. yöntemiyle başka bir fidana dönüşür. Bu mantık ile aşılama operatörü tanımlanmıştır. Fidanlar çiçek Bu işlemden çiftleşme operatörü olarak tanımlanır. Bütün bu tanımlamalar sonucunda genel amaçlı bir hesapsal yöntem ortaya konulmuştur.

Bu açıklamalar etrafında fidan gelişim algoritması bir çok probleme uygulanabilmesine olanak sağlar.

Örneğin; İnsan vücudu bir çok hücrelerden oluşmaktadır. Hücreler ise vücudun hangi parçasını oluşturuyorsa o bölgede ihtiyaç duyulan boyut ve şekle sahip olurlar. Bu yapı içerisinde proteinler önemli bir yere sahiptir. Birçok bilim adamı bu proteinler üzerinde çalışmaktadırlar. Bu çalışmalara rağmen hala proteinlerin keşfedilemeyen yönleri araştırmacıların ilgisini çekmektedir. Bizde bu makalede proteinlerin açılarının tahmini için fidan gelişim algoritmasını kullanarak bir çözüm önerdik. Bu algoritmayı kullanmamızın nedeni hem yeni bir algoritma olması hem de NP türü problemlerin çözümünde kolaylık sağlamasıdır.

Bu tez çalışması aşağıdaki gibi organize edilmiştir. İkinci bölümde Biyoinformatik biliminden ve çağımızda bilgisayar teknolojilerinin biyolojiye uygulanmasından bahsedilmiştir. Üçüncü bölümde proteinler, protein yapıları ve çeşitleri hakkında geniş bilgiye yer verilmiştir. Dördüncü bölümde ise fidan gelişim algoritmasının esaslarından bahsedilmiştir. Beşinci bölümde fidan gelişim algoritmasının temel aşamalarının anlatılmış olup, bir sonraki bölümde Proteinlerin ikincil yapı tahmini için daha önce yapılmış çalışmalardan ve bu çalışmaların performansından bahsedilmiştir. Altıncı bölümde fidan gelişimi algoritmasının proteinlerin ikincil yapılarının tahminine uyarlanması verilmiştir. Son bölümde ise elde edilen sonuçlar tartışılmıştır.

2.BİYOİNFORMATİK

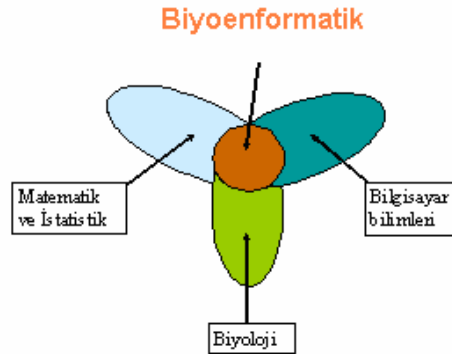
Biyoinformatik genel olarak biyolojik problemlerin çözümünde bilişim teknolojilerinin kullanılması olarak tanımlanabilir. Enformatik, matematik ve yaşam bilimlerini birleştiren ve işlevsel genomik ile birlikte, gen ve protein işlevlerini anlamaya yönelik bilim dalıdır.

Biyoinformatik ; temelde algoritma (herhangi bir sorunun çözümü için izlenecek yol) çıkarmak ve veri tabanı işlemekle uğraşan , protein ve gen dizilimleri ile ilgili bilgilerin işlenmesini ve derlenmesini hedefler. Başka bir deyişle çeşitli biyoloji veri bankalarından gelen bilginin anlaşılır ve organize hale getirilmesi için informatik tekniklerinin kullanımıdır.[4]

20. yüzyıldan sonra biyoloji biliminde köklü değişiklikler meydana gelmiştir. Bu yüzyıldan sonra biyoloji bilimi laboratuarda yapılan bir bilim olmakla kalmayıp bilgi teknolojisinden de üst düzeyde faydalanan bir bilim haline gelmiştir.

Bilgisayarların moleküler biyolojide kullanımı üç boyutlu moleküller yapıların grafik temsili, moleküler dizilimler ve üç boyutlu moleküler yapı veritabanları oluşturulmasıyla başlamıştır. Kısa sürede çok yüksek miktarda veri üreten, endüstri düzeyinde gen ekspresyonu, protein protein ilişkisi, biyolojik olarak aktif molekül araştırmaları, bakteri, maya, hayvan ve insan genom projeleri gibi biyolojik deneylerin doğurduğu talep bu alanda gelişmeleri hızla arttırmıştır.

Biyolojinin bilişim teknolojileri ve bilgisayar bilimini bir araya getiren bir bilim dalı haline gelmesi ile birlikte, bu bilimin en büyük amacı genom düzeyinde saklanan ve gerektiğinde kullanılan yaşam şifresinin keşfedilmesi olmuştur.



Şekil 2.1 Biyoinformatik ve diğer bilim dalları arasındaki ilişki

2.1 .Biyoinformatik Veri Tabanları

İnternet teknolojisi bilimsel bilginin oldukça hızlı yayılmasına neden olmuştur. Özellikle biyoloji araştırmacıları internetin gücünden faydalananarak biyolojik kaynaklar oluşturmuşlardır. Bu

kaynakların en önemli sınıfını internet üzerinden erişilebilen veritabanları oluşturmaktadır. Biyoinformatik veritabanlarının gelişimi şu şekilde olmuştur.

- İlk Biyolojik Veritabanı - Protein Identification Resource Margaret Dayhoff tarafından 1972’de kuruldu.
- Dayhoff ve arkadaşları protein ailelerini organize ettiler.
- Protein dizilerinin karşılaştırılması fikri doğdu.
- Şu an iki büyük protein bankası var : SwissProt, PIR(Protein information Resource)
- İlk DNA veritabanı 1979’da kuruldu. Şu an çok sayıda veritabanı var: GenBank, EMBL(the European Molecular Biology Laboratory), DDBJ(DNA Data Bank of Japan).....

İnternette en çok rastlanılan veritabanları genomik veritabanlarıdır. Sekanslama çalışmaları tüm dünyada, farklı ülkelerde pek çok laboratuarda birden gerçekleştirildiği için ortaya çıkan genomik bilginin bir araya getirilmesi oldukça güç bir iştir. Tüm dünyada tanımlanan bu genomik bilginin toplanması adına ciddi uluslararası çabalar bulunmaktadır. Son derece güncel bir konu olan İnsan Genom Projesi çerçevesinde Amerika’da bulunan National Science Foundation isimli bir kuruluşun ilgi alanları arasında biyolojik veri tabanları geliştirmek vardır. Japonya’da ki DDJB ve Avrupa’da ki EMBL önemli gen veri bankalarıdır. Bununla birlikte Amerika’da Ulusal Hayvan Genomu Araştırma Programının (The National Animal Genome Research Program) amacı temel hayvan grupları (domuz, hayvan, sığır ve kümes hayvanları) için veri tabanı programları geliştirmektir. Bunlar, belirtilen hayvanın genlerinin tam düzenlenişini içeren genomik veri bankaları olarak hizmet göreceklerdir. Böylelikle veri tabanında saklanacak olan bilgi; hayvanlar arasında karşılaştırmalı çalışmalar yapmak için, genler ve fonksiyonları arasındaki korelasyonu kolaylaştırmak için ve ayrıca DNA’daki genlerin nispeten pozisyonlarını tanımlamak için bir temel oluşturacaktır. Bu çerçevede hazırlanan uspigbase isimli domuz veri tabanı programı şimdi kullanılabilir durumdadır.

İnternetteki diğer önemli veritabanları taksonomik bilgi içeren veritabanlarıdır. Bu veritabanları ile dünyanın biyolojik çeşitliliği internet ortamına aktarılmaktadır. Örnek olarak Agricultural Research Service tarafından hazırlanan Grin Taxonomy adındaki veritabanında, özellikle çoğunluğunun ekonomik bitkilerin oluşturduğu, 17.000 cins kayıtlıdır . Amerika’da Beyaz Saray Biyolojik Çeşitlilik ve Ekosistem Dinamikleri Alt Komitesi, sistematik çalışmalarının ekosistem yönetiminin ve biyolojik çeşitliliğin korunmasını için esas olduğunu ortaya koymuş ve bu yönde Birleştirilmiş Taksonomik Bilgi Sistemi adı altında bir program

kurulmuştur . Dünyanın Memeli Türleri (Mamalian Species of the World) isimli veritabanı 4.629 türü içermektedir ve sınıflandırma araştırmalarda bir çevrimiçi referans olmayı amaçlamaktadır . Veri tabanı programlarının metin, resim, ses ve görüntü biçimlerinde bilgi saklayabilmesi nedeniyle etkinliği oldukça yüksektir. Bir veri tabanı programı *tablolar, formlar, sorgular* ve *rapor* olmak üzere dört temel unsura dayanmaktadır. Tablolar, toplanan bilgilerin ayrı ayrı saklanması için özel olarak hazırlanmaktadır. Formlar, veri tabanı ile kullanıcı arasında iletişim sağlayan ara birimdir. Veri tabanındaki bilgiler formlar aracılığı ile kullanıcıya sunulmaktadır. Ayrıca kullanıcı da formlar ile veri tabanındaki bilgiler içinde gezinebilmekte ve bilgi girişi yapabilmektedir. Sorgular toplanan bilgiler ile ilgili hazırlanır. Sorgular, veri tabanındaki bilgileri işler ve istenen sonuçları hazırlar. Raporların kullanılmasıyla ise sorgu ile elde edilen sonuçlar okunabilir ve basılabilir formatlarda kullanıcıya sunulur . Günümüzde kolay kullanılabilirliği, depolanan verileri indeksleyerek saklayabilmesi ve hızlı olması nedeniyle veri tabanları web ortamında da kullanılmaktadır. İçeriği fazla olan hemen hemen bütün siteler, veri tabanı kullanılarak hazırlanmaktadır. Böylelikle sitelerin güncellenmesi için statik sayfaların tekrar yapılmasına gerek kalmamaktadır.

İnternet tabanlı veri tabanı programlarına ulaşabilmek kolay olduğu için konu ile ilgili olan araştırmacı ve öğrencilere büyük faydalar sağlamaktadır. Ancak pek çok araştırmacı kendi bilimleri ile oldukça meşgul oldukları için kullanacakları veritabanı programlarının kullanışlı ve güncel olmasının yanında kompleks ara yüzlerini kullanmayı öğrenmek zorunda bırakmamasını da tercih etmektedir . Bununla beraber internetteki yurt dışı kaynaklı veri tabanları konu ile ilgilenenlere geniş bir bilgi kaynağı sunmaktadır ancak ülkemizdeki araştırmacılar ve öğrenciler dil problemi yüzünden bu veri tabanlarından yeterince faydalanamamaktadır. Bununla birlikte internette Türkçe veri tabanları da oldukça azdır. İyi tasarlanmış bir ara yüz kullanıcının programı anlaması için gerekli zamanını indirgemesinin yanında meydana gelebilecek hataları ve yanlış anlamaları da minimuma indirecektir. Sonuç olarak ulaşılabilirlik programın kullanışlı olmasına bağlıdır, kullanışlılık ise programın üretkenliği ile ciddi bir şekilde ilişkilidir. Dünya, sanayi toplumundan ya da eski ekonomi sisteminden bilgi toplumuna ya da yeni ekonomi sistemine geçişe neden olan hızlı ve baş döndürücü bir değişim yaşamaktadır. Önümüzdeki yıllar ülkeler, ülkeler içinde kurumlar ve hatta bireyler arasında bilgi sistemleri (BS) ve bilgi teknolojileri (BT) alanlarında kıyasıya bir rekabete ve yarışa sahne olacak gibi görünmektedir. Avrupa Birliği'nin bilgi toplumu projelerine verdiği önem ve ayırdığı kaynaklar, İngiltere Başbakanı Blair ve Almanya Başbakanı Schröder'in geçtiğimiz yıl boyunca yapmış olduğu açıklamalar, son G-8 zirvesinde alınan kararlar, ABD Başkanı Clinton'un bizzat açıkladığı yeni ekonomi paketi bu değişim rüzgarının ilk adımları durumundadır. Bu konuya somut bir örnek Almanya'nın bu yıl içinde Hindistan, Ermenistan, Bulgaristan vb. ülkelere 80

bin BS/BT uzmanına kapılarını açmasıdır . Biyoinformatik veri tabanları programları bilgi kavramı için önemli sayılan aşağıdaki özellikleri kazandırmaktadır.

1. Gelişen erişim olanaklarının elvermesi ile bilgi bölünmeden ortaklaşa kullanılabilir. kullanılabilmektedir.
2. Etkileşimli ortamlarda oluştuğu anda bilgisayara aktarılan bilgi sürekli olarak kendiliğinden artmaktadır.
3. Büyük boyutlardaki bilgi içerisinden gerekli olana erişim gibi oldukça önemli bir problemi ortadan kaldırmaktadır.
4. Veriyi işlemek, yeni bilgi oluşturmak, ondan yararlanabilmek veritabanları ile daha kolay hale gelmektedir.
5. Bilginin güncellenmesi, her zaman en son durumu göstermesi veri tabanlarının önemli bir özelliğidir.

Milyonlarca nükleotidin depolanması ve organizasyonu için veritabanlarının oluşturulması, araştırmacıların bu bilgilere ulaşabilmeleri ve yeni veriler girebilmeleri için ilk aşamadır. Biyoinformatik'te nükleotid dizi bilgilerinin organizasyonu ve depolanması görevini üstlenmiş üç kuruluş vardır. Genbankası (GenBank), Avrupa Moleküler Biyoloji Laboratuvarı (EMBL) ve DNA Japonya veritabanıdır (DDBJ). Bu üç kuruluş, araştırmacıların yararlanmasına açık, nükleotid dizi bilgilerinin toplanması ve yayılmasında işbirliği içinde çalışmaktadır. Gen Bankası ABD'de Maryland, Bethesda'da, Avrupa Moleküler Biyoloji Laboratuvarı, İngiltere'deki Hinxton'da ve DNA Japonya veritabanı ise Japonya'da Mishima'da yeni dizi bilgilerinin alışverişinde,İnternet üzerinde günlük olarak e-mail, ortak kullanılan ftp ve www üzerinden hizmet sunmaktadırlar. Protein dizi verileri ile ilgili başlıca hizmet sağlayıcılar ise GenBank, EMBL, PIR International ve Swiss-Prot'tur. NIH'in National Center for Biotechnology Information merkezi, biyoinformatik gereci sunan başlıca web sayfalarından biridir. Ve günde 3 milyon araştırmacı tarafından ziyaret edilmektedir.

Veritabanlarındaki bazı sorunlar; vektöriyel dizilerle kirlilik, bir gene ait dizi parçacıklarının veritabanına birden çok kez girilmesi ile ortaya çıkan kalabalık, aynı gene ait birden fazla EST (Ekspresed Sequence Tag) içeren EST veritabanlarının olması gibi durumlardır. Bu durumlar; genom projelerinin ileri aşamalarını oluşturan UNIGENE, VecScreen gibi projelerle ortadan kaldırılmaya çalışılmaktadır.

2.2 Biyolojik Bilgi Analizi

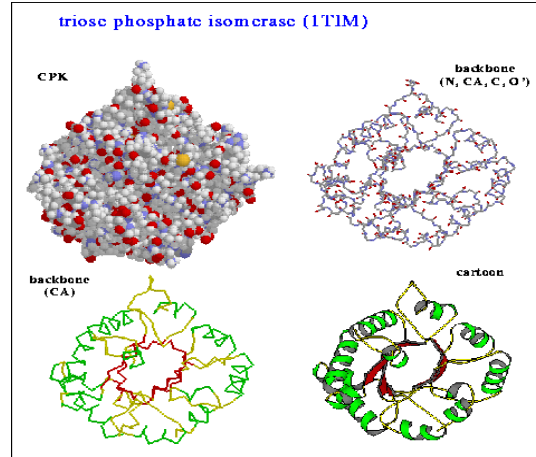
Gün geçtikçe globalleşen veya küreselleşen dünya, bilginin iktidar olduğu bir döneme girmiş bulunmaktadır. Gelecekte dünyanın bilgi ile yönetileceği; bilgi zenginleri ve bilgi fakirleri arasındaki farkın inanılmaz boyutlara varacağı bir gerçektir. Bunu fark eden liderler ve

yöneticiler ülkelerinin ve kurumlarının vizyonlarını çizmektedirler. Bu hızlı değişim sürecinde Devlet ve kurumlarının, teknolojiye, bireylerdeki ve iş dünyasındaki gelişimi önceden görmesi, algılaması ve onlara belirli konularda destek olurken ve vizyon çizerken, bazı konularda da önlem alması gerekmektedir .Bu yüzden biyoinformatikte bilgi analizi çok önemlidir.

Protein dizisinden Yapısına

APRKFFVGGNWKMNKD
KKSLLGELIHTLNGAKL
SADTEVVCGAPSIYLD
FARQKLDKIGVAAQN
CYKVPKGAFTGEISPA
MIKDIGAAWVILGHSE
RRHVFGEDELIGQKV
AHALAEGLGVIACIGE
KLDEREAGITEKVVFE
QTKAIADNVKDWSKVV
LAYEPVWAIGTGKTAT
PQQAQEVHEKLRGWLK
SHVSDAVAQSTRIYG
GSVTGGNCKELASQHD
VDGFLVGGASLKPEFV
DIINAKH

=



Şekil 2.2 Protein dizileri ve yapısı [5]

Biyoenformatiğin ikinci özelliği; saklanan biyolojik bilgilerin analizidir. Analiz kapsamına giren konular .

1. Çeşitli organizmalardaki DNA dizilerinin hangi genlere ait olduğunu belirlenmesi
2. Yeni keşfedilen proteinlerin ve RNA dizilerinin yapı işlev ilişkilerinin belirlenebilmesi için yöntem geliştirilmesi
3. Protein dizilerinin ilgili gen ailelerine kümelernmesi ve protein modellerinin geliştirilmesi
4. Benzer proteinlerin sıralanarak gelişimsel ortaya çıkaracak filogenetik ailelerin oluşturmak[5].

2.3 Biyoinformatik Çalışma Alanları

Biyoinformatik genel olarak biyolojik problemlerin çözümünde bilişim teknolojilerinin kullanılması olarak tanımlanabilir. En dar tanımı ile genomik sekansları destekleyen biyolojik veritabanlarının oluşturulması ve işletilmesi, en geniş tanımı ile de mevcut tüm bilgisayar

uygulamalarının biyolojik problemlerin çözümünde kullanılması olarak anlaşılır. Biyoinformatik modern biyolojinin iki temel bilgi akışını kapsar.

1. Genetik bilgi akışı: Bir organizma'nın DNAsı incelenerek özelliklerinin belirlenmesinden, incelenen bu organizma türünün oluşturduğu toplulukların karakteristik özelliklerine kadar olan bilgi akışı. Elde edilen DNA bilgisi tekrar genetik havuzun tanımlanması için kullanılır.

2. Deneysel bilgi akışı: Biyolojik olaylar gözlenerek elde edilen enformasyon, açıklayıcı modeller ile tarif edilir, daha sonra bu modellerin doğruluğu yeni deneyler ile test edilir

Son yirmi yılda temel biyolojik araştırmaların klinik tıp uygulamaları ve klinik tıp bilgi sistemleri üzerindeki etkisi daha da belirleyici olmuş ve bugün yeni kuşak epidemiyolojik, tanı, teşhis ve tedavi amaçlı modüllerin ortaya çıkmasına yol açmıştır. Biyoinformatik çalışmalar temel bilimsel araştırmalara yönelik görünmekle beraber önümüzdeki on yıl içinde klinik bilişim için vazgeçilmez olacaktır. Örneğin hastaların medikal formlarında giderek artan bir sıklıkla DNA dizilim bilgileri yer almaya başlayacaktır. Bugün ABD'de bazı sigorta şirketleri, risk primleri belirlenirken mevcut genetik tarama test sonuçlarını talep edebilmektedir. Biyoinformatik araştırmalar için geliştirilen algoritmaların çok yakında klinik bilişim sistemlerine entegre olması beklenmektedir[6].

Biyoinformatik araçları genel olarak araştırma çalışmalarında kullanılmaktadır. Bu araştırma konuları metodolojik çalışmalar ve biyolojik çalışmalar olmak üzere başlıca iki başlık halinde toplanabilir.

Metodolojik Çalışmalar:

- DNA dizilim araştırmaları
- Protein dizilim araştırmaları
- Makromoleküler yapıların üç boyutlu dizilim araştırmaları
- Küçük moleküllerin ligandlarıyla etkileşiminin araştırılması
- Heterojen biyolojik veritabanlarının entegrasyonu
- Biyolojik bilginin paylaşımının kolaylaştırılması
- Bilgisayar ile otomize edilmiş veri analizi
- Etkileşimde bulunan gen ürünleri için bilgi ağları oluşturulması

- Biyolojik faaliyet sürecinin simülasyonu
- Büyük çaplı projelerden çıkan sonuçların değerlendirilmesi (insan genom projesi)
- Metabolik yol izleri ve hücre algılama modellemesi
- Protein-protein ilişkileri
- Protein familyalarının nasıl evrimleştiğinin mekanizmasının anlaşılması
- Değişik hücre ve dokulardaki proteinlerinin haritalarının çıkarılması

Biyolojik çalışmalar:

- Protein yapı ve fonksiyonunun belirlenmesi
- Herhangi bir biyolojik fonksiyonu artıran veya engelleyen küçük moleküllerin tasarlanması
- Karmaşık genetik fonksiyon yada regülasyon faaliyetlerinin tanımlanması
- Tıbbi yada endüstriyel amaçlı yeni makro moleküller üretmek
- Genetik faktörlerin, hastalık yatkınlığına etkilerini ortaya çıkarmak

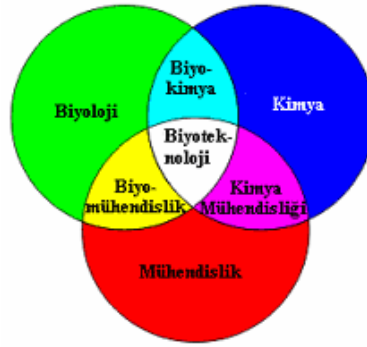
Yapılan dizi analizi çalışmaları; yeni genlerin bulunması, genlerin yapı analizinden fonksiyonlarının tayini ve bir genin yapısındaki değişimin hastalıklarla ilişkisinin araştırılmasında fayda sağlar. Moleküler modülleme çalışmaları ise; bir proteinin üç boyutlu yapısının, fonksiyonu ile nasıl bir bağlantısının olduğunu anlamaya yardımcı olur.

Günümüzde Biyoinformatik insan genomundaki genlerin sekanslanması ve haritalanmasından elde edilen yeni bilgilerin analizlenmesi ile yoğun olarak uğraşmaktadır. Yapılan bu çalışmalarla elde edilen bilgiler değişik genetik ve diğer hastalıkların daha iyi anlaşılmasına ve yeni ilaç hedeflerinin belirlenmesine fayda sağlayacaktır[6].

Sonuç olarak Biyoinformatik ;ilaç tasarımı,gen terapisi,diagnostik tarımsal rekolte (1 yıllık ürün), biyokimyasal prosesler gibi önde gelen biyoteknoloji alanlarında uygulama bulan bir disiplin olarak kendini gösterir[6].

3.BİYOKİMYA VE PROTEİNLER

Klasik bir tarife göre biyokimya, canlı organizmaların kimyasal yapısını ve hayatın devamı boyunca canlının içinde meydana gelen kimyasal olayları konu olarak ele alan ve inceleyen bir bilim dalıdır. Tıp ve sağlık bilimlerini de kapsamına alan biyoloji bilimi çok karmaşık ve komplike problemler içermektedir. Biyokimya bilim dalı ise deneysel metodlar ve aletler kullanarak bu sorulara ve problemlere çözüm aramaktadır.[7] Biyokimya bilimi, yaşamın en küçük birimi olan hücrenin kimyasal yapısını, canlının meydana gelişindeki, hayatının devamındaki ve nihayet yok oluşundaki kimyasal mekanizmaları konu olarak ele alan ve inceleyen bir bilimdir. Bu nedenle biyokimya büyük oranda metabolizma ile ilgilenmektedir. Bununla beraber bazı anormal patolojik koşullar altında metabolizmada görülen ayrıcalıklar, yalnız hastalıkların teşhisinde görülen ayrıcalıklar, yalnız hastalıkların teşhisinde rol oynamakla kalmayıp, aynı zamanda normal olayların daha derinliğine aydınlatılmasına da ışık tutacaktır.[7]



Şekil 3.1. Biyoloji, Kimya ve Mühendislik Bilimleri ve birlikte kullanım alanları

3.1 PROTEİNLER

3.1.1. Proteinin Tanımı

Proteinler doğada bulunan 300 den fazla amino asitden 20 tanesinin birbirine peptit bağlarıyla bağlanarak oluşturduğu dev yapılı moleküllerdir. Bu moleküllerin ağırlığı Dalton cinsinden ölçülmektedir (1 Dalton=1H atomu ağırlığı) Bir protein molekülü birkaç bin ile birkaç milyon Dalton ağırlığında olabilmektedir.

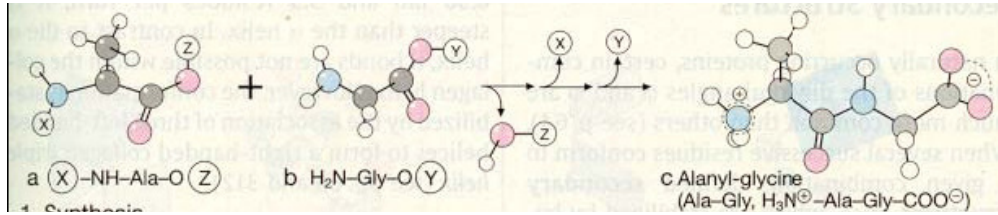
Proteinler, yapı ve fonksiyon maddeleri olarak, biyokimyasal açıdan çok önemli bir yere sahiptirler. Canlı organizmanın doku hücrelerinde proteinlerin yer almadığı biyokimyasal bir olaya rastlamak hemen hemen imkansızdır. Aslında bu moleküller canlı bünyesinde en fazla işleve sahip moleküllerdir. Enzimlerin ve polipeptit hormonların yapısında bulunmak suretiyle vücuttaki metabolizmanın düzenlenmesini sağlaması, kastaki kontraktıl proteinler sayesinde hareketin sağlanması kemikteki kollojen isimli proteinlerin bir betondaki çelik demirler gibi davranarak bu yapıyı güçlendirmesi , dolaşımda hemoglobin ve plazma albümini gibi

proteinlerin hayat için gerekli molekülleri taşınması, immunglobulinlerin enfeksiyöz bakteri ve virüsleri yok etmesi gibi daha sayılamayacak kadar çok yararlarının olması proteinler hakkında yukarıdaki söylenenleri haklı çıkarmaktadır[1].

3.1.2. Proteinlerin Yapılarındaki Bağlar

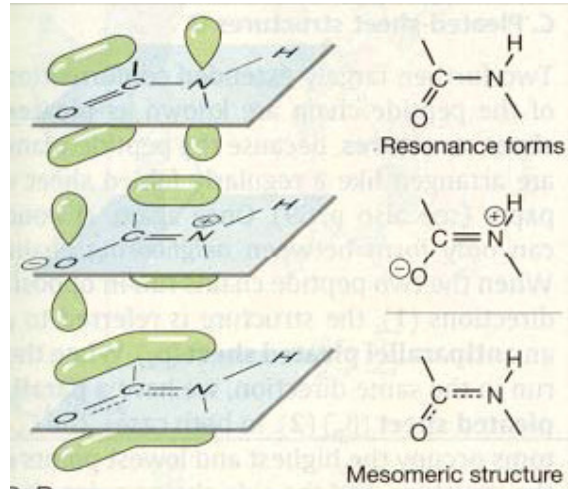
Proteinlerin yapılarında kovalent bağlar ve kovalent olmayan bağlar vardır. Proteinlerin yapılarındaki kovalent bağlar, peptit bağları ve disülfid bağlarıdır; kovalent olmayan bağlar ise hidrojen bağları, iyon bağları ve apolar bağlardır[8].

Peptit bağları: Bir amino asidin α -karboksil karbonu ile bir başka amino asidin α -amino azotu arasında oluşan C-N bağıdır.



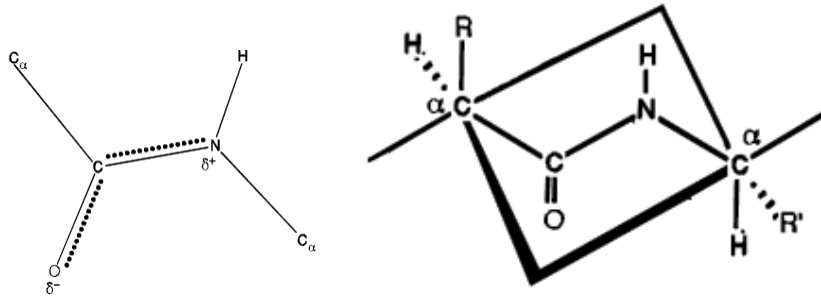
Şekil 3.2 Karbon ve Azot atomları arasındaki peptit bağları

Peptit bağının uzunluğunun C=N çift bağının uzunluğundan büyük C-N tek bağının uzunluğundan küçük olması nedeniyle peptit bağının kısmen çift bağ olduğu kabul edilir. Bunun da nedeni, peptit bağlarında rezonans veya mezomeri denen durum olmasıdır:



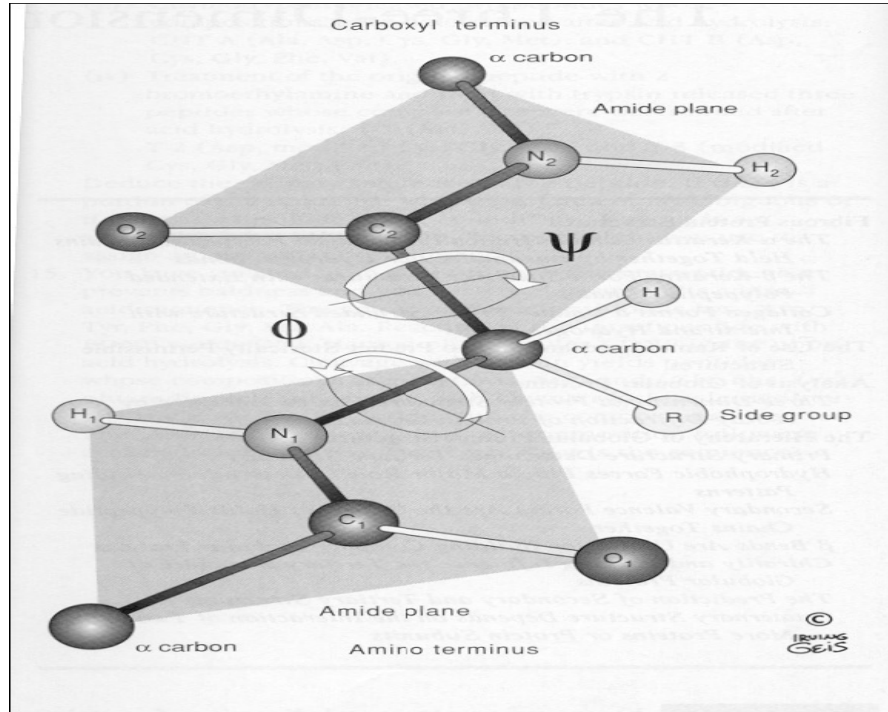
Şekil 3.3 Atomlar arasındaki mezomeri yapısı

Peptit bağlarındaki rezonans veya mezomeri nedeniyle OC - N bağı %50 çift bağ niteliği kazanmıştır.



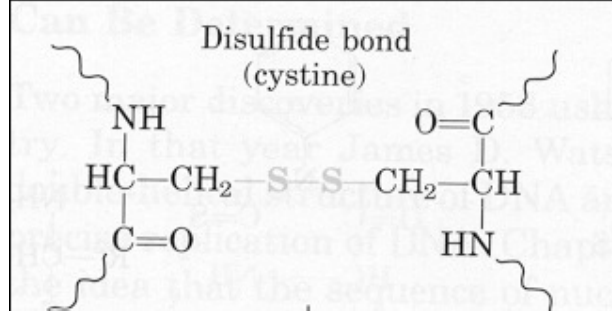
Şekil 3.4 Atomlar arasındaki OC-N bağları

Peptit bağının iki yanındaki α -karbon atomları, tek bağ etrafında ϕ ve ψ ile gösterilen açılarla dönüş yapabilirler. ϕ , α C-N bağının dönüş açısıdır; ψ ise α C-C bağının dönüş açısıdır.



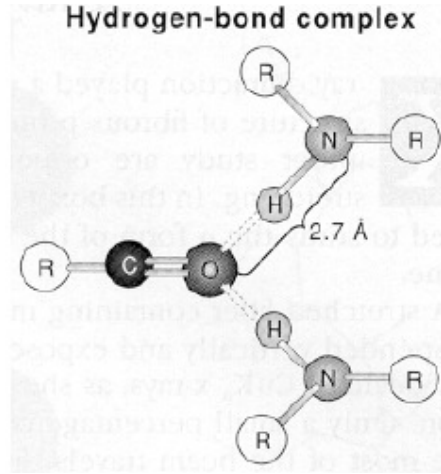
Şekil 3.5 Peptit bağları ve atomlar arasında dönüş açıları

Disülfid Bağları: İki sistein kalıntısı arasında, sülfhidril (tiyol, -SH) gruplarının H kaybetmeleri sonucu oluşan S-S bağlarıdır[8]. Disülfid bağlarının bir protein molekülünün şeklinin oluşmasında ve korunmasında önemli etkisi vardır. Disülfid bağları, bir polipeptit zinciri içerisinde kurulabilir veya çeşitli polipeptit zincirlerinin birbirine bağlanmasını sağlayabilir. Disülfid bağları, ribonükleaz, oksitosin ve vazopressinde aynı polipeptit zincirinde bulunur; insülinde ise iki ayrı polipeptit zincirini birbirine bağlar. [8]



Şekil 3.6 Disülfid bağları

Hidrojen bağları: Polipeptit zinciri oluşturan peptit bağlarındaki rezonans veya mezomeri durumundan dolayı, oksijenlerin bilinen keto gruplarından daha negatif, azotların ise pozitif özellik taşımasının sonucu olarak, bir polipeptit zincirdeki bir peptit düzleminde bulunan oksijen atomu ile bir başka peptit bağı veya düzlemindeki azot atomu arasında, aradaki uzaklık yaklaşık $2,7 \text{ Å}$ olduğunda, hidrojen köprüsü şeklinde ($\text{C=O} \cdots \text{H} \cdots \text{N}$) oluşan bağlardır:



Şekil 3.7 Hidrojen bağları

İyon bağları: Polipeptit zincirlerindeki asidik ve bazik amino asit kalıntılarının fonksiyonel gruplarının fizyolojik pH'da tamamen veya kısmen iyonlaşmış halde bulunmalarının sonucu olarak, elektronegatif ve elektropozitif gruplar arasında gelişen elektrostatik çekim kuvveti ile ($\text{COO}^- \cdots \cdots \text{H}_3\text{N}^+$) oluşan bağlardır.

Apolar bağlar (hidrofob bağlar): Polipeptit zincirindeki amino asit kalıntılarının metil grubu, alifatik grup, sıklık grup gibi apolar kısımlarının birbirlerine yeter derecede yakın olmaları halinde geçici bir polarite göstermelerinin sonucu ortaya çıkan ve Van der Waals-London çekme kuvveti diye bilinen zayıf çekme kuvveti ile ($\text{CH}_3 \cdots \cdots \text{CH}_3$) oluşan bağlardır.

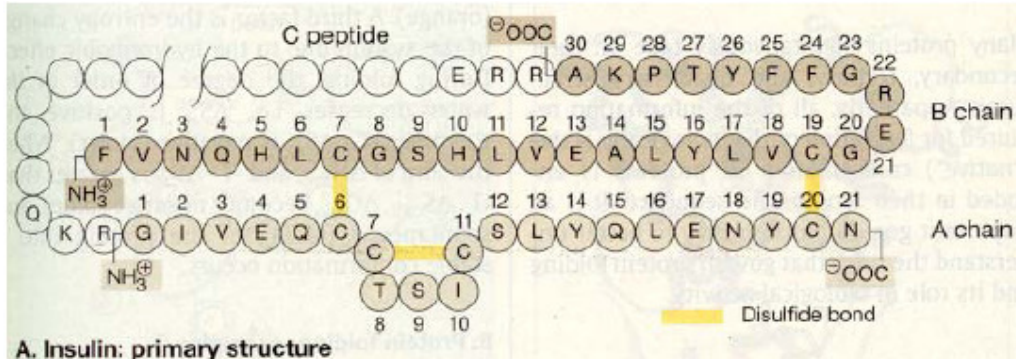
Hidrofobik etkileşimler gerçek bağ değildirler; elektron paylaşımı yoktur. Hidrofobik etkileşimler, proteinlerin iç kısımlarının kararlı olarak devamlılığının sağlanmasında rol oynar.

3.1.3. Protein Moleküllerinin Yapısı ve Konformasyonu

Proteinlerde birinci (birincil), ikinci (sekonder), üçüncü (tersiyer) ve dördüncü (kvarterner) yapı diye dört yapı tanımlanır.

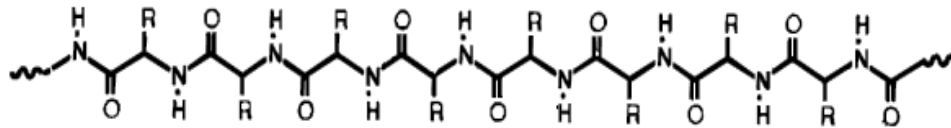
Bir proteinin birincil (birincil) yapısı

Bir proteinin birincil (birinci) yapısı, bir protein için karakteristik ve genetik olarak tespit edilmiş olan amino asit dizilişidir; belirli türde, belirli sayıda, belirli diziliş sırasında amino asitlerin birbirlerine peptit bağlarıyla bağlanarak oluşturdukları bir polipeptit zinciri biçimindeki yapısıdır:



Şekil 3.8 Proteinlerin birincil yapısı

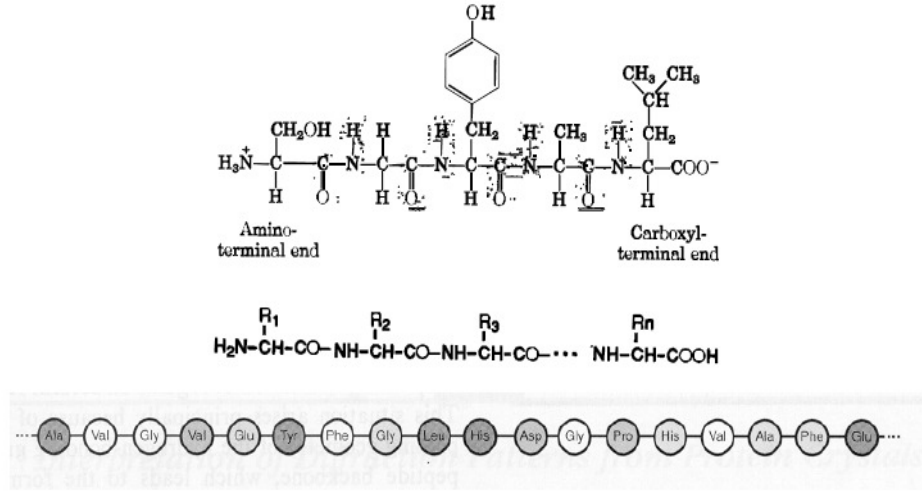
Bir polipeptitteki amino asit üniteleri, sıklıkla amino asit kalıntıları olarak isimlendirilirler. Basit olarak bir polipeptit zincirden ibaret olan protein, zincir başındaki amino asit kalıntısında serbest bir α -amino grubuna sahiptir; zincir sonundaki amino asit kalıntısında ise serbest bir α -karboksil grubuna sahiptir. Buna göre bir protein polipeptit zincirinin bir ucu amino terminal uç veya N-terminal uç; diğer ucu karboksil terminal uç veya C-terminal uç olarak isimlendirilir.



Şekil 3.9 Birincil yapıdaki tek bağlarla bağlı atomlar zinciri

Bir proteinin birincil yapısının oluşmasını ve sürdürülmesini sağlayan, peptit bağlarıdır. Bir proteinin birincil yapısında, peptit bağları ile birbirine tek bağlarla bağlı bir atomlar zinciri

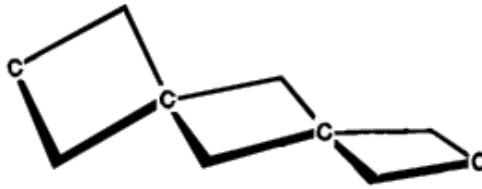
oluşturmuştur; diğer atom ve kalıntılar bu atomlar zincirinin dışında ve değişik yönlerde bulunurlar.



Şekil 3.10 Proteinlerin birincil yapısı ve atomlar arasındaki bağ yapısı

Peptit bağındaki rezonans veya mezomeri nedeniyle OC-N bağı %50 çift bağ niteliği kazanmıştır; çift bağların eksen etrafında dönmeleri sınırlı olduğundan peptit bağı oluşumuna katılan grupların atomları (3C, O, N ve H atomları) bir düzlemde bulunurlar. Peptit bağının iki yanındaki α-karbon atomları, tek bağ etrafında φ ve ψ ile gösterilen açılarla dönüş yapabilirler; φ, αC-N bağının dönüş açısıdır; ψ ise αC-C bağının dönüş açısıdır. Polipeptit zincirinin omurgasının uzaydaki düzeni, bu iki dönüş açısıyla belirlenir. Peptit bağının iki yanındaki α-karbon atomlarının tek bağ etrafında φ ve ψ ile gösterilen açılarla dönüş

yapmalarının sonucu olarak polipeptit zincirde art arda gelen küçük peptit düzlemleri aynı düzlemde bulunmazlar.



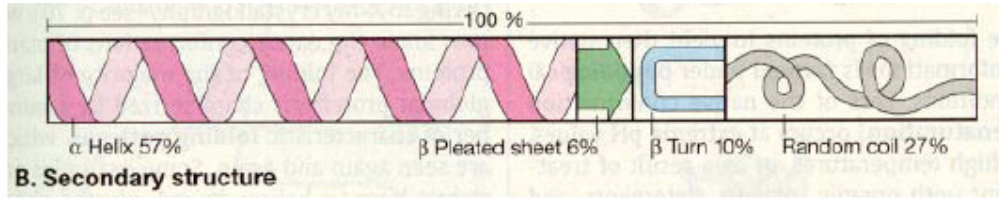
Şekil 3.11 Polipeptit zincirde ardı ardına gelen küçük peptit düzlemleri

Polipeptit zincirinin omurgasındaki atomların 1/3'ünün hareketli, 2/3'ünün hareketsiz oluşu yüzünden omurgada bir yarı sertlik vardır.

Bir proteinin ikincil yapısı

Bir proteinin ikinci (sekonder) yapısı, yarı sertleşmiş polipeptit zincirlerinin bükülmeler ve katlanmalarla oluşturdukları özgün kangallar biçimindeki yapısıdır. Bir proteinin ikinci yapısının oluşmasını ve sürdürülmesini sağlayan, birincil yapı ile meydana gelen polipeptit omurgasının özelliği ve yan bağlardan özellikle hidrojen bağlarıdır.

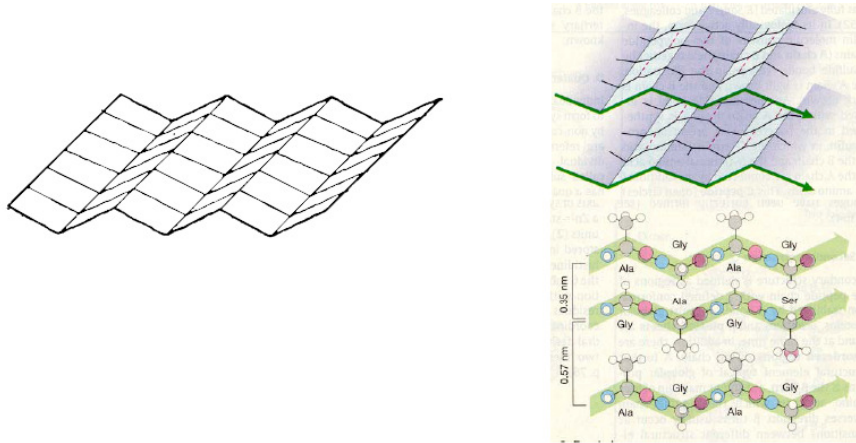
Bir proteinin birincil yapısı yani polipeptit zinciri omurgası oluşur oluşmaz, özgün diziliş sırasındaki amino asit kalıntılarının R- yan gruplarından uzanan özel kimyasal gruplar, özgün katlanmalar yönetirler. Proteinler için, gelişigüzel kangallanım, α -heliks yapısı ve β -konformasyonu veya kırmalı tabaka yapısı olmak üzere üç değişik ikincil yapı tanımlanır.



Şekil 3.12 Proteinin ikincil yapısı

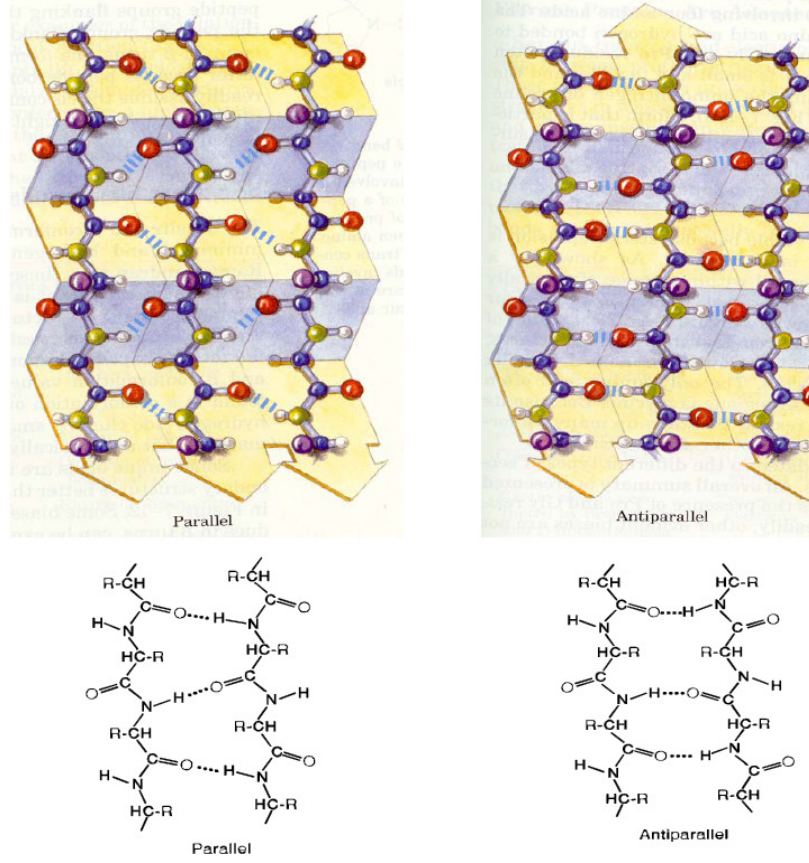
Proteinlerin gelişigüzel kangallanım tipi ikincil yapısında, polipeptit zincirin R- kalıntıları, α -karbonlar etrafında dönüşler yaparlar; fakat polipeptit zinciri boyunca tekrarlanmış bir örneğine rastlanmayacak biçimde davranışlar olur. Gelişigüzel kangallanımda hidrojen bağları ve diğer yan bağlar rol almaz; peptit bağları düzlemleri arasında kurucu bir ilişki yoktur.

Proteinlerin β -konformasyonu veya kırmalı tabaka yapısı tipi ikincil yapısında, molekülün şekli, kırmalı tabakalı görünümündedir.



Şekil 3.13 Proteinlerin β -konformasyonu veya kırmalı tabaka yapısı

Proteinlerin kırmalı tabaka yapısı tipi ikincil yapısında, polipeptit zincirdeki amino asit kalıntılarının R- grupları, ya aynı yönde sıralanma ile paralel dizilme gösterirler, ya da zıt yönlerde sıralanma ile antiparalel dizilme gösterirler.



Şekil 3.14 Kırmalı yapıdaki paralel ve antiparalel dizilimler

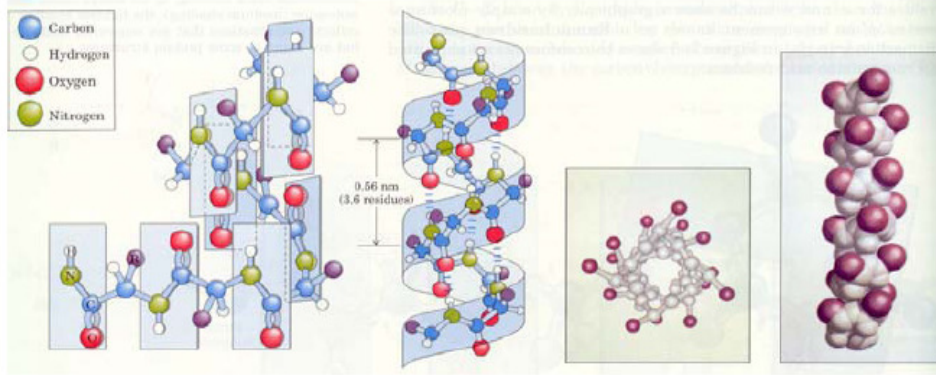
Proteinlerin kırmalı tabaka yapısı tipi ikincil yapısında, hidrojen bağları önemli rol almıştır.

Proteinlerin α -heliks yapısı tipi ikincil yapısında, polipeptit omurgası, oluşması mümkün olan bütün hidrojen bağlarının oluşması için, kıvrımları sağa dönen bir heliks biçiminde bükülmüştür:



Şekil 3.15 Proteinlerin α -heliks tipi ikincil yapısı

Proteinlerin α -heliks yapısı tipi ikincil yapısında, α -heliksin her kıvrımında 3,6 amino asit kalıntısı bulunur ve bir kıvrımın yüksekliği 0,56 nm kadardır; polipeptit zincirdeki amino asit kalıntılarının R- grupları, heliks yüzeyinden dışarı sarkmışlardır:



Şekil 3.16 Karbon, Hidrojen, Oksijen atomlarının α -heliks yapısı

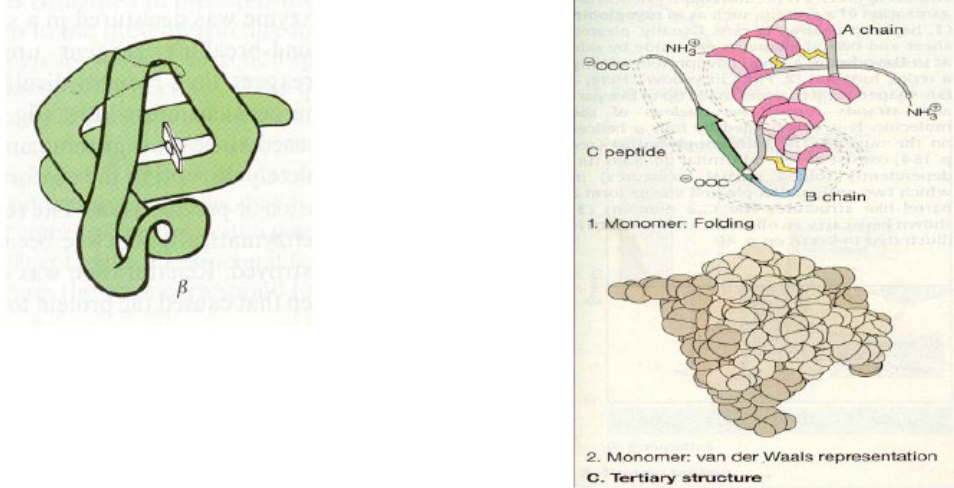
Polipeptit zincirdeki bazı amino asit kalıntıları, aynı zincirin iç hidrojen bağlarının oluşumuna, yapısı ve bulunduğu koşullar nedeniyle olanak vermez; heliks oluşumuna engel olabilir. Yüklü bazı amino asitler heliks oluşumunu engellerler. Glutamik asitten zengin bölgeler, komşu glutamik asitlerin negatif yüklü karboksil gruplarının birbirini itmesi nedeniyle hidrojen bağlarını stabilize edemez ve α -heliks oluşturamaz. Aynı şekilde, pozitif yüklü R grupları olan lizin ve arjinin de α -heliks oluşumunu önler. α -heliks yapısında genellikle pozitif yüklü amino asitler, negatif yüklü amino asitlerden üç kalıntı uzakta bulunur; iki aromatik halkalı amino asit sıklıkla üç kalıntı uzaklıktadırlar ve hidrofobik etkileşime neden olurlar. Prolinin peptit bağına giren azot atomu rijit halka yapısının bir parçası olduğundan, α -heliks oluşumunda prolin, N-C bağının dönüşünü engeller; bu nedenle prolin, nadiren α -heliks içinde yer alır. Heliks kıran veya kalıcılığını bozan glisin, prolin, serin, treonin, asparajin gibi amino asit kalıntılarının bulunduğu yerlerde gelişigüzel kangallanım meydana gelir. [P5]

Bir proteinin tersiyer (üçüncül) yapısı

Bir proteinin tersiyer (üçüncü) yapısı, polipeptit zincirinin, ikincil yapı oluşumundan sonra, daha önce açıklanan bağlayıcı güçlerin hepsinin toplamı ile uzayda daha ileri katlanmalar veya lifler halinde düzenlenme sonucu oluşan globüler veya fibriler yapısıdır. Bir proteinin tersiyer yapısının oluşmasına ve bu yapının sürdürülmesine, birincil ve ikincil yapının oluşmasına katılan bağlardan başka Van der Waals çekimleri ve iyon bağları da katılır. Böylece, üç boyutlu, tam konformasyonlu ve yoğunlaşmış protein molekülü meydana gelir.

Bir proteinin üç boyutlu yapı şekli, sulu çözeltide çeşitlilik gösterir. Globüler proteinlerde eksenleri arasındaki oran 2:1'e varan rotasyon elipsoid şekil karakteristiktir. Fibriler

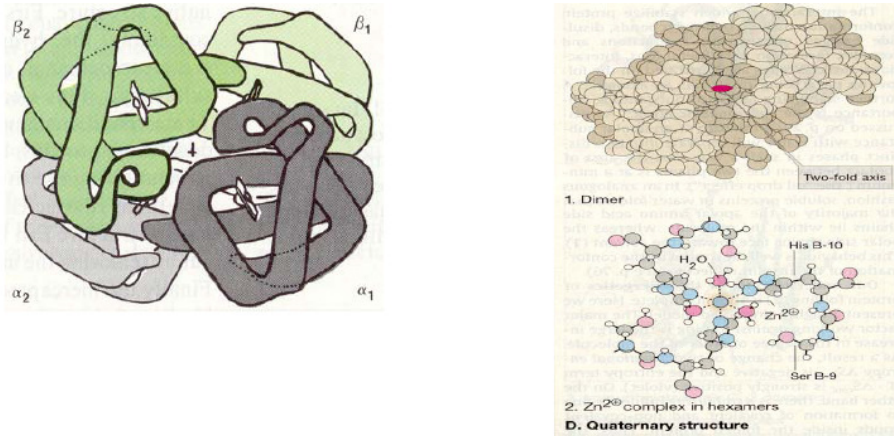
proteinlerde eksenleri arasındaki oran 30:1'e varan çok gerilmiş elipsoid şekil karakteristiktir. Kan serumundaki lipoproteinlerde küre şeklinde yapı gözlenir.



Şekil 3.17 Bir proteinin tersiyer (üçüncül) yapısı

Bir proteinin kuarterner (dördüncü) yapısı

Bir proteinin kuarterner (dördüncü) yapısı, birincil, ikinci ve üçüncü (tersiyer) yapıya sahip polipeptit zincirlerinin daha büyük yapıli agregatlar halinde biraraya gelmesiyle oluşan yapıdır:



Şekil 3.18 Bir proteinin kuarterner (dördüncü) yapısı

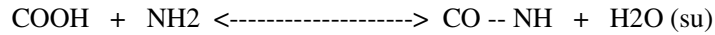
Her proteinin kuarterner yapısı olmayabilir, fakat molekül ağırlığı 100.000'nin üzerinde olan bir protein genellikle kuarterner yapıya sahiptir. Bir proteinin kuarterner yapısını oluşturan polipeptit zincirlerinin her birine alt birim veya monomer denir; bu monomerler, hidrojen bağları, Van der Waals çekmeleri ve iyon bağları etkisiyle polimerize olmuşlardır. Bir proteinin

kuarterner yapısını oluşturan monomerlerin reverzibl düzleşmeleri veya ayrılmaları, konformasyon değişikliğine yol açar. Proteinlerin spesifik biyolojik fonksiyonları bunların konformasyonlarına bağlı olduğundan, konformasyonda meydana gelen değişiklik, proteinin biyolojik aktivitesinin kaybolmasına neden olabilir.

3.2. Aminoasitler

Canlı organizmaların temelini nasıl hücreler meydana getiriyor ise, hücrelerin temelini de proteinler meydana getirir. Protein molekülleri hücreyi inşaa eden birer tuğla gibidir. Amino asitler ise proteinleri meydana getiren daha küçük moleküllerdir. Yani amino asitler uzun zincirler oluşturarak proteinleri, proteinlerde kompleks bir şekilde organize olarak hücreyi meydana getirir. Doğal olarak karmaşık bir yapıya sahip olan hücre yalnızca proteinlerden oluşmaz. Bunun yanında karbonhidratlar, yağlar, glikolipidler, fosfolipidler ve DNA - RNA molekülleri gibi kimyasal maddeler de hücrenin yapısına katılırlar. Fakat proteinsiz bir hücre düşünmek mümkün değildir.

Amino asit molekülleri, bir ucunda "amino grubu (NH₂) " diğer ucunda ise "karboksil (COOH)" grubu taşırlar.İşte amino asitlerin yan yana gelip zincirler oluşturarak proteinleri sentezlemesi, bu iki grubun aralarında kovalent veya iyonik bağ yapmasıyla gerçekleşir. İki amino asit yan yana geldiklerinde COOH ve NH₂ grupları arasında bağlanma meydana gelir ve bu bağa "peptid bağı" adı verilir.Bağlanma sırasında ise bir su molekülü sebest kalır. İki amino asitin yalnızca uç kısımlarını yani karboksil ve amino gruplarının bağlanması şu şekilde olur;

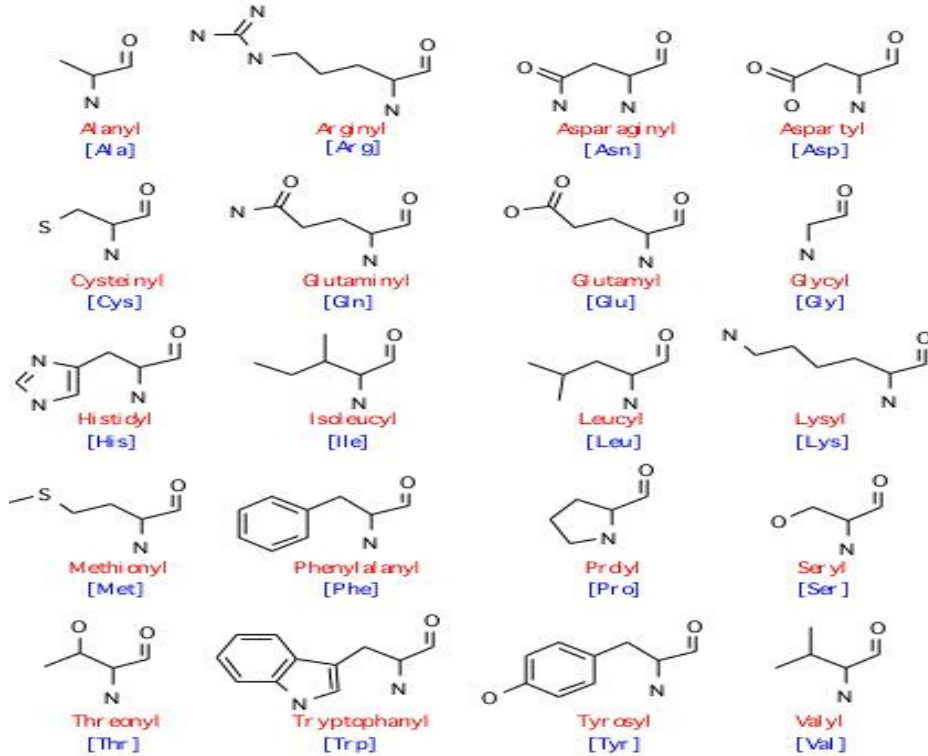


Denklemden COOH 1. aminoasitin bir ucu, NH₂ ise 2.amino asitimizin diğer ucunu temsil etmektedir. Bu uçlar yan yana geldiklerinde COOH grubundan bir oksijen ve NH₂ grubundan bir hidrojen serbest kalır. Böylelikle serbest kalan bu atomlar aralarında bağ yaparak suyu oluşturur. İki amino asitin yan yana gelmesiyle oluşan peptid bağına "dipeptid", üç veya daha fazla (yüzlerce yada binlerce) amino asitin yan yana gelmesiyle oluşan zincirdeki peptid bağlarına ise "polipeptid" adı verilir. Proteinler düz amino asit zincirlerinden meydana gelmesine rağmen oldukça karmaşık yapılara sahiptir.Bunun nedeni ise zincirdeki bazı amino asitlerin birbirleriyle ikinci veya üçüncü bir bağ yapmasındandır.

Amino asitler üzerlerinde belirli miktarlarda elektrik yükü taşırlar.Bu elektrik yükleri (+ veya -), asit veya baz özelliği gösteren bir ortama girdiklerinde nötrleşmeye başlar ve bu nötrleşme ortamın pH'ına bağlıdır. Bir amino asit ancak belirli bir pH noktasında nötr hale gelebilir ki, bu pH seviyesine o amino asitin "izoelektrik noktası" denir. Örneğin, histidin amino asiti, ancak pH'ı 7,47 olan hafif bazik bir sıvı içerisinde nötr hale gelebilir.

Protein molekülü, yalnız düz peptid zinciri şeklindeyse buna proteinlerin birincil yapısı denir. Moleküller, polipeptidler farklı kimyasal bağlarla ve değişik biçimde tutunarak proteinlerin ikincil ve üçüncül yapılarını oluştururlar. İkincil yapı helezon, üçüncü yapı ise küresel biçimdedir.

Vücutta bazı amino asitler birbirine çevrilebilir ve böylece amino asitlerin bir kısmı diğerinden oluşabilir. Bazı amino asitler ise bu şekilde yapılamaz yada yeterli miktar hızda oluşamaz. Vücutta sentezlenemeyen ve besinsel proteinlerle alınması zorunlu amino asitlere esansiyel amino asit denir. Esansiyel amino asitlerin tümünün gereksinmeyi karşılayacak miktarda ve düzenli olarak hayvanlara verilmesi zorunludur. Bunların tümü gerektiğinde vücut tarafından dışarıdan alınmazsa vücut proteinleri yeterli sentezlenemez, protein dengesi kurulamaz, hücre çalışmasında ve büyümede yetersizlik olur. Esansiyel amino asitler şunlardır; arginin, histidin, isoleusin, losin, lisin, methionin, fenilalanin, treonin, triptofan, valin.

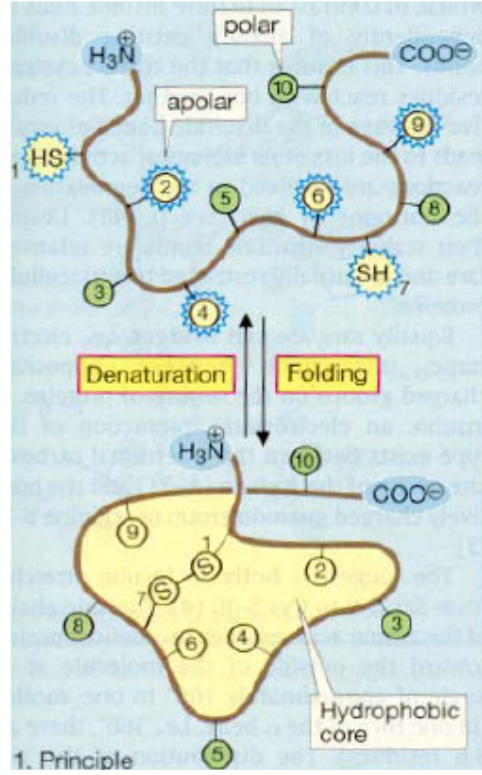


Şekil 3.19. Bazı Aminoasitlerin yapısı

3.3. Proteinlerin özellikleri

1) Proteinler, çeşitli etkilerle denatüre olurlar. Bir proteinin denatürasyonu, molekülündeki yan bağların yıkılması ile polipeptit zincirin katlarının açılması, gelişigüzel kangallanım yapısına dönüşmesi, sonra yeni bir biçimde yeniden katlanması olayıdır. Bir proteinin denatürasyonu, proteinin tersiyer yapısının bozulması, ikincil ve birincil yapısının

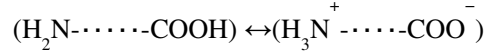
korunması biçiminde olursa reversibl (geri dönüşümlü)'dür. Denatüre olmuş bir proteinin tekrar eski haline dönmesine renatürasyon denir.



Şekil 3.20 Proteinlerde Denatürasyon

Bir proteinin denatürasyonu, proteinin tersiyer ve ikincil yapısının bozulması, yalnızca birincil yapısının korunması biçiminde olursa irreversibl (geri dönüşümsüz)'dür. Bir proteinin denatüre olmasıyla fiziksel ve kimyasal özelliklerinde değişimler görülür. Proteinin çözünürlüğü çok azalır, biyolojik aktivitesi kaybolur. Bir proteinin denatürasyonu, çoğu kez hidrojen bağlarını yıkan etkilerle olur. Bir proteinin denatürasyonuna neden olan etkiler şunlardır: Isı, X-ışını ve UV ışınlar, ultrason, uzun süreli çalkalamalar, tekrar tekrar dondurup eritmeler, asit etkisi, alkali etkisi, organik çözücülerin etkisi, derişik üre ve guanidin-HCl etkisi, salisilik asit gibi aromatik asitlerin etkisi, dodesil sülfat gibi deterjanların etkisi.

2) Proteinler, amfoter maddeler yani amfoter elektrolit veya amfolittirler; hem asit hem baz gibi davranma özellikleri vardır. Bir protein molekülü, her protein için farklı ve karakteristik olan, proteindeki elektriksel yüke sahip R- gruplarının sayıları ve elektriksel yüklerinin çeşidi tarafından belirlenen ve izoelektrik nokta diye tanımlanan bir pH değerinde iyonlaşmış fakat dış ortama karşı elektriksel yönden nötral bir yapıdadır.



Bir protein molekülü, izoelektrik noktasından düşük pH ortamında pozitif yüklü katyon ($H_3N^+ - \dots - COOH$) şeklinde bulunur; izoelektrik noktasından yüksek pH ortamında ise negatif yüklü anyon ($H_2N - \dots - COO^-$) şeklinde bulunur.

3) Proteinler, polipeptit zincirindeki peptit bağlarının su girişi ile yıkılması sonucu hidroliz olurlar. Proteinlerin kısmi hidrolizi ile proteozlar, peptonlar ve peptitler oluşur; tam hidrolizi ile amino asitler oluşur. Proteinlerin hidrolizi, kaynatma, asit etkisi ve enzim etkisiyle olabilir.

3.4. Proteinlerin Yapılarına Göre Sınıflandırılması

Basit proteinler

Basit proteinler, yalnızca amino asitlerden oluşmuş; hidroliz olduklarında sadece amino asitleri veren, polipeptit zincirleri yapısındaki proteinlerdir. Basit proteinler, değişik niteliklerine göre alt gruplara ayrılarak incelenirler:

1) Globüler proteinler: Molekülünün üç boyutlu şekli rotasyon elipsoid biçiminde olan proteinlerdir. Globüler proteinler de albüminler, globülinler, globinler, glutelinler, prolaminler, protaminler, histonlar gibi alt gruplara ayrılırlar.

Albüminler, suda ve sulu tuz çözeltilerinde çözünürler; ısı ile denatüre olurlar; sulu çözeltilerde amonyum sülfat ile doyurulmuş bir ortamda çökerler; molekül ağırlıkları genel olarak 100.000'in altındadır; glisince fakirdirler. Yumurta akında bulunan ovalbümin, kandaki serum albümin ve sütteki laktalbümin, hayvansal kökenli albüminlerdir; baklagillerdeki legumelin, hububattaki löykosin ise bitkisel kaynaklı albüminlerdir[8].

Globülinler, suda çözünmezler; sulu nötr tuz çözeltilerinde çözünürler; sulu çözeltilerinden, çözeltinin amonyum sülfat ile yarı doyurulması suretiyle çöktürme suretiyle ayrılabilirler; ısı ile de denatüre olurlar; molekül ağırlıkları 100.000'den yüksektir; glisince zengindirler. Yumurtadaki ovoglobülin, sütteki laktoglobülin, kan plazmasındaki α -, β -, γ -globülinler hayvansal kökenli globülinlerdir; kendir tohumundaki edestin, baklagillerdeki legumin, fasülyedeki faseolin ise bitkisel kaynaklı globülinlerdir. Globinler, Genellikle bileşik halde, başlıca hemoglobin yapısında bulunurlar. Glutelinler, bitkisel kökenli basit proteinlerdir; suda ve sulu nötr tuz çözeltilerinde çözünmezler; çok sulu asit ve alkalilerde çözünürler; ısı ile denatüre olurlar. Buğdayda bulunan glutenin, arpada bulunan hordenin ve pirinçte bulunan orizenin tipik glutelinlerdir; glutenin molekül ağırlığı 2-3 milyon kadardır.

Prolaminler, bitkisel kökenli basit proteinlerdir; suda, nötral tuzlarda ve alkolde çözünmezler; ancak %70-80'lik alkolde çözünürler; adlarını, çok fazla içerdikleri prolin amino asidinden alırlar; sistin ve lizin yönünden fakirdirler. Prolaminler, özellikle taneli bitkilerde çok bulunurlar; buğdayda bulunan gliyadin, mısırda bulunan zein, arpadaki hordein, önemli prolaminlerdir. Buğday unu hamuru akar su altında nişastasını kaybedince, geriye çok elastik bir madde kalır; Gluten adını alan bu madde, gliyadin ve glutenin karışımıdır.

Protaminler, suda, seyreltik asit ve alkalilerde, seyreltik amonyum hidroksit çözeltisinde çözünürler; fazla miktarda arjinin içermelerinden dolayı kuvvetli bazik karakterde basit proteinlerdir; tirozin, triptofan ve kükürtlü amino asit içermezler. Protaminler, proteinlerin en kısa zincirli olanlarıdır; molekül ağırlığı 1000-5000 arasındadır. Protaminler, dokularda asitlerle, özellikle nükleik asitlerle birleşmiş olarak bulunurlar; türlü protaminler, balık spermalarından büyük miktarlarda elde edilmişlerdir; uskumrudan skombrin, som balığından salmin, ringa balığından klüpein bunlara örnektir. Histonlar, protaminler gibi, fakat daha büyük molekülü, daha az bazik basit proteinlerdir. Uskumru balığındaki skombron, timus bezinden elde edilen timohiston önemli örneklerdir.

2) Fibriler proteinler: Molekülünün üç boyutlu şekli çok gerilmiş elipsoid biçiminde olan proteinlerdir. Skleroproteinler (Albüminoidler), suda, nötral tuz çözeltiğinde, seyreltik asit ve alkalilerde ve saf alkolde çözünmezler; pepsin ve tripsin gibi enzimlere dirençlidirler; hayvansal kaynaklıdır. Boynuz, kıl, yün, saç ve tırnaklarda bulunan keratin; bağ doku, kemik, kıkırdak ve tendonlarda bulunan, organizma proteinlerinin yarısından çoğunu oluşturan kollajen; ligament ve diğer destek dokularda bulunan elastin; ipek fibroini, önemli skleroproteinlerdir. Keratin, bir sistein dimeri olan sistin bakımından zengindir; normalde α -heliks yapısına sahiptir, fakat ıslak durumda β -kırmalı tabaka yapısını alır. Kollajen, glisin, prolin ve 4-hidroksiprolince zengindir; triptofan içermez. İpek fibroini, glisin, prolin, serin ve tirozince zengindir. Fibrinojen, kan plazması içinde çözülmüş olarak bulunur; kanın pıhtılaşması sırasında görev alır. Miyozin, kasta bulunur; kasın kasılmasında görev alır.

Bileşik proteinler (Konjuge proteinler)

Bileşik proteinler (konjuge proteinler), amino asitlerden oluşmuş polipeptit zincirlerinin prostetik grup denen yapılara bağlanmasıyla oluşmuş; hidroliz edildiklerinde amino asitlerden başka değişik nitelikte kimyasal maddeler de veren proteinlerdir.

1) Glikoproteinler: Prostetik grubu karbonhidrat olan bileşik proteinlerdir; %1-80 arasında değişen oranda karbonhidrat içerirler. Glikoproteinlerin genellikle karbonhidrat oranı %4'ten az olanlarına glikoprotein; %10-20 arasında değişenlere mukoprotein, karbonhidrat oranı proteinden fazla olanlarına mukoid adı verilir. Kan plazması proteinlerinden bazı taşıyıcı

proteinler ve immunoglobülinler; kemikteki osseomukoprotein, tendonlardaki tendomukoprotein, kıkırdaktaki kartilagomukoprotein, tükürükteki müsin, prostetik grupları karbonhidrat olan proteinler yani glikoproteinlerdir.

2) Proteoglikanlar: %80-95 gibi çok yüksek oranda karbonhidrat içeren bileşik proteinlerdir; polisakkaritler konusunda ayrıntılı olarak incelenmişlerdir.

3) Lipoproteinler: Proteinlerin lipidlerle oluşturdukları bileşik proteinlerdir; değişik oranlarda trigliserid, kolesterol ve fosfolipid içerirler. Lipoproteinler, önemli oranlarda lipid içermelerine karşın suda çözünürler; böylece kandaki lipidleri taşırlar. Lipoproteinin protein kısmına apolipoprotein veya apoprotein denir. Lipoproteinler, lipidler konusunda ayrıntılı olarak inceleneceklerdir. Proteinlerin fosfolipidlerle oluşturdukları suda çözünmeyen bileşik proteinler, proteolipidlerdir; özellikle miyelinin yapısında bulunurlar.

4) Fosfoproteinler: Prostetik grup olarak fosfat içeren bileşik proteinlerdir. Fosfoproteinlerde, proteinin yapısındaki serin, tirozin ve treonin gibi amino asidi kalıntılarının hidroksil grupları fosforik asitle esterleşmiştir. Sütte kazein; yumurtada vitellin, livetin ve fosvitin; balık yumurtasında ihtulin önemli fosfoprotein örnekleridirler.

5) Nükleoproteinler: Protaminler, histonlar ve diğer basit proteinlerin nükleik asitlerle bağlanması sonucu oluşmuş bileşik proteinlerdir. Nükleoprotaminler, en basit nükleoproteinlerdir; nükleik asit ile proteinler, arjinin-fosfat bağı ile bağlanmıştır; balık spermalarında boldurlar. Nükleohistonlarda da nükleik asit ile proteinler, arjinin-fosfat bağı ile bağlanmıştır; balık spermalarında ve kuş eritrositlerinde boldur. Yüksek nükleoproteinler, ribozomlar, kovalent bağlı RNA-protein, DNA-protein bileşiminde sitoplazma ve mitokondrilerde bulunurlar.

6) Metalloproteinler: Prostetik grup olarak Fe, Cu, Zn gibi ağır metalleri içeren bileşik proteinlerdir. Demirli metalloproteinlerden ferritin ve transferrin, bakırlı metalloproteinlerden seruloplazmin, önemli metalloprotein örnekleridirler.

7) Kromoproteinler: Metal-porfirin kompleks sistemleri ile oluşmuş bileşik proteinlerdir. Hemoglobin, miyoglobin, sitokromlar, peroksidaz, demir içeren önemli kromoprotein örnekleridirler.

Türev proteinler

Türev proteinler, ilk iki protein grubunda yer alan proteinlerin belirli etkilerle değişimleri sonucu oluşan proteinlerdir; birincil türev proteinler ve ikincil türev proteinler olmak üzere iki alt grupta incelenirler.

1) Birincil türev proteinler: Peptit bağlarına dokunmadan, asit, baz ve ısı gibi etkililerle protein moleküllerinin değişmesi sonucu oluşmuş türev proteinlerdir; denatüre tip proteinler olarak da adlandırılırlar. Suda çözünmeyen bu proteinlerin seyreltik asitler ve enzim etkisiyle oluşanlarına protean denir; asit ve alkalilerin sürekli etkisiyle oluşanlarına metaprotein denir; kaynatma, çalkalama, UV ışınları ve etanol etkisiyle oluşanlarına pıhtılaşmış protein veya koagule proteinler denir.

2) İkincil türev proteinler: Peptit bağlarını kısmen yıkan asit veya enzimlerin etkisiyle oluşan türev proteinlerdir. Peptit bağlarının bu şekilde parçalanmasıyla protein molekülleri, gitgide daha küçük parçalara bölünürler. Böyle bir parçalanmada büyük parçalara proteoz (albüminoz) denir; küçük parçalara pepton denir; daha küçük zincirler de polipeptitler ve peptitlerdir. İkincil türev proteinler, kaynatmakla çökelmez veya ısı ile pıhtılaşmazlar.

3.5. Proteinlerin Biyolojik Rollerine Göre Veya Fonksiyonel Olarak Sınıflandırılması

Katalitik proteinler(Enzimler): Biyokimyasal reaksiyonları katalize eden enzimler, yüksek derecede spesialize proteinlerdir. Amilaz, pepsin, lipaz önemli katalitik protein veya enzim örnekleridirler.

Taşıyıcı proteinler (transport proteinleri): Spesifik molekülleri veya iyonları bağlayıp bir organdan bir başka organa veya hücre membranının bir tarafından diğer tarafına transport eden proteinlerdir. Serum albümin, en iyi bilinen taşıyıcı proteindir; bilirubin, kalsiyum, yağ asitleri ve birçok ilaç serum albümine bağlanarak taşınır. Hemoglobin, oksijen taşıyan; lipoproteinler, lipid taşıyan; transferrin, demir taşıyan önemli taşıyıcı protein örnekleridirler. Bütün organizmaların plazma membranlarında ve intrasellüler membranlarında bulunan taşıyıcı proteinler, glukoz, amino asitler ve diğer maddeleri bağlarlar; bunları membranın bir tarafından diğer tarafına taşırlar.

Besleyici ve depo proteinler: Yumurta akının esas proteini ovalbümin, sütün esas proteini kazein besleyici proteinlerdir; bir çok bitki tohumu da çimlenen tohumun büyümesi için gerekli besleyici proteinleri depolamıştır. Ferritin, demir depolayan proteindir.

Kontraktil proteinler: Kasılabilen veya kendiliğinden hareket edebilen proteinlerdir. Miyozin ve aktin, iskelet kaslarının kontraktil sisteminde ve aynı zamanda bir çok kas olmayan hücrede işlev görür. Tubulin, mikrotubilleri oluşturan proteindir. Hücrelerde bulunan mikrotubuller, hücreleri yürütmek için kamçı ve kirpiklerdeki dynein proteini ile birlikte hareket eder.

Yapısal proteinler: Tendonların ve kıkırdağın esas yapısını, çok yüksek gerilme gücüne sahip kollajen oluşturmuştur; kösele, hemen hemen saf kollajendir. Ligamentler, iki

boyutta gerilme yeteneğinde bir yapısal protein olan elastin içerirler. Saç, tırnak ve tüyler, keratin içerirler. İpek liflerinin ve örümcek ağlarının esas komponenti fibroindir. Bazı böceklerin kanat eksenleri, resilinden yapılmıştır.

Savunma (defans) proteinleri: Organizmaları diğer türler tarafından istilaya karşı savunan, organizmayı hasardan koruyan proteinlerdir. İmmünoglobülinler, omurgalıların lenfositleri tarafından yapılan, spesialize (özüğüleşmiş) proteinlerdir; organizmayı istila eden bakterileri, virüsleri veya başka türe ait yabancı proteinleri (antijenler) tanıyabilirler ve presipite edebilirler (çöktürebilirler) veya nötralize edebilirler. Fibrinojen ve trombin gibi kan pıhtılaşma proteinleri, vasküler sistem yaralandığında yaralanan yerin kan pıhtısı ile kapatılarak kan kaybının önlenmesini sağlarlar. Yılan zehirleri, bakteriyel toksinler ve risin gibi toksik bitki proteinleri, aynı zamanda savunucu fonksiyonlara sahip gibi görünmektedirler.

Savunma proteinlerinin fibrinojen, trombin ve bazı zehirler dahil bazıları, aynı zamanda enzimdirler.

Düzenleyici proteinler: Sellüler düzenleme veya fizyolojik aktiviteye yardım eden proteinlerdir. İnsülin, büyüme hormonu gibi bazı hormonlar, düzenleyici proteinlerdir; insülin, şeker metabolizmasının düzenlenmesinde etkilidir; büyüme hormonu ise büyümenin düzenlenmesinde etkilidir. Bir çok hormonal sinyal için sellüler yanıt, sıklıkla G proteinler denen, GTP-bağlayan proteinler sınıfı aracı olur. Bazı düzenleyici proteinler, DNA'yı sarar; enzimlerin ve RNA moleküllerinin biyosentezini düzenlerler.

Diğer proteinler: Fonksiyonları henüz daha fazla bilinmeyen ve kolayca sınıflandırılmayan çok sayıda proteindir.

4. FİDAN GELİŞİM ALGORİTMASI

Gün geçtikçe zor mühendislik problemlerini çözmek için yeni yöntemler geliştirilmekte, ve geliştirilen yöntemlerde doğadan esinlenilmektedir. Özellikle optimizasyon problemlerinde Genetik algoritmalar, karınca koloni algoritmaları, yapay sinir ağları, parçacık sürü optimizasyonu ve yapay bağışıklık sistemleri gibi teknikler yoğun olarak kullanılmaktadır. Bu tezde da fidanların gelişiminden esinlenerek geliştirilen hesapsal bir yöntem olan Fidan Gelişimi Algoritması kullanılmıştır. Fidan Gelişimi Algoritması hem yerellik ve hem de küresellik özelliğine sahip yeni bir hesapsal yöntemdir. Bu algoritmanın ilk yapısını Karci ve diğerleri aşağıdaki gibi etraflıca açıklamışlardır[11].

Fidan Gelişimi algoritmasına göre bir ormanda bir fidandan yalnızca bir tane bulunabilmektedir. Ayrıca Fidanlar yetiştirilecekleri ortama düzgün dikilmelidirler. Bu Fidanların çözüm uzayına düzgün dağılması olarak ta ifade edilebilir. Yöntemde diğer bir operatör olarak Fidanların büyürken dalların birbirine çok yakın olmaması olayı ele alınmıştır. Aşılama Yöntemiyle bir fidanın başka bir fidana dönüşmesi sağlanır. Çiftleşme ile yeni fidanların üremesi sağlanır. Çiftleşme işlemi çiftleşme operatörüyle temsil edilir[11].

Geliştirilmiş olan yöntemdeki operatörler, fidan gelişiminin mümkün olduğunca en iyi bir şekilde taklit edilmesi amacıyla fidanların gelişim esnasında nasıl bir değişikliğe uğradıkları operatörler olarak ele alınmıştır. Bundan dolayı, fidan ekimi bir operatör, eşleştirme bir operatör, dallanma bir operatör ve aşılama da bir operatör olarak ele alınmıştır. Fidan ekimi operatörü arama uzayına düzgün dağılmış fidanların (çözüm) oluşturulmasını sağlamaktadır. Dallanma operatörünün yerel arama, eşleştirme operatörünün ise küresel aramayı sağlaması beklenmektedir. Aşılama operatörü ise benzer fidanlar arasında bilgi değişimini sağlamaktadır [11].

Böylece geliştirilmiş yöntemde, aday çözümlerin başlangıçta arama uzayında düzgün dağıtılması sağlandıktan sonra hem yerel ve küresel arama gerçekleştirilmekte ve benzer çözümler arasında bilgi değişimi imkânı olmaktadır.

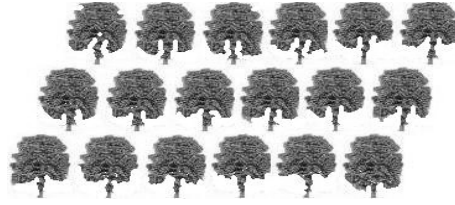
Kısaca, günümüzde kullanılan yumuşak hesaplama tekniklerine alternatif bir yöntem geliştirilmiştir. Tavlama benzetimi, tabu arama ve tepe tırmanma algoritmalarında olduğu gibi geçici çözümün komşularını incelemeyi içerdiği gibi birden fazla geçici çözüm ile başlamayı ve bu çözümleri çözüm uzayına düzgün olarak dağıtmaya çalışmayı ve işlemler devam ederken küresel değişimleri de içermektedir. Tüm bunlardan başka, genetik algoritmalarındaki gibi bir uygunluk fonksiyonu belirleme zorunluluğu yoktur; direkt olarak amaç fonksiyon değeri kullanılır. Bütün bu özelliklerinden dolayı daha güçlü bir hesapsal yöntem olması beklenmektedir.

4.1. Fidanların ekimi

Düzenli Populasyon [12]: Çözüm uzayı Z olarak kabul edilsin ve bu uzay s elemanını içermiş olsun. Z uzayı $Z=\{z_0, z_1, \dots, z_{s-1}\}$ şeklinde yazılabilir. İlk olarak çözüm uzayının ikili tabanda olduğu kabul edilsin ve çözüm bir fidan veya fidan kümesi olacaktır. Eğer her fidanın iki dalı varsa,

$$Z=\{00,01,10,11\}$$

şeklinde olur. Eğer her fidanda n tane dal varsa, çözüm uzayının boyutu 2^n olur ve s elemanı 2^n tane farklı durumu olur. Z uzayının bazı $\{01,10\}$ olur ve diğer durumların bunların lineer kombinasyonundan elde edilebilir[10].



Şekil 4.1. Fidanların bahçeye düzenli dağıtılması.

Çözüm uzayının boyutu problemlere göre farklı olabilir ve buna göre başlangıç çözümleri çözüm uzayının boyutuna göre üretilir. Birçok sayısal yöntemde başlangıç çözümleri rasgele olarak üretilirler, fakat bu çalışmada başlangıç çözümleri rasgele olarak üretilmemektedir ve Şekil 1’ de görüldüğü gibi başlangıç çözümleri çözüm uzayına düzgün bir şekilde dağıtılır.

Bir fidanın dallarının alacağı değerlerin alt ve üst sınırları vardır. D_i bir dal olmak üzere değer alabileceği aralık $[l_i, u_i]$ olsun. Bu durumda S_1 ve S_2 birer fidan olmak üzere $S_1=\{u_1, u_2, \dots, u_n\}$, $S_2=\{l_1, l_2, \dots, l_n\}$ şeklinde üretilir (Şekil 2), n ise bir fidanda olabilecek maksimum dal sayısıdır. Bir populasyon faktörü tanımlanır ve bu faktör k olmak üzere $k=2$ için

$$S_3=\{l_1+(u_1-l_1)*r, l_2+(u_2-l_2)*r, \dots, l_n/2+(u_n/2-l_n/2)*r, l_n/2+1+(u_n/2+1-l_n/2+1)*(1-r), l_n/2+2+(u_n/2+2-l_n/2+2)*(1-r)\} \text{ ve },$$

$$S_4=\{l_1+(u_1-l_1)*(1-r), l_2+(u_2-l_2)*(1-r), \dots, l_n/2+(u_n/2-l_n/2)*(1-r), l_n/2+1+(u_n/2+1-l_n/2+1)*r, l_n/2+2+(u_n/2+2-l_n/2+2)*r\}$$

şeklinde iki tane daha fidan elde edilir ve $0 \leq r < 1$

Üretilen populasyon P olmak üzere ve başlangıçta $|P|$ tane fidan içermek üzere, S_1 , S_2 , S_3 , S_4 elemanlarının eklendiği küme R olsun. Eğer $|R| < |P|$ ise, $k=3$ yapılı ve bu durumda $2^3=8-2=6$ fidan üretilmelidir.

$$S_5=\{l_1+(u_1-l_1)*r, \dots, l_2n/3+(u_2n/3-l_2n/3)*r, l_2n/3+1+(u_2n/3+1-l_2n/3+1)*(1-r), \dots, l_n+(u_n-l_n)*(1-r)\}$$

$S6=\{l_1+(u_1-l_1)*r, \dots, l_n/3+(u_n/3-l_n/3)*r, l_n/3+1+(u_n/3+1-l_n/3+1)*(1-r), \dots, l_{2n}/3+(u_{2n}/3-l_{2n}/3)*(1-r), l_{2n}/3+1+(u_{2n}/3+1-l_{2n}/3+1)*r, \dots, l_n+(u_n-l_n)*r\}$
 $S7=\{l_1+(u_1-l_1)*r, \dots, l_n/3+(u_n/3-l_n/3)*r, l_n/3+1+(u_n/3+1-l_n/3+1)*(1-r), \dots, l_n+(u_n-l_n)*(1-r)\}$
 $S8=\{l_1+(u_1-l_1)*(1-r), \dots, l_n/3+(u_n/3-l_n/3)*(1-r), l_n/3+1+(u_n/3+1-l_n/3+1)*r, \dots, l_n+(u_n-l_n)*r\}$
 $S9=\{l_1+(u_1-l_1)*(1-r), \dots, l_n/3+(u_n/3-l_n/3)*(1-r), l_n/3+1+(u_n/3+1-l_n/3+1)*r, \dots, l_{2n}/3+(u_{2n}/3-l_{2n}/3)*r, l_{2n}/3+1+(u_{2n}/3+1-l_{2n}/3+1)*(1-r), \dots, l_n+(u_n-l_n)*(1-r)\}$

şeklinde 6 tane fidan üretilir ve R kümesine eklenir. Eğer $|R|$ hâlâ $|P|$ 'den küçük ise, k değeri 1 artırılır. Bu durumda, 14 tane fidan üretilir ve R kümesine eklenir. Fidan üretim algoritması aşağıdaki gibidir.

Algoritma 1. Başlangıç Populasyon Üretilmesi

// P başlangıç popülasyonu, I endeks kümesi ve I_e genişletilmiş endeks kümesi.

İki tane fidan üretilir: $P[0]$ dalların üst sınırlarını ve $P[1]$ dalların alt sınırlarını içersin.

Endeks $\leftarrow 2$

$k \leftarrow 2$

P dolmadığı sürece devam et

i_e elemanı I_e kümesinin elemanı olsun ve i_e elemanı bitleri genişletilmiştir (bitler parçaları ilgilendirir).

$i \leftarrow 1$

P dolmadığı sürece ve belli k değeri için bütün fidanlar üretilmediği sürece devam et (ve $i \leq 2^k - 2$)

i , k -bit numarası ve i_e ise i 'nin genişletilmiş şekli.

Bütün i bitleri genişletilir ve her biri $P[0]$ ve $P[1]$ uzunluğu kadar genişletilir.

$j \leftarrow 1, \dots, n$

Eğer i_e 'nin j . biti 1 ise, $P[\text{Endeks}]$ 'in j . dalı $P[0]*r$ eşit olur.

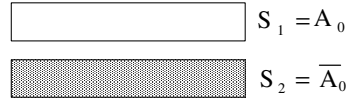
Değilse $P[\text{Endeks}]$ 'in j . dalı $P[1]*r$ eşit olur.

(r bir rasgele sayı olup $[0,1]$ aralığında değer alır ve gerçel bir sayıdır)

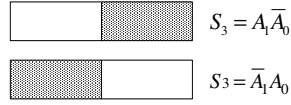
Endeks $\leftarrow$ Endeks + 1

$i \leftarrow i + 1$

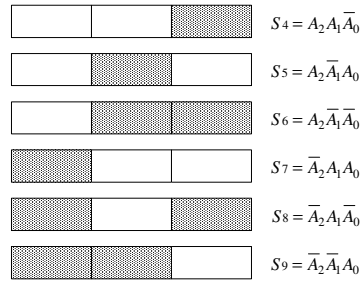
$k \leftarrow k + 1$



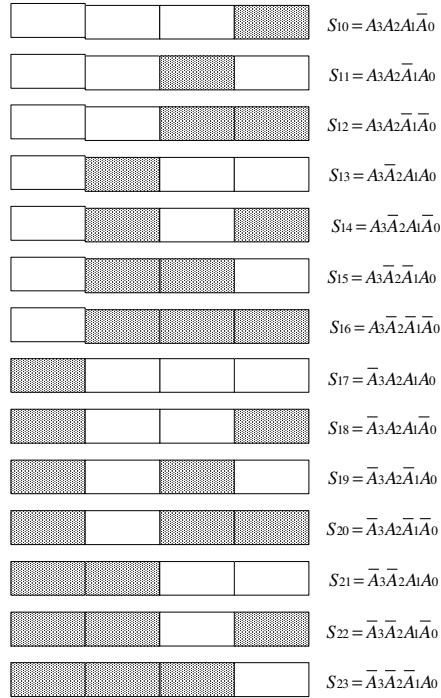
Şekil 4.2. Başlangıç fidanları



Şekil 4.3. $k=2$ için yeni fidanlar



Şekil 4.4 $k=3$ için yeni fidanlar



Şekil 4.5. $k=4$ için yeni fidanlar

4.2. Fidanların Büyümesi

Fidanların Büyümesi iki operatörle meydana gelir. Bu operatörle eşleştirme ve Karıştır-Geliştir(Mix and Improve) operatörleridir.

a) Eşleştirme (Küresel arama operatörü): Eşleştirme operatörünün amacı ($\otimes$ ile gösterilecek), fidanlar arasında genetik bilginin değişimini sağlamaktır. Bu çalışmada bir eşleştirme faktörü vardır ve bu faktör fidanlar arasındaki uzaklık ile ilgili olarak tanımlanan bir olasılıktır.

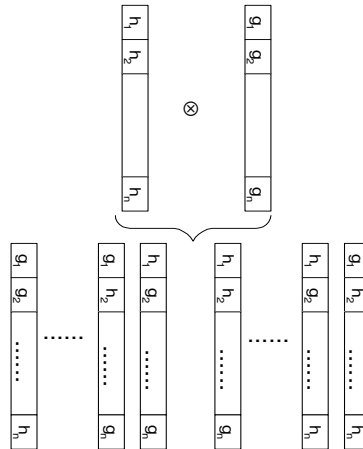
$G=g_1g_2...g_i...g_n$ ve $H=h_1h_2...h_i...h_n$ iki tane fidan olsun. $P(G,H)$, G ve H fidanlarının eşleşme olasılığı olsun ve $P_m(G,H)$ ise, eşleşme olasılığı olsun. Bu durumda bu olasılıklar denklem 4.1 deki gibi hesaplanır.

$$P(G,H)=\frac{\left(\sum_{i=1}^n(g_i-h_i)^2\right)^{1/2}}{R} \text{ ve } R=\left(\sum_{i=1}^n(u_i-l_i)^2\right)^{1/2} \quad (4.1)$$

$$P_m(G,H)=1-\frac{\left(\sum_{i=1}^n(g_i-h_i)^2\right)^{1/2}}{R} \quad (4.2)$$

Burada u_i üst sınır ve l_i alt sınırdır. Çünkü rüzgâr ya da doğadaki bazı etkiler eşleştirme olasılığını etkilemektedir.

$G=g_1g_2...g_i...g_n$ ve $H=h_1h_2...h_i...h_n$ fidan olmak üzere, $G\otimes H$ sonucunda $2n$ tane yeni fidan oluşturulur, her fidan ya kendi dalını eşleştiği fidana verir ya da eşleştiği fidanın dalını alır. Şekil 4.6, bir dal kopyalama durumunu göstermektedir. Eşleştirme operatörü eğer $P_m(G,H)$ eşleştirme şartını sağlıyorsa her çifti (g_i ve h_i) için uygulanır.



Şekil 4.6. G ve H eşleşmesi: Birer dal değişimi

Algoritma 2. Eşleştirme(G,H) //tek dal
değişim durumu
 $j \leftarrow 1, \dots, n$

$$P_m(G,H) = 1 - \frac{\left(\sum_{j=1}^n (g_j - h_j)^2 \right)^{1/2}}{R} \text{ hesapla}$$

 $i \leftarrow 1, \dots, n$
eğer $P_m(G,H) \geq \text{rasgele}[0,1]$ ise
 $G \leftarrow G - g_i$, ve $H \leftarrow H - h_i$
 $G \leftarrow G + h_i$, ve $H \leftarrow H + g_i$
// $G \leftarrow G + h_i$, ve h_i dalı g_i dalının yerine
eklenir, ve $H \leftarrow H + g_i$ ve g_i dalı h_i dalının yerine
eklenir,

b) Karıştır-Geliştir: Genel olarak iki versiyonu kullanılmaktadır. Bu Operatörün amacı Eşleşen 2 fidanı iyileştirmektir. Bu operatör S_1 ve S_2 fidanları arasındaki öklit uzaklığı tabanlıdır. S_{ij} S_i fidanının j. dalını ifade etsin. S_1 ve S_2 arasındaki öklit uzaklığı denklem 4.3 teki gibi ifade edilir.

$$R_1 = \sqrt{\sum_{i=1}^n (S_{1,i} - S_{2,i})^2} \quad (4.3)$$

Herhangi bir populasyon için $R_{\max}$ maksimum uzaklık aşağıdaki gibi hesaplanır. $S_{\min}$ değerinin bütün değişkenler için alt sınırı içerdiği farz edilirse. $S_{\max}$ değeri ise tüm değişkenler için üst sınırı ifade eder.

$$R_{\max} = \sqrt{\sum_{i=1}^n (S_{\max,i} - S_{\min,i})^2} \quad (4.4)$$

Versiyon1: O_{ort} bahçenin ortalama amaç değeridir. O_i ise i. fidanın amaç değeridir. Ortalama amaç değeri denklem 4.5 deki gibi elde edilir.

$$O_{\text{ort}} = \frac{\sum_{i=1}^{|G|} O_i}{|G|} \quad (4.5)$$

S_{yeni} amaç değeri hesaba katılarak yeni fidanların elde edilmesi denklem 4.6 ve denklem 4.7 daki gibi hesaplanabilir.

$$S_{yeni,i} = \frac{O_1}{O_{ort}} \frac{R_1}{R} S_{1,i} + \frac{O_2}{O_{ort}} \left(1 - \frac{R_1}{R}\right) S_{2,i} \quad (4.6)$$

$$S_{yeni,i} = \frac{O_1}{O_{ort}} \left(1 - \frac{R_1}{R}\right) S_{1,i} + \frac{O_2}{O_{ort}} \frac{R_1}{R} S_{2,i} \quad (4.7)$$

Versiyon2: S_1 ve S_2 fidanlarına Karıştır-Geliştir operatörünün uygulanmasıyla yeni fidan (S_{yeni}) elde edilir. Karıştır-Geliştir operatörünün 2.versiyonunda yeni fidanın herhangi bir dalı amaç değeri kullanılmaksızın denklem 4.8 ve denklem 4.9 daki gibi elde edilir.

$$S_{yeni,i} = \frac{R_1}{R} S_{1,i} + \left(1 - \frac{R_1}{R}\right) S_{2,i} \quad (4.8)$$

$$S_{yeni,i} = \left(1 - \frac{R_1}{R}\right) S_{1,i} + \frac{R_1}{R} S_{2,i} \quad (4.9)$$

Karıştır-Geliştir operatörü kullanılırken hem versiyon1 hemde versiyon2 denklemleri her bir iterasyonda iki çocuk üretmek için kullanılabilir. Bu durumda iterasyon sayısı $|G|/2$ üst sınırı ile sınırlandırılır.

Fidan Gelişimi Algoritması bütün bu tanımlamalar sonucunda genel amaçlı bir hesapsal yöntem olarak ortaya konulmuş olup, Geliştirilmiş olan bu yöntem optimizasyon problemleri, çözüm arama problemleri ve çok amaçlı çözüm arama problemleri kapsamaktadır.

5. PROTEİN YAPI TAHMİNİ İÇİN UYGULANAN DİĞER YÖNTEMLER

Geçmişten günümüze proteinlerin yapı tahmini ile ilgili olarak bir çok çalışma yapılmıştır. Bu çalışmaların bir çoğunda katlanmanın şekli esas alınmıştır.[10]

Bu yöntemleri başarı oranlarını esas alarak sıralayacak olursak:

1-Biodirectional Dynamic for protein Secondary Structure Prediction[12] . Bu çalışmada helix ve coil yapılar üzerinde çalışılmıştır.Yaklaşık %75 başarı elde edilmiştir.

2-Use of Multiple Alignment in Protein Secondary Structure Prediction [13]. Bu çalışmada da Helix, Coil, ve Extended Standart yapılar üzerinde çalışılmıştır. Brook haven databank'tan alınan 20 protein örneği kullanılmıştır.Yaklaşık %69 başarı elde edilmiştir.[10]

3-Two-Stage Multi Class Support Vector Machines To Protein Secondary Structure Prediction[14]. Bu çalışmada PSI-BLAST yöntemi ile RS126 ve CB396 örnekleri üzerinde Helix, Coil, ve Extended Strand yapıları üzerinde çalışılmıştır. Yaklaşık %77 başarı elde edilmiştir.[10]

4-HYPROSP:A hybrid Protein Secondary Structure Prediction algoritma-a knowledge-based approach [15]. Bu çalışmada PSI-BLAST yöntemi ile Helix, Coil, ve Extended Strand ve diğer yapılar üzerinde çalışılmıştır.Yaklaşık %80 başarı elde edilmiştir.[10]

5-Context Sensitive Vocabulary And its Application in Protein Secondary Structure Prediction [16]. Bu çalışmada BLAST yöntemi ile RS126 ve CB513 protein örnekleri üzerinde çalışılmıştır.Yaklaşık %70 başarı elde edilmiştir.[10]

6- Prediction of Protein Secondary Structure Using Genetic Programming[17]. Bu çalışmada Helix, Coil, ve Extended Standart yapılar üzerinde çalışılmıştır. Yaklaşık %76 başarı elde edilmiştir[10]

7-Parallel Stochastic Search for Protein Secondary Structure Prediction[18]. Bu çalışmada EDA algoritması yardımıyla CB13 data seti üzerinde çalışılmış ve ikincil yapının olasılık tahminleri üzerinde durulmuştur. Yaklaşık %70 oranında başarı elde edilmiştir.

8- Protein Secondary Structure Prediction Using Decision Lists[19] Bu çalışmada Greedy decision list learner algoritması CB513 data seti üzerinde uygulanmıştır. Kural tabanları oluşturularak işlem yapılmıştır.Yaklaşık %69 başarı elde edilmiştir.

9- Protein Secondary Structure Prediction with a neural network [20]. Bu çalışmada protein yapı tahmini için bir neural network yapısı mevcuttur. Bu yöntem ile yaklaşık % 79 başarı elde edilmiştir.

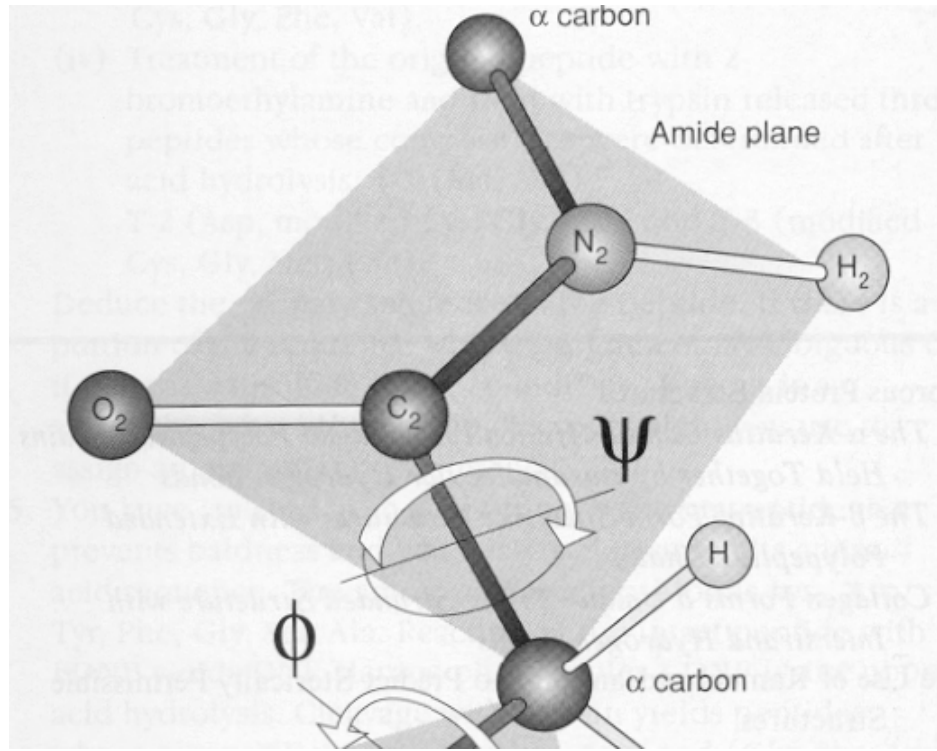
10-Secondary Structure prediction Of Hemoglobin By using Combined Neural Networks. Bu çalışmada hemoglobin proteininin yapısını tahmin için birleştirilmiş yapay sinir ağları metodu kullanılmıştır.[10]

11-Yumuşak hesaplama teknikleri ile protein yapı tahmini [10]. Bu çalışmada protein yapı tahmini için yapay sinir ağları kullanılmıştır. Bu çalışmada çalışmada bizimde üzerinde çalıştığımız 1a9a,18be,1bbe kollejen proteinleri kullanılmış ve yaklaşık %70 başarı elde edilmiştir.

6.GELİŞTİRİLEN YÖNTEM VE SONUÇLARI

Yapılan çalışmalarda proteinlerin ikincil yapılarının tahmini için peptitlerin merkezci alfa karbonları arasında oluşan bağ açılarından faydalanılmıştır. Tezde değişik proteinlerden alınan açı örnekleri Fidan gelişimi Algoritması kullanılarak eğitime tabi tutulmuştur. Açılar üzerinde fidan Gelişimi algoritmasının 4.bölümde anlatılmış olan her iki versiyonu üzerinde de eğitim ve test işlemi yapılmış performans ölçümleri elde edilerek ve sonuçlar değerlendirilmiştir.

1bbe, 1a89 ve 1a9a kollajen proteinlerine ait açıları elde etmek için Swiss PDB Viewer programı kullanılmıştır. Program kullanılarak PDB uzantılı protein örnekleri elde edilmiştir. Bu veriler eğitim ve test verisi olarak kullanılmıştır[10].



Şekil 6.1. Merkezci karbon atomları arasındaki bağ açıları

Bir protein için peptitlerdeki merkezci karbon atomları arasındaki bağ açıları o protein için ikincil yapısının belirleyiciliği açısından önemlidir. Şekil 6.1 de görülen merkezci karbon atomları arasındaki bağ açıları değerlendirilirse bir proteinin tüm sıralı merkezci karbonlarının oluşturmuş oldukları bağ açıları tahmin edilebilir. Bu sayede proteinin ikincil yapısı ortaya konulur.

Fidan gelişimi Algoritmasında eğitim için elimizde bulunan 1bbe, 1a89 ve 1a9a kollajen proteinleri'nin aşağıdaki verileri kullanılmıştır. Test için kullanılan proteinlerin içermiş olduğu 11 değişik protein mevcuttur. Bu nedenle tüm amino asitleri kullanmaya ihtiyaç duyulmamıştır. Sadece lazım olan 11 farklı amino asit için bir değişim tablosu hazırlanmıştır.

Tablo 6.1. Amino asitlerin nümerik değerlere dönüşüm tablosu

Amino Asidin Kısa İsmi	Nümerik Kodu
Ala	1
Arg	2
Asn	3
Gln	4
Glu	5
Gly	6
Leu	7
Lys	8
Pro	9
Thr	10
Val	11

Protein data banklardan alınan farklı açi örneklerimiz 109.10 ile 135.10 arasındadır. Bu aralığı 11 farklı amino asit sayısına bölüp taban fonksiyon aldığımızda 3 derece aralıklarla açi sınıflarına bölünerek 9 farklı açi sınıfı elde edilir. Elde edilen açi sınıflarını Tablo 6.2 deki gibi yazabiliriz.

Tablo 6.2. Açı sınıfları

Açı aralıkları	Açı sınıfları
109-112	A1
112-115	A2
115-118	A3
118-121	A4
121-124	A5
124-127	A6
127-130	A7
130-133	A8
133-136	A9

Fidan Gelişimi Algoritmasında kullanılan operatörler yapılan çalışmada da problem çözümünde sırayla uygulanmıştır. Bölüm 4'te de anlatıldığı gibi Fidan Gelişimi Algoritması'nın Karıştır-Geliştir operatörü iki versiyondan oluşmaktadır. Bu versiyonlardan her ikisinde tez çalışmasında probleme uygulanmış ve farklı sonuçlar elde edilmiştir.

1a9a, 1a89 ve 1bbe proteini için başlangıç popülasyonu açılış değerleri sırasıyla aşağıdaki tablolarda gösterildiği gibidir.

Tablo 6.3. 1a9a proteini için başlangıç popülasyonu açılış değerleri

A1	116.34	118	118.68	118.24	119.72	117.9	116.86	117.66	116.28	117.49
A2	118.2	117.06	119.15	117.45	116.81	117.59	119.57	117.55	117.41	117.8
A3	116.34	118	118.68	118.24	119.72	117.9	116.86	117.66	116.28	117.49

Tablo 6.4. 1a89 proteini için başlangıç popülasyonu açılış değerleri

A1	119.55	116.29	118.05	117.26	116.71	117.8	118.29	115.3	118.95	119.07
A2	117.8	117	120.24	118.58	116.92	121.82	121.25	117.19	117.77	119.02
A3	118.2	118.31	116.74	119.36	119.53	118.17	119.6	117.98	118.07	119.03

Tablo 6.5. 1bbe proteini için başlangıç popülasyonu açılış değerleri

A1	121.53	119.85	120.55	121.53	119.88	120.54	121.54	119.85	120.55	121.53
A2	121.53	119.86	120.53	121.51	119.85	120.52	121.53	119.86	120.52	121.53

Başlangıçta kullanılan bu açılış verileri eğitim işlemi için kullanılmıştır. Öncelikle Başlangıç popülasyonu eşleştirme işlemine tabi tutulmuştur. Eşleştirme işlemi için her iterasyonda rasgele olarak 1 – 10 arasında farklı bir çaprazlama noktası seçilmiş ve bu işlemle genetik bilginin fidanlar arasında değişimi hedeflenmiştir.

Eşleştirme işlemi sonucunda tüm fidanlar birbiriyle rasgele noktalardan çaprazlanarak 6 yeni fidan elde edilmiştir. 200 iterasyon sonucunda elde edilen yeni fidanlar 1a9a, 1a89 ve 1bbe proteinleri için aşağıdaki tablolarda sırasıyla tablo 6.6 , tablo 6.7 ve tablo 6.8 de verilmiştir.

Tablo 6.6. 1a9a proteinin eşleştirmesi sonucunda oluşan yeni fidanlar

Yeni fidan1	116.34	118	118.68	118.24	119.72	117.59	119.57	117.55	117.41	117.8
Yeni fidan2	118.2	117.06	119.15	117.45	116.81	117.9	116.86	117.66	116.28	117.49
Yeni fidan3	116.34	118	118.68	118.24	119.72	117.9	116.86	117.66	116.28	117.49
Yeni fidan4	116.34	118	118.68	118.24	119.72	117.9	116.86	117.66	116.28	117.49
Yeni fidan5	118.2	117.06	119.15	117.45	116.81	117.9	116.86	117.66	116.28	117.49
Yeni fidan6	116.34	118	118.68	118.24	119.72	117.59	119.57	117.55	117.41	117.8

Tablo 6.7. 1a89 proteinin eşleştirmesi Sonucunda oluşan yeni fidanlar

Yeni fidan1	117.21	115.06	120.24	118.58	116.92	121.82	121.25	117.19	117.77	119.02
Yeni fidan2	117.8	117	120.24	119.36	119.53	118.17	116.02	119.69	117.75	119.02
Yeni fidan3	117.21	115.06	120.24	119.36	119.53	118.17	119.6	117.98	118.07	119.03
Yeni fidan4	118.2	118.31	116.74	119.36	119.53	118.17	116.02	119.69	117.75	119.02
Yeni fidan5	117.8	117	120.24	119.36	119.53	118.17	119.6	117.98	118.07	119.03
Yeni fidan6	118.2	118.31	116.74	118.58	116.92	121.82	121.25	117.19	117.77	119.02

Tablo 6.8. 1bbe proteinin eşleştirmesi Sonucunda oluşan yeni fidanlar

Yeni fidan1	121.53	119.86	120.53	121.51	119.85	120.52	121.53	119.86	120.52	121.53
Yeni fidan2	121.53	119.86	120.53	121.51	119.85	120.52	121.53	119.86	120.52	121.53
Yeni fidan3	121.53	119.86	120.53	121.51	119.85	120.52	121.53	119.86	120.52	121.53
Yeni fidan4	121.53	119.86	120.53	121.51	119.85	120.52	121.53	119.86	120.52	121.53
Yeni fidan5	117.8	117	120.24	118.58	116.92	118.17	119.6	117.98	118.07	119.03
Yeni fidan6	118.2	118.31	116.74	119.36	119.53	121.82	121.25	117.19	117.77	119.02

Eşleştirme işlemi sonucunda probleme uygulanan ikinci operatör Karıştır-Geliştir operatörüdür. Bu operatörde eşleşme operatörüyle elde edilen yeni fidanlardan amaç değeri en küçük olan 2 fidanın her bir dalının mutlak uzaklığı alınarak kullanılarak R1 değeri hesaplanır. S_1 ve S_2 fidanları eşleştirme işlemi sonrasında elde edilen en iyi iki olmak üzere $S_1[Di]$ S_1 fidanın i.dalını ifade eder.

Amaç değeri, kullanılan Algoritmada her iterasyonda seçim kriteri olarak kullanılır. Operatörün 1. ve 2. versiyonlarında kullanılan S_1 ve S_2 fidanları seçilirken amaç değeri en küçük olan fidanlar seçilir. Bu Çalışmada amaç değerini belirlemek için aşağıdaki fonksiyon denklem 6.1 kullanılmıştır.

$$\sum_{i=0}^n w_i(A[i] - ort(A[i])) \quad (6.1)$$

$w(i)$ değerleri 0-1 arasında olup her iterasyonda rasgele olarak hesaplanmıştır. S_1 ve S_2 fidanları arasındaki öklit uzaklığı ise denklem 6.2 deki gibi belirlenir.

$$R1 = \sqrt{\sum_{i=1}^n (S_1[Di] - S_2[Di])^2} \quad (6.2)$$

$R_{\max}$, değeri ise eğitim için kullanılan her bir açı değerinin maksimum ve minimum değerlerinden oluşan iki fidanın dallarının 1 kereye mahsus olarak öklit uzaklığı denklem 6.3 teki gibi hesaplanarak edilir.

$$R_{\max} = \sqrt{\sum_{i=1}^n (S_{\max,i} - S_{\min,i})^2} \quad (6.3)$$

Bir sonraki nesile aktarılabacak olan en iyi fidan seçilirken aşağıdaki formül kullanılarak en iyi fidanın seçilme işlemi yapılır.

Bir sonraki nesile aktarılabacak yeni fidanlar için 2 farklı versiyon kullanılmıştır. Kullanılan ilk versiyonda denklem 6.4 deki fonksiyon kullanılmıştır. Burada O_{ort} bahçenin ortalama amaç değeridir. O_i ise i.fidanın amaç değeridir.

$$S_{\text{yeni},i} = \frac{O_1}{O_{\text{ort}}} \frac{R_1}{R} S_{1,i} + \frac{O_2}{O_{\text{ort}}} \left(1 - \frac{R_1}{R}\right) S_{2,i} \quad (6.4)$$

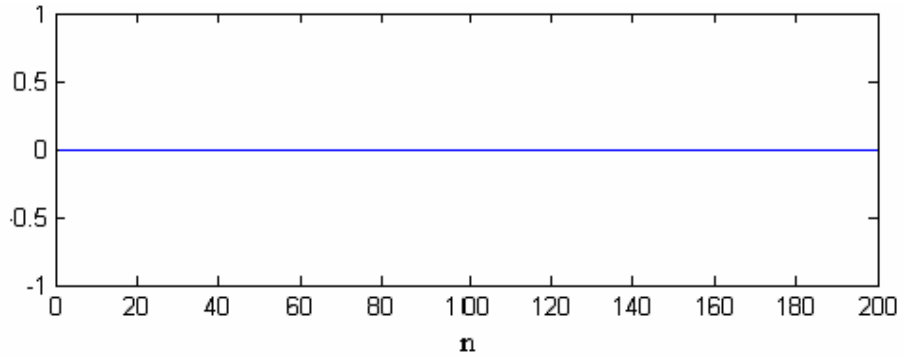
Burada O_{ort} bahçenin ortalama amaç değeridir. O_i ise i.fidanın amaç değeridir. Ortalama amaç değeri denklem 6.5 deki gibi elde edilir.

$$O_{\text{ort}} = \frac{\sum_{i=1}^{|G|} O_i}{|G|} \quad (6.5)$$

Bu işlem sonucunda ilk versiyon için elde edilen en iyi fidan belirlenmiş ve eğitim işlemi tamamlanmıştır. Bundan sonra proteine ait elimizde bulunan başka bir bağ açı değerleri kullanılarak test işlemine tabi tutulacaktır. Bu işlemi yaparken test için kullanılan fidanın açı sınıflarıyla elde edilen fidanın açı sınıfları karşılaştırılır ve baz olarak fidanlar arasındaki açı uyuşma sayısı kullanılır. Bağlı hata elde edilen en iyi fidanın her iterasyonda hata farkını vermektedir.

$$\text{Bağlı Hata Değeri}(\epsilon) = \sum_{i=1}^n |\text{Test açı değeri} - \text{En iyi fidanın açı değeri}| \quad (6.6)$$

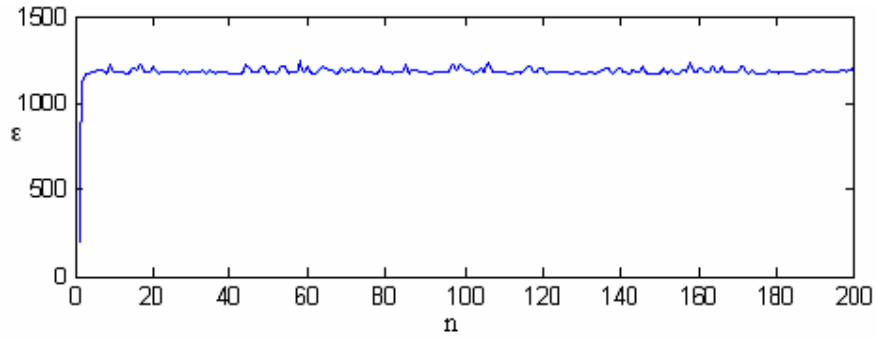
Versiyon1 için fidanlar arasındaki uyuşma sayısı ve bağlı hatanın iterasyon sayısına göre değişimi şekil 6.2 deki değişmiştir.



Şekil 6.2. Versiyon 1 için 1a9a proteinine ait açı sınıfı uyuşma sayısı

Tablo 6.9. Versiyon1, 1a9a proteini için test açı sınıfı ve en iyi fidan için açı sınıfı değerleri

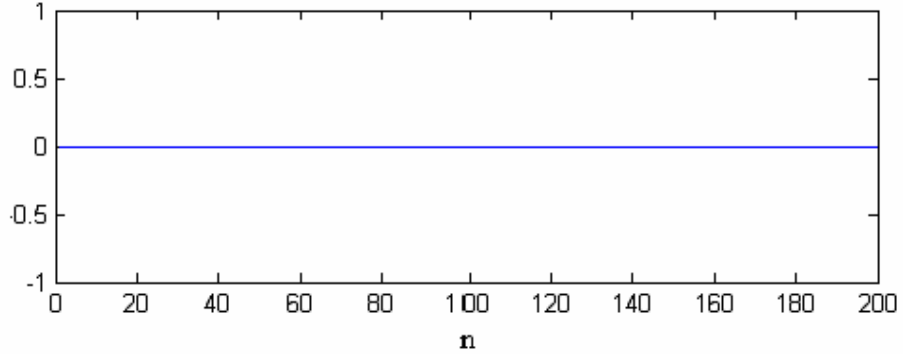
1a9a(1)	117.75	118.68	120.37	117.81	118.43	115.86	117.8	119.04	118.12	118.13
Açı sınıfı	A3	A4	A4	A3	A4	A3	A3	A4	A4	A4
En iyi fidan(1)	-	-	-	-	-	-	-	-	-	-
	20.115	22.599	21.629	22.527	24.448	22.091	24.265	21.699	22.637	22.072
Açı sınıfı	-	-	-	-	-	-	-	-	-	-



Şekil 6.3. Versiyon1, 1a9a proteinine ait bağıl hata değişimi

Tablo 6.10. Versiyon1, 1a9a proteini için test açı değerleri ve en iyi fidana ait açı değerleri

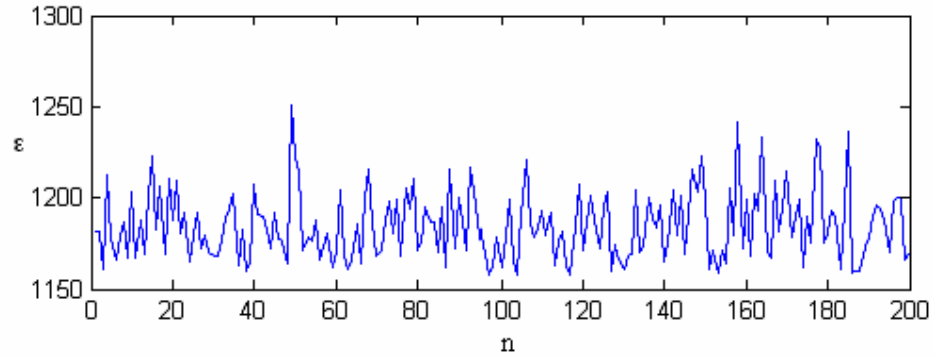
1a9a(1)										
Açı Değeri	117.75	118.68	120.37	117.81	118.43	115.86	117.8	119.04	118.12	118.13
En iyi fidan(1)	-	-	-	-	-	-	-	-	-	-
Açı Değeri	20.115	22.599	21.629	22.527	24.448	22.091	24.265	21.699	22.637	22.072



Şekil 6.4. Versiyon1, 1a89 proteinine ait aç sınıfı uyuşma sayısı

Tablo 6.11. Versiyon1, 1a89 proteini için test aç sınıfı ve en iyi fidan için aç sınıfı değerleri

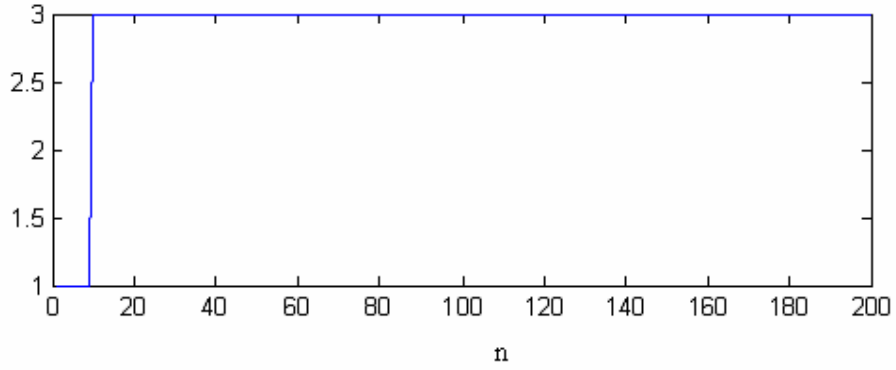
1a89(1)	117.96	120.76	118.11	118.01	118.58	116.67	117.23	118.73	117.7	117.65
Aç sınıfı	A3	A4	A4	A4	A4	A3	A3	A4	A3	A3
En iyi fidan(1)	11.433	13.313	0.801	12.852	14.825	0.799	10.279	12.724	12.209	11.996
Aç sınıfı	-	-	-	-	-	-	-	-	-	-



Şekil 6.5. Versiyon1 için 1a89 proteinine ait bağl hata değişimi

Tablo 6.12. Versiyon1, 1a89 proteini için test aç ı değ erleri ve en iyi fidana ait aç ı değ erleri

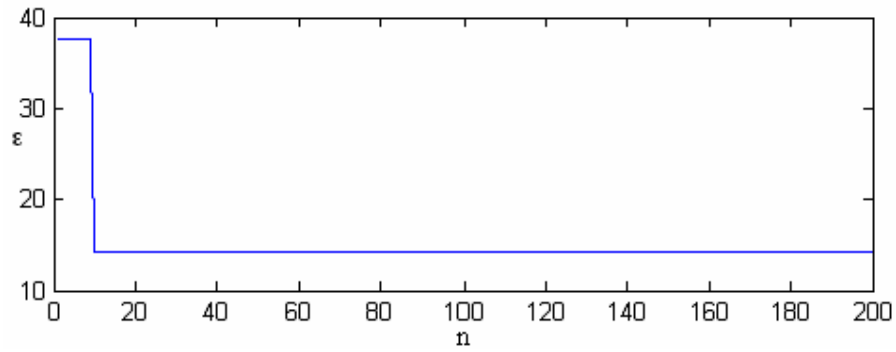
1a89(1)										
Aç ı	117.96	120.76	118.11	118.01	118.58	116.67	117.23	118.73	117.7	117.65
Değ eri										
En iyi										
fidan(1)										
Aç ı	11.433	13.313	0.801	12.852	14.825	0.799	10.279	12.724	12.209	11.996
Değ eri										



Şekil 6.6. Versiyon1, 1bbe proteinine ait aç ı sınıf ı uyuş ma sayısı

Tablo 6.13. Versiyon1, 1bbe proteini için test aç ı sınıf ı ve en iyi fidan için aç ı sınıf ı değ erleri

1bbe(1)	121.54	119.87	120.54	121.51	119.85	120.54	121.53	119.86	120.53	121.53
Aç ı	A5	A4	A4	A5	A4	A4	A5	A4	A4	A5
sınıf ı										
En iyi	119.47	117.83	118.49	119.46	117.82	118.48	119.47	117.83	118.48	119.47
fidan(1)										
Aç ı	A4	A3	A4	A4	A3	A4	A4	A3	A4	A4
sınıf ı										

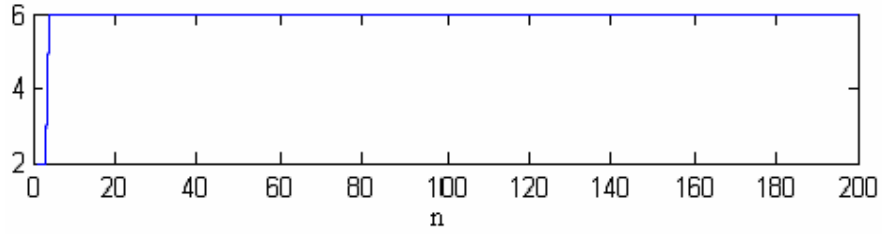


Şekil 6.7. Versiyon1, 1bbe proteinine ait bağı l hata değ iş imi

Tablo 6.14. Versiyon1 için lbbe proteini için test aç ı değerleri ve en iyi fidana ait aç ı değerleri

lbbe(1)										
Aç ı değeri	121.54	119.87	120.54	121.51	119.85	120.54	121.53	119.86	120.53	121.53
En iyi fidan(1)										
Aç ı değeri	119.47	117.83	118.49	119.46	117.82	118.48	119.47	117.83	118.48	119.47

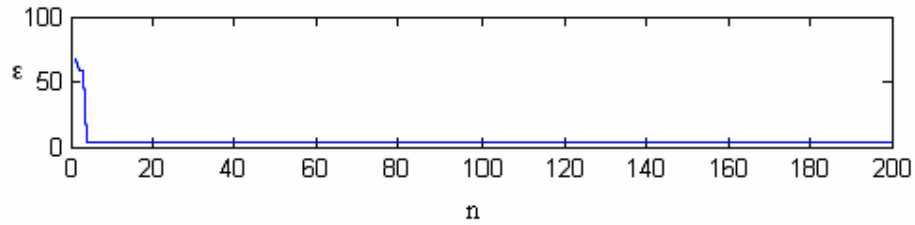
Fidan Gelişimi Algoritmasında Kullanılan Karıştır-Geliştir operatörünün birinci versiyonu ile her bir protein için yukarıdaki sonuçlar alınmıştır. Her bir protein için aç ı sınıfları doğru tahmin edilemediğinden elde edilen sonuçlar başarılı değildir. Önerdiğimiz ikinci Karıştır-Geliştir operatörü ile daha başarılı sonuçlar elde edilmiştir. Her bir protein için bağıl hata, eşleşen aç ı sayısı ve sınıflar aşağıda sırasıyla verilmiştir.



Şekil 6.8. Versiyon2, 1a9a proteinine ait aç ı sınıfı uyuşma sayısı

Tablo 6.15. Versiyon2, 1a9a proteini için test aç ı sınıfı ve en iyi fidana ait aç ı sınıfı değerleri

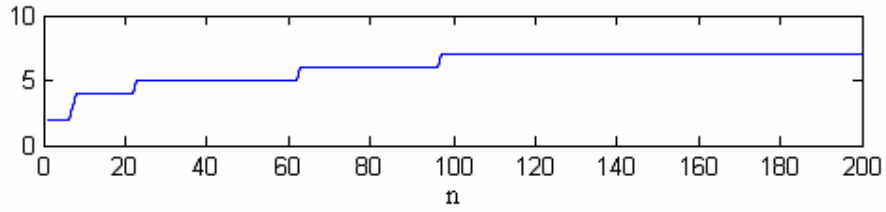
1a9a(2)	117.75	118.68	120.37	117.81	118.43	115.86	117.8	119.04	118.12	118.13
Aç ı Sınıfı	A3	A4	A4	A3	A4	A3	A3	A4	A4	A4
En iyi fidan(2)	116.34	118	118.68	118.24	119.72	117.9	116.86	117.66	116.28	117.49
Aç ı Sınıfı	A3	A4	A4	A4	A4	A3	A3	A3	A3	A3



Şekil 6.9. Versiyon 2 için 1a9a proteinine ait bağıl hata değ iş imi

Tablo 6.16. Versiyon2, proteini için test aç değeri ve en iyi fidana ait aç değeri

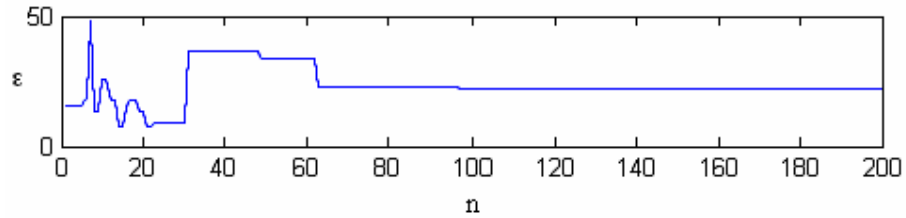
1a9a										
Aç değeri(2)	117.75	118.68	120.37	117.81	118.43	115.86	117.8	119.04	118.12	118.13
En iyi fidan	116.34	118	118.68	118.24	119.72	117.9	116.86	117.66	116.28	117.49
Aç değeri(2)										



Şekil 6.10. Versiyon2, 1a89 proteinine ait aç sınıfı uyuşma sayısı

Tablo 6.17. Versiyon2, 1a89 proteini için test aç sınıfı ve en iyi fidana ait aç sınıfı değeri

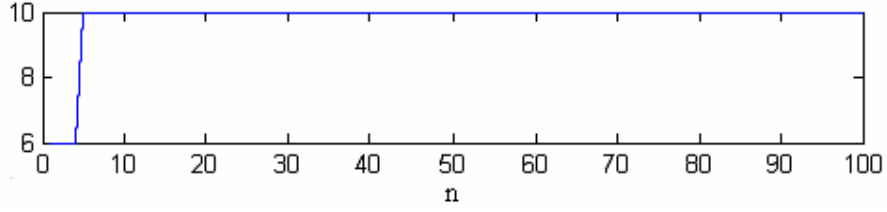
1a89(2)	117.96	120.76	118.11	118.01	118.58	116.67	117.23	118.73	117.7	117.65
Aç Sınıfı	A3	A4	A4	A4	A4	A3	A3	A4	A3	A3
En iyi fidan (2)	116.62	118.31	116.74	119.36	119.53	135.85	116.02	119.69	117.77	119.02
Aç Sınıfı	A3	A4	A3	A4	A4	A9	A3	A4	A3	A4



Şekil 6.11. Versiyon2, 1a89 proteinine ait bağıl hata değeri

Tablo 6.18. Versiyon2, 1a89 proteini için test aç değeri ve en iyi fidana ait aç değeri

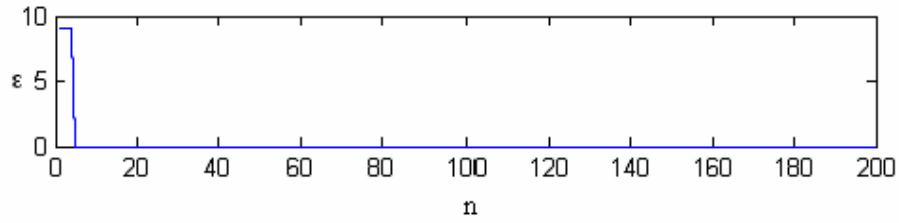
1a89(2)										
Aç değeri	117,96	120,76	118,11	118,01	118,58	116,67	117,23	118,73	117,7	117,65
En iyi fidan(2)										
Aç değeri	116,62	118,31	116,74	119,36	119,53	135,85	116,02	119,69	117,77	119,02



Şekil 6.12. Versiyon2, 1bbe proteinine ait aç sınıfı uyuşma sayısı

Tablo 6.19. Versiyon2 1bbe proteini için test aç sınıfı ve en iyi fidana ait aç sınıfı değerleri

1bbe(2)	121.54	119.87	120.54	121.51	119.85	120.54	121.53	119.86	120.53	121.53
Aç sınıfı	A5	A4	A4	A5	A4	A4	A5	A4	A4	A5
En iyi fidan (2)	121.53	119.86	120.53	121.51	119.85	120.52	121.53	119.86	120.52	121.53
Aç sınıfı	A5	A4	A4	A5	A4	A4	A5	A4	A4	A5



Şekil 6.13. Versiyon2, 1bbe proteinine ait bağıl hatanın değişimi

Tablo 6.20. Versiyon2, 1bbe proteini için test aç değeri ve en iyi fidana ait aç değeri

1bbe(2)										
Aç değeri	121,54	119,87	120,54	121,51	119,85	120,54	121,53	119,86	120,53	121,53
En iyi fidan(2)										
Aç değeri	121,53	119,86	120,53	121,51	119,85	120,52	121,53	119,86	120,52	121,53

1a9a, 189 ve 1bbe proteinleri için, Fidan Gelişim Algoritmasının önerdiğimiz hem versiyon1 hem de versiyon2 sonuçları elde edilmiştir. Elde edilen sonuçlar eşleşen aç değerlerine göre değerlendirilmelidir. Tüm proteinler için önerdiğimiz versiyon2 yöntemi daha iyi sonuçlar vermiştir. Elde edilen sonuçlar her iki versiyon için test edilen aç sayısı uyuşan aç sayısı ve başarı oranına göre değerlendirilmiştir.

Tablo 6.21. Versiyon 1 ve Versiyon 2 için performans değerleri

1a9a proteinine ait performans değerleri			
	Test açısı sayısı	Uyuşan açi sayısı	Başarı Oranı
Versiyon1	10	0	%0
Versiyon2	10	6	%60
1a89 proteinine ait performans değerleri			
	Test açısı sayısı	Uyuşan açi sayısı	Başarı Oranı
Versiyon1	10	0	%0
Versiyon2	10	7	%70
1bbe proteinine ait performans değerleri			
	Test açısı sayısı	Uyuşan açi sayısı	Başarı Oranı
Versiyon1	10	3	%30
Versiyon2	10	10	%100

Tablo 6.21 de proteinlerin ikincil yapılarının tahmininde versiyon ikinin her üç proteinin açi değeri içinde daha iyi performans gösterdiği gözlenmiştir.

7. SONUÇ

Geçen 20-30 yıl içerisinde Genomik teknolojilerin hızlı ilerlemesiyle, moleküler biyoloji alanında bir bilgi patlaması yaşandı. Genomik bilginin bu denli hızlı büyümesi; bilginin depolanması, organize edilmesi, indekslenmesi, özel programlar aracılığı ile görüntülenip analiz edilebilmesi için, bilgisayar tabanlı veri bankalarının oluşturulması gerekliliğini ortaya çıkardı.

Biyoinformatik biliminin amacı, yeni keşiflerin ışığı altında, biyolojinin ilkelerini global bir bakış açısıyla görebilmek ve bu ilkeleri bütünleştirebilmektir. Veri tabanlarının gelişimi sadece dizayn edilmesi ile değil bunun yanında karmaşık ara yüzlerin oluşturulması, veri girişleri, bildirimler ve verilerin revizyonlarının sağlanması ile gerçekleştirilmiştir. Sonuçta tüm bu çalışmalar normal hücresel aktivitenin ayrıntılı bir resmini ortaya çıkardığı gibi, araştırmacılara farklı hastalık durumları hakkında bilgi de sağlamaktadır.

Proteinlerin fonksiyonel karakterizasyonu biyolojideki en önemli problemlerden biridir. Bir proteinin yapısı fonksiyonunu belirlemekte anahtar rol oynar. Proteinlerin regülasyonları, hücre içindeki yerleşimleri, motifleri, anlatımları, yapıları, post-translasyonel modifikasyonları, etkileşimleri yaşamsal anlamda çok önemlidir. Diziler değerli bilgiler sağlamasına rağmen, onların yüksek değişkenliği fonksiyonel olarak yeterli bilgi edinilmesini olanaksız kılar. Bu amaçla proteinlerin üç boyutlu yapıları belirlenmeye çalışılmaktadır. Proteinlerin ikincil yapısı, aminoasitler arası etkileşimlerle (hidrojen bağı) oluşan α -helix ve β -tabaka katlanmalarını içerir. Bu katlanmalar belli aminoasitlerin varlığında oluşur. Örneğin prolin ve glisin aminoasitleri helix yapısı oluşumunu bozarken metiyonin, alanin, lösin, ... gibi aminoasitler bu yapının oluşumuna katkıda bulunur.

İkincil yapı tahmini yöntemleri α ve β katlanmaları proteinin aminoasit dizisinden (birincil yapısından) oluşturmaya çalışır. Pauling 1951 yılında diziden α ve β katlanmalarının tahminini doğru bir şekilde yapmayı başarmışlardır. Bu tarihten sonra ikincil yapı tahmin yöntemleri Pauling'in hesaplamalarından yola çıkılarak geliştirilmiştir. Ayrıca proteinlerin sınıflandırılmasında (α proteinler, β proteinler, α/β proteinler) ve 3 boyutlu yapı tahmin yöntemi olan fold recognition'da da bu katlanmalardan yararlanır. İkincil yapı tahmini için geliştirilen çok çeşitli yöntemler vardır. Bu yöntemlerden bölüm 5 te ayrıntılı olarak bahsedilmiştir.

Bu tezde proteinlerin ikincil yapılarının tahmini için Günümüzde kullanılan yumuşak hesaplama tekniklerine alternatif bir yöntem olan Fidan gelişimi Algoritması uygulanmıştır. Tüm mühendislik dallarında optimizasyon problemleri bir amaç fonksiyonuna sahip doğadaki algoritmalar ile çözülür. Fidan Gelişimi Algoritması hem yerellik ve hem de küresellik

özelliğine sahip yeni bir hesapsal yöntemdir. En iyi çözüm için tüm algoritmalarda bir amaç fonksiyonu vardır. Bu yaklaşım algoritmalarının mühendislik problemlerine uygulanmasının nedeni, problemin matematiksel modelinin kurulmasının zor veya çözüm zamanı maliyetinin çok yüksek olmasıdır. Bizim problemimizi çözmek için kullandığımız fidan gelişimi algoritması ise doğadaki fidanların gelişimini esas alan bir algoritmadır. Geliştirilmiş olan yöntemde, aday çözümlerin başlangıçta arama uzayında düzgün dağıtılması sağlandıktan sonra hem yerel ve küresel arama gerçekleştirilmekte ve benzer çözümler arasında bilgi değişimi imkânı olmaktadır. Tavlama benzetimi, tabu arama ve tepe tırmanma algoritmalarında olduğu gibi geçici çözümün komşularını incelemeyi içerdiği gibi birden fazla geçici çözüm ile başlamayı ve bu çözümleri çözüm uzayına düzgün olarak dağıtmaya çalışmayı ve işlemler devam ederken küresel değişimleri de içermektedir. Tüm bunlardan başka, genetik algoritmalarındaki gibi bir uygunluk fonksiyonu belirleme zorunluluğu yoktur; direkt olarak amaç fonksiyon değeri kullanılır. Bütün bu özelliklerinden dolayı daha güçlü bir hesapsal yöntem olması beklenmiştir.

Tablo 7.1 Versiyon1 ve Versiyon2 başarı oranları

1a9a proteinine ait performans değerleri			
	Test açısı sayısı	Uyuşan açısı sayısı	Başarı Oranı
Versiyon1	10	0	%0
Versiyon2	10	6	%60

1a89 proteinine ait performans değerleri			
	Test açısı sayısı	Uyuşan açısı sayısı	Başarı Oranı
Versiyon1	10	0	%0
Versiyon2	10	7	%70

1bbe proteinine ait performans değerleri			
	Test açısı sayısı	Uyuşan açısı sayısı	Başarı Oranı
Versiyon1	10	3	%30
Versiyon2	10	10	%100

Yapılan çalışmalarda diğer çalışmaların aksine proteinlerin coil helix ve b-sheet yapıları üzerinde dekil merkezci karbon açıları arasındaki beta açıları kullanılmıştır. Daha önce yapılan çalışmalarda yaklaşık olarak %60 -%80 arasında başarı gözlenmiştir. Örneğin Yapay sinir ağlarıyla yapılan Proteinlerin ikincil yapı tahmininde tezde kullandığımız 3 test datası kullanılmış ve aynı açılar üzerinde %70, %62 ve % 49.12 oranlarında başarı elde edilmiştir[10].

Bizim çalışmamızda ise öncelikle Fidan Gelişimi Algoritmasının 1.versiyonu uygulanmıştır. Versiyon1’de elde edilen sonuçlar aşağıdaki gibidir. Tablo 7.1’de de görüldüğü gibi versiyon1 sonucunda istenilen düzeyde başarı elde edilememiştir. Versiyon1 3 test datası için %0 ,%0, %30 oranlarında başarı göstermiştir. Bu nedenle amaç değerini belirleyen fonksiyonda bir değişiklik yapılmış ve versiyon2 kullanılmıştır.4. bölümde de anlatıldığı gibi Versiyon2 versiyon1den farklı olarak karıştır geliştir operatörünü kullanır. Fidan gelişimi algoritmasının ikinci versiyonu kullanılarak tablo 7.1’de ki sonuçlar elde edilmiştir.Versiyon1 de test dataları üzerinde %0, %0, %30 oranında başarı elde edilirken Versiyon2 için %60, %70, %100 oranlarında başarı elde edilmiştir. Bu sonuçlar literatürdeki daha önceki çalışmalarla karşılaştırıldığında sonuçların daha iyi olduğu gözlenmiştir.

KAYNAKLAR

- [1]-<http://www.Ödevsel.com>
- [2]. 89-1-07-pdf
- [3]- A. KARCI, B. ALATAŞ, E.AKIN Fidan Gelişim Algoritması
- [4]- ogrenci_sunumlari_genmuh10-4
- [5] biyoloji.ksu.edu.tr/biyoinfo_dosyalar/ biyoinformatik_uzerine.pdf
- [6] CEBECİ, Z., BEK, Y., 1999. Tarımsal Enformatik Eğitimi Mümkün mü? 3. Tarımda Bilgisayar Uygulamaları Sempozyumu (Çağrılı Bildiri). 3-6 Ekim 1999, Çukurova Üniversitesi, Adana
- [7]- unite01-GÖZÜKARA, Engin M. Biyokimya, Ankara, Ofset Pepianet LTD.Şt. 1990.
- FEHMÜĞER, Albert L. Biochemistry, New York, Nort Publishers, Inc 1988.
- [8]- 89-1-07-[P1]
- [10] Murat DEMİR, Yumuşak Hesaplama Teknikleri ile Protein Yapı Tahmini, Fırat Üniversitesi Yüksek Lisans tezi, 2006
- [11] Ali KARCI ,Bilal ALATAŞ, Erhan AKIN Fidan Gelişim Algoritması
- [12] A. Karcı, *Novelty in the Generation of Initial Population for Genetic Algorithms*, KES2004, 20-24, Wellington, New Zealand, 2004.
- [13] Pierre Baldi¹,Seren Breunak², Paolo Frasconi³, Gianluca Pollastri¹, and Giovanni Soda³, Bidirectional Dynamics for protein Secondary Structure Prediction, Sequence Learning,
- [14] V.Di Farnesco ,P.J.Munson and J.Garnler,Use of Multiple Alignment in Protein Secondary Structure Prediction,Proceedings of the 28th Annual Hawaii International Conference on System Sciences,1995.
- [15] M.N. Nguyen and J.C.Rajapakse,Two-Stage Multi-Class Support Vector Machines to Protein Secondary Structure Prediction,Pacific Symposium on Biocomputing 10:346-357,2005.
- [16] Kuen-Pin Wu,Hsin-Nan Lin,Jia-Ming Chang,Ting-Yi Sung and Wen-Lian Hsu,Hyprosp:A Hybrid Protein Secondary Structure Prediction Algoritm,Instute of Information Science ,Academia Sinica ,Taipei Taiwan,2004.
- [17] Yav Liv,Jaime Carbonell,Judith Klein-Seetharaman, Vanatti Gopalaknishnan,ContextSensitive Vocabulary and its Appliation in Protein Secondary Structure Prediction,Sheffeld, South Yorkshire, U.K, 2004
- [18] Varun Aggorwal, Prediction of Secondary Structure Using Genetic Programming, Summer Internship Project Report Stockholm University, Sweeden 2003.

- [19] Victor Robles, maria S.Perez ,Vanessa herves Jose M.Pena and Pedro Larranaga,Paralel Stochastic Search for protein Secondary Structure Prediction ,Springer-Verlag Berlin Heidelberg, 2004
- [20] Volkan KURT, Protein Structure Prediction Using Decision Lists,Master Thesis Koç University, 2005
- [21] Howard Holley and Martin Karplus, Protein Secondary Structure Prediction with a neural network , Department of Chemistry ,Harvard University Cambridge ,1989
- [22]İrem ERSÖZ,Secondary Structure Prediction of Hemoglobin by Using Combined Neural Networks, Master Thesis Çukurova University, 2003

ÖZGEÇMİŞ
SEDA ARSLAN TUNCER

Fırat Üniversitesi
Mühendislik Fakültesi
Bilgisayar Mühendisliği Bölümü
23119, ELAZIĞ

E_posta : seda@ebk.gov.tr

1983	Kahramanmaraş'ta doğdu.
1996-2000	Kahramanmaraş Atatürk Y.D.A.L mezun oldu.
2000-2005	Fırat Üniversitesi Bilgisayar Mühendisliği Bölümünden mezun oldu.
2006-2008	EBK Genel Müdürlüğü'nde Bilgisayar Mühendisi olarak Görev yapmaktadır.