

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**PEPTİT BAĞ AÇILARINI KULLANARAK YAPAY SİNİR AĞI
TABANLI PROTEİNLERİN SEKONDER YAPI TAHMİNİ**

Murat DEMİR

Tez Yöneticisi :
Yrd.Doç.Dr. Ali KARCI

YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

ELAZIĞ, 2006

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**PEPTİT BAĞ AÇILARINI KULLANARAK YAPAY SİNİR AĞI
TABANLI PROTEİNLERİN SEKONDER YAPI TAHMİNİ**

Murat DEMİR

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 : Prof. Dr. Bilal ÜSTÜNDAĞ

Üye : Yrd. Doç. Dr. Mehmet KAYA

Üye:

Üye:

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

TEŞEKKÜR

Tez çalışmalarım esnasında, ilgi ve yardımlarını esirgemeyen danışmanım Sayın Yrd.Doç.Dr. Ali KARCI'ya ve desteklerini sürekli hissettiğim aileme teşekkürlerimi sunarım.

İÇİNDEKİLER

TEŞEKKÜR

İÇİNDEKİLER.....	I
ŞEKİLLERİN LİSTESİ.....	IV
TABLoların LİSTESİ.....	VI
ÖZET.....	VII
ABSTRACT.....	VIII
1. GİRİŞ.....	1
2. BİYOİNFORMATİK.....	14
2.1. Biyoinformatik	14
3. BİYOKİMYA VE PROTEİNLER.....	17
3.1. Biyokimya	17
3.2. Proteinler	17
3.2.1. Enzimler.....	18
3.2.2. Transport Proteinleri.....	18
3.2.3. Lipoproteinler.....	18
3.2.4. Besin ve Depo Proteinleri.....	19
3.2.5. Kontraktil Sistemde Bulunan Proteinler.....	19
3.2.6. Yapısal Proteinler.....	19
3.2.7. Savunma Proteinleri.....	20
3.2.8. Regülasyonu Sağlayan Proteinler.....	20
3.3. Amino Asitler.....	20
3.4. Proteinlerin Yapıları.....	22
3.4.1. Primer Yapı.....	22
3.4.2. Sekonder Yapı.....	22
3.4.2.1. α - Heliks Yapı.....	23
3.4.2.2. Beta Kırmalı Tabaka.....	24
3.4.2.3. Beta Bendler Yapı.....	26
3.4.2.4. Tesadüfi Kıvrılamalar.....	26
3.4.3. Tersiyer Yapı.....	27
3.4.4. Kuaterner Yapı.....	28
3.5. Peptit Bağı.....	28
4. YAPAY SİNİR AĞLARI.....	30
4.1. Biyolojik Sinir Sistemi.....	30

4.2. Yapay Sinir Ağı.....	31
4.2.1. Yapay Sinir Ağının Özellikleri.....	33
4.2.1.1. Makine Öğrenmesi Gerçekleştirme.....	33
4.2.1.2. Bilgi Saklama.....	33
4.2.1.3. Klasik Yöntemlere Benzememe.....	33
4.2.1.4. Öğrenme.....	33
4.2.1.5. Doğrusal Olmama.....	34
4.2.1.6. Genelleme.....	34
4.2.1.7. Uyarlanabilirlik.....	34
4.2.1.8. Hata Toleransı.....	34
4.2.1.9. Eksik Bilgi ile Çalışabilme.....	34
4.2.1.10. Donanım ve Hız.....	35
4.2.1.11. Analiz ve Tasarım Kolaylığı.....	35
4.2.2. Yapay Sinir Hücresi.....	35
4.2.2.1. Girdiler.....	36
4.2.2.2. Ağırlıklar.....	36
4.2.2.3. Toplama Fonksiyonu.....	36
4.2.2.4. Aktivasyon Fonksiyonu.....	37
4.2.2.5. Çıktı.....	39
4.2.3. Yapay Sinir Ağına Yapısı.....	39
4.2.3.1. Giriş Katmanı.....	40
4.2.3.2. Ara Katman.....	40
4.2.3.3. Çıkış Katmanı.....	40
4.2.4. Yapay Sinir Ağı Yapıları.....	40
4.2.4.1. İleri Beslemeli Yapay Sinir Ağları.....	41
4.2.4.2. Geri Beslemeli Yapay Sinir Ağları.....	42
4.2.4.3. Bellek Hücreli Yapay Sinir Ağları.....	43
4.2.4.4. Radyal Tabanlı Fonksiyon Ağları.....	44
4.2.4.5. Fonksiyonel Link Ağları.....	46
4.2.5. Yapay Sinir Ağlarında Öğrenme.....	47
4.2.5.1. Eğitici Öğrenme.....	48
4.2.5.2. Eğitici Öğrenme.....	49
4.2.5.3. Takviyeli Öğrenme.....	49
4.2.5.4. Öğrenme Algoritmaları.....	50
5. PROTEİN YAPI TAHMİNİ İÇİN UYGULANAN DİĞER YÖNTEMLER.....	55

6. GELİŞTİRİLEN YÖNTEM VE SONUÇLARI.....	59
7. SONUÇ.....	78
KAYNAKLAR.....	80
ÖZGEÇMİŞ.....	82

ŞEKİLLERİN LİSTESİ

Şekil 1.1. tRNA'ya doğru amino asidin bağlanması.....	3
Şekil 1.2. DNA'dan polimeraz enzimi ile bilginin alınması.....	5
Şekil 1.3. Gereksiz bilginin bükülüp atılması.....	8
Şekil 3.1. Amino asit genel yapısı.....	21
Şekil 3.2. Primer yapı.....	22
Şekil 3.3. α - heliks yapı	23
Şekil 3.4. Hidrojen Bağları.....	24
Şekil 3.5. Anti paralel ve paralel beta kırmalı yapı.....	25
Şekil 3.6. Beta kırmalı yapıda hidrojen bağları.....	25
Şekil 3.7. Beta bendler yapı.....	26
Şekil 3.8. Tesadüfi kıvrılma yapısı.....	26
Şekil 3.9. Tersiyer yapı.....	27
Şekil 3.10. Kuaterner yapı.....	28
Şekil 3.11. Peptit bağı oluşumu.....	29
Şekil 4.1. Biyolojik sinir sistemi blok diyagramı.....	30
Şekil 4.2. Biyolojik sinir hücre yapısı.....	31
Şekil 4.3. Yapay sinir hücresi modeli.....	35
Şekil 4.4. Eşik aktivasyon fonksiyonu.....	37
Şekil 4.5. Doğrusal aktivasyon fonksiyonu.....	38
Şekil 4.6. Logaritma sigmoid aktivasyon fonksiyonu.....	38
Şekil 4.7. Tanjant hiperbolik aktivasyon fonksiyonu.....	39
Şekil 4.8. Yapay sinir ağı modeli.....	40
Şekil 4.9. İleri beslemeli yapay sinir ağı modeli.....	41
Şekil 4.10. Geri beslemeli yapay sinir ağı modeli.....	42
Şekil 4.11. Bellek hücreli yapay sinir ağı modeli.....	43
Şekil 4.12. Radyal tabanlı fonksiyon ağları modeli.....	45
Şekil 4.13. Fonksiyonel link ağları modeli.....	46
Şekil 4.14. Eğitici öğrenme blok şeması.....	48
Şekil 4.15. Eğitici öğrenme blok şeması.....	49
Şekil 4.16. Takviyeli öğrenme blok şeması.....	49
Şekil 4.17. İleri beslemeli çok katmanlı sinir ağı modeli.....	50
Şekil 5.1. Bayesian Network Modeli.....	55
Şekil 5.2. Yapı pozisyon tablosu.....	56

Şekil 6.1. Tripeptit yapıda kullanılan bağ açıları.....	59
Şekil 6.2. Kullanılan yapay sinir ağı modeli.....	60
Şekil 6.3. Yapay sinir ağı eğitimi birinci örnek şekil.....	68
Şekil 6.4. Yapay sinir ağı eğitimi ikinci örnek şekil.....	68
Şekil 6.5. Yapay sinir ağı eğitimi üçüncü örnek şekil.....	69
Şekil 6.6. Yapay sinir ağı eğitimi dördüncü örnek şekil.....	69
Şekil 6.7. Yapay sinir ağı eğitimi beşinci örnek şekil.....	70

TABLoların LİSTESİ

Tablo 4.1. Toplam fonksiyonu örnekleri.....	36
Tablo 6.1. Amino asitlerin nümerik değere dönüşüm tablosu.....	61
Tablo 6.2. Açık sınıfları.....	61
Tablo 6.3. Eğitim örnekleri.....	62
Tablo 6.4. 1bbe proteini açık değeri	70
Tablo 6.5. 1bbe proteini açık sınıfları.....	71
Tablo 6.6. 1a89 proteini açık değeri.....	71
Tablo 6.7. 1a89 proteini açık sınıfları.....	72
Tablo 6.8. 1a9a proteini açık değeri.....	72
Tablo 6.9. 1a9a proteini açık sınıfları.....	73
Tablo 6.10. 1a89 proteini için 0,2 derece hata payı sınıf tablosu.....	74
Tablo 6.11. 1a89 proteini için 0,5 derece hata payı sınıf tablosu.....	74
Tablo 6.12. 1a89 proteini için 1 derece hata payı sınıf tablosu.....	75
Tablo 6.13. 1a9a proteini için 0,2 derece hata payı sınıf tablosu.....	75
Tablo 6.14. 1a9a proteini için 0,5 derece hata payı sınıf tablosu.....	76
Tablo 6.15. 1a9a proteini için 1 derece hata payı sınıf tablosu.....	76

ÖZET

Yüksek Lisans Tezi

PEPTİT BAĞ AÇILARINI KULLANARAK YAPAY SİNİR AĞI TABANLI PROTEİNLERİN SEKONDER YAPI TAHMİNİ

Murat DEMİR

Fırat Üniversitesi

Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

2006, Sayfa : 82

İnsan bedeni, yeryüzündeki en karmaşık makinedir. Hayatımız boyunca bu bedenle görür, işitir, nefes alır, yürür, koşar ve zevk alırız. Vücudumuzdaki her şey milimetrenin binde biri büyüklüğündeki hücrelerden oluşur. Hücreler vücudun hangi parçasını oluşturuyorlarsa bu bölgede ihtiyaç duyulan boyuta ve şekle sahip olurlar. Bu yapı içerisinde önemli bir yere sahip olan proteinler ise eşsiz tasarımlarıyla insanın aklını hayrete düşürecek kadar karmaşık bir o kadarda kusursuz bir yapıya sahiptir.

Unutmamak gerekir ki, bugün dünyada yüzbinlerce bilim adamı protein konusunda çalışmaktadır. Buna rağmen bu kusursuz makromoleküller hala akıllara durgunluk verecek kadar muhteşem yapısıyla keşfedilemeyen yönleriyle araştırmacıların ilgilerini çekmektedirler.

Yapıları itibarıyla proteinler birincil, ikincil, üçüncül ve dördüncül yapı olmak üzere dört düzeye sahiptirler. Birincil yapı amino asit ardışıklarından oluşan düz zincir yapıdır. İkincil yapı bölgesel şekillenmelerden meydana gelen yapıdır. Üçüncül yapı ikincil yapıların bir araya gelmesi ile oluşur. Dördüncül yapı ise bir veya daha fazla polipeptit zincirinin bir araya gelmesi ile arada oluşan etkileşimler sonucunda meydana gelen yapıdır.

Bu tezde proteinlerin ikincil yapılarının üzerinde çalışılmış ve ikincil yapıyı ortaya koyabilmek için peptitlerdeki merkezî karbon atomları arasındaki bağ açılarından yola çıkarak ikincil yapının tahmini yapılmaya çalışılmıştır. Bunu gerçekleştirebilmek için yapay sinir ağlarından faydalanılmıştır. İleri beslemeli yapay sinir ağı modeli üzerinde Levenberg-Margquardt öğrenme algoritması ile sonuçlar elde edilmiştir. Sonuçta üç farklı test datası üzerinde %70 , %68.42 ve % 49.12 oranlarında başarı elde edilmiştir.

Anahtar Kelimeler : Proteinler , Peptitler, Yapay Sinir Ağları

ABSTRACT

Master Thesis

ESTIMATION OF PROTEIN SECONDARY STRUCTURE BASED ON ARTIFICIAL NEURAL NETWORKS BY USING PEPTIDE BOND ANGLES

Murat DEMİR

Fırat University

Institute of Natural and Applied Science

Computer Engineering Division

2006, Pages : 82

Human body is the most complex machine on the earth. That's why, we can see, hear, breathe, run and enjoy along our life. All the parts of our body constitute from cells. Cells have different sizes and shapes with respect to organs they constitute. There are complex and amazing important structures in the cells and they are called proteins.

They are a lot of researchers study on the proteins on the earth. Although these researches, these perfect macromolecules have a lot of undiscovered aspects, and they attract researchers.

Proteins have four different structures such as primary, secondary, tertiary and quaternary. The primary structure consists of amino acids sequence and chain. Protein secondary structure refers to certain common repeating structures found in proteins. There are two types of secondary structures: alpha-helix and beta-pleated sheet. Tertiary structure is the full 3-dimensional folded structure of the polypeptide chain. Quaternary structure is only present if there is more than one polypeptide chain. With multiple polypeptide chains, quaternary structure is their interconnections and organization.

In this thesis, we studied on the secondary structures of proteins and in order to depict the secondary structure, we tried to estimate the angle between polypeptide bonds between alpha-carbons. At this aim, we used artificial neural networks. The results obtained by using feedforward neural network with Levenberg-Margquardt learning algorithm. At the consequence, we have obtained success rates for three different data set as 70% , 68.42% and 49.12%.

Keywords : Proteins , Peptides, Artificial Neural Networks.

1. GİRİŞ

İnsanoğlu yaratıldığından bu yana sürekli ihtiyaçlarına çözümler aramakta ve bunları karşılamak amacıyla sürekli bilimin her dalında gelişmeler kaydetmektedir. Bu bilim dallarından ikisi de biyokimya ve günümüzde artık ayrılmaları imkansız hale gelen biyoinformatiktir.

Bu bilim dallarına olan ilgi, gerek insanın kendi yapısı ile ilgili bilgilerle uğraşmasından; gerekse bilgisayar tekniklerinin git gide artan kolay ve hızlı çözümler sunma özelliklerinden dolayı günümüzde giderek artmaktadır.

Biyoinformatik adı verilen ve biyoloji ve biyokimya bilimi ile bilgisayar biliminin tekniklerini birleştirip, bu alanlardaki problemlere çözüm arayan bilim dalı, günümüzde çok popüler bir bilim dalı haline gelmiştir.

Biyokimya biliminin içerisinde ayrı bir yeri olan proteinler ise, canlılığın temel taşlarından biridir. Bu makromoleküller aynı zamanda hastalıklar ve hastalıkların tedavisinde geliştirilen ilaçların sahası içerisinde de etkili olduklarından giderek de ilgi duyulan bir konu olmaktadır. Günümüzde hedef proteinleri belirlemeden ilaç geliştirme programına başlanması nadirdir. Bunun için de, biyolojik yanıt mekanizmasının çok iyi bilinmesi gerekmektedir. [1]

Proteinler DNA'nın gerekli bölgesindeki şifrenin kodlanması ve bunun ribozomlarda okunarak uygun amino asitlerin birbirleriyle birleşmesi sonucu oluşturdukları organik bir yapıdır. Proteinler bir organizma ve hücredeki yapısal ve fonksiyonel reaksiyonların temelini oluştururlar. Proteinler, amino asitler arasında peptid bağlarının oluşmasıyla kurulurlar ve daha sonra uygun konformasyonları kazanırlar.

Proteinler organizmanın yapısını ve fonksiyonunu temin ederler ancak, rasgele protein eylemleri “yasam” sağlayamazlar. Bilim adamları kompleks davranışlara yol açan protein fonksiyonlarını oluşturan mekanizmayı belirlemek ihtiyacındaydılar. Onların araştırmaları proteinlerin değişken oldukları (stabil olmadıkları) gerçeği üzerine kuruluydu. Tıpkı arabanın parçaları gibi, kullanıldığı zaman proteinlerde “aşınmaya tabi” idi. Eğer belli bir patikadaki bireysel protein aşınır ve yenisi ile değiştirilmezse patika duracaktı. Fonksiyonunu tamamlayabilmesi için proteinin değiştirilmesi lazımdı. Buna bağlı olarak davranışsal fonksiyonları, patikaları oluşturan proteinlerin varlığı veya yokluğunu regüle ederek (düzenleyerek) , kontrol edilebileceği düşünülüyordu. Protein parçalarını değiştirmede kullanılacak kaynak ise soyaçekim için gerekli ”karakterin” bir sonraki nesle geçirilmesini sağlayan “hafıza-bellek” faktörlerine bağlıydı.

Protein sentezini kontrol eden soyaçekim faktörlerinin araştırılması ise DNA'ya götürdü. 1953 yılında Watson ve Crick “genetik kod” gizemini aydınlığa kavuşturdular. DNA'nın bir proteini oluşturan amino asit sıralamalarını belirleyen moleküler bir “anakopya-blueprint” olduğunu açıkladılar. Her bir proteinin anakopyası olan DNA'ya ise gen dediler. Proteinler bir organizmanın karakterini belirlediğinden ve bu protein yapılarının da DNA'nın içinde şifrelenmiş olmasından dolayı da biyologlar DNA'nın öncelliği (Primacy) olarak bilinen dogmayı oluşturdular. Bu bağlamda, öncelik “kontrolün ilk seviyesi” anlamına geliyordu.

Kısaca, DNA'nın canlı organizmaların yapısını ve davranışını “kontrol ettiği” sonucuna varıldı. DNA bir organizmanın karakterini “belirlediğine” göre genetik belirleyicilik kavramını ortaya koymak uygundu. Bu görüşe göre organizmanın yapısını ve davranışlarını genler belirliyordu.

Canlıların yaşamlarını sürdürebilmelerinde hayati bir önemi olan proteinlerin, hücre içinde üretimleri için dünya üzerindeki hiçbir üretim sistemiyle kıyaslanamayacak komplekslikte ve düzende, kusursuz bir sistem bulunmaktadır. Bu kompleks üretim sisteminde hiçbir hataya yer yoktur. Herhangi bir aşamada meydana gelen bir aksaklık, hemen güvenli kontrol sistemi sayesinde düzeltilir. Böylece canlının yaşamını sürdürmesini sağlayacak olan proteinler hiçbir aksama olmadan, tam gerektiği zamanda tam gereken yerde ve şekilde üretilir.

Protein üretiminin bir diğer mucizevi özelliği de çok yüksek bir hızda gerçekleşmesidir. Örneğin 100 amino asit taşıyan bir protein molekülü, E. coli bakterisinin hücresi tarafından 5 saniyede sentezlenir. Bu öylesine büyük bir hızdır ki, böyle bir hızda bütün üretim sürecini kusursuz biçimde tamamlayabilen bir fabrika, yeryüzünde mevcut değildir. Bu hız canlı için çok önemlidir, çünkü hücrelerde canlılığın sürdürülebilmesi için her an birçok proteine ihtiyaç duyulur.

RNA'daki bilginin proteine çevrilmesi işlemine translasyon denir. mRNA'da sadece dört farklı Nükleotid bulunurken, proteinlerde 20 değişik amino asit bulunur. Bu nedenle, mRNA ile proteinler arasında bire bir ilişki olamaz. Genlerdeki ve dolayısıyla mRNA'daki nükleotid dizisi ile proteinlerdeki amino asit dizisi arasındaki ilişki genetik şifre'dir.

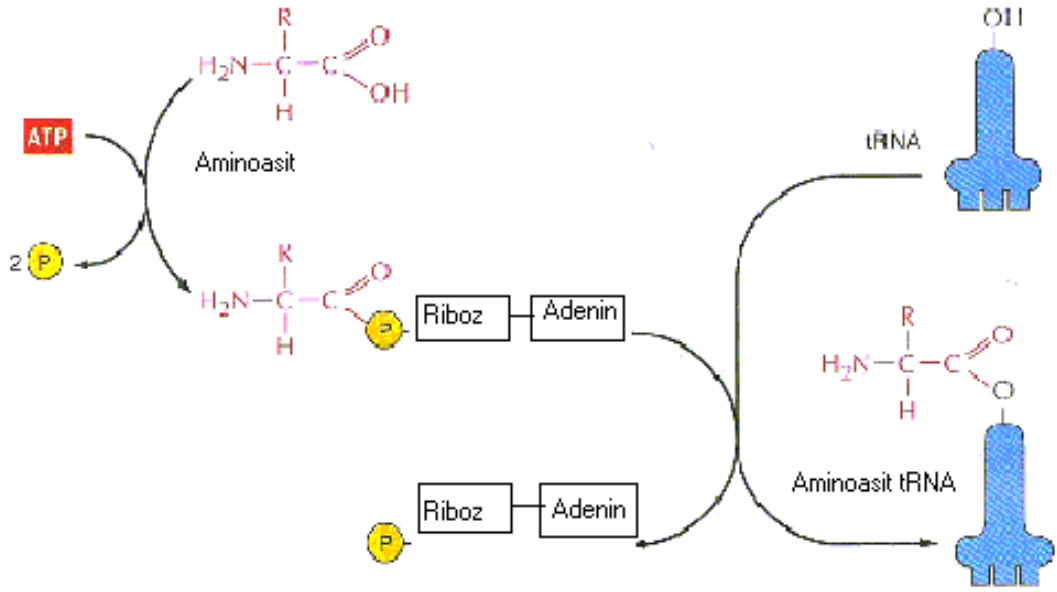
4 değişik Nükleotidin 4^3 olası dizilimi, 64 olası kombinasyon oluşturur ki bu sayı, 20 çeşit amino asit şifresi olmaya fazlasıyla yeterlidir.

Nükleotidlerin mRNA üzerindeki üçlü dizilimlerine kodon adı verilir. Her bir kodon bir amino asidi şifreler. Bazı amino asitlerin birden fazla şifresi bulunur (kodon gevşekliliği).

Formilmethionin tRNA'nın tanıdığı AUG, protein sentezinde başlama kodonu iken, UAA, UAG ve UGA kodonları dur kodonu'dur.

Bazı tRNA antikodonlarının 5'pozisyonundaki baz, kodondaki baz ile gevşek baz eşleşmesi (Wobble baz etkileşimi) yapar. Bunun amacı, bir tRNA tarafından tanınan kodonların sayısının artırılmasıdır. Bu durumda örnek verilecek olursa 5'pozisyonundaki nadir bir baz olan inosin, A, C ve U ile baz eşleşmesi yapabilir. Bu nedenle bakteride 20 amino aside karşılık, 61 kodon ve yaklaşık 31 çeşit tRNA bulunur. tRNA'ların sayısı bir türden diğerine değişir. Örneğin insanda 497 tane tRNA geni bulunurken, sadece 48 değişik antikodon bulunmaktadır.

Her bir tRNA'ya doğru amino asidin tanınıp, bağlanması gerekir. Ayrıca, bu süreç enerji gereksinen bir sentez sürecidir. Bu nedenle, her hücrede her bir amino aside özgü aminoasit tRNA sentetaz enzimleri bulunur. Enzim katalizli tepkimede ATP hidrolizi ile aktive edilen amino asit, tRNA'nın 3' ucundaki adenin bazına ait riboz şekerin 3'-OH grubuna bağlanır.



Şekil 1. 1. tRNA'ya doğru amino asidin bağlanması

Protein üretimi sırasında ise birçok protein aynı anda faaliyet gösterir. Hücrelerin içinde protein üretimi için gereken bütün parçalar eksiksiz biçimde bir arada çalışırlar. 80'in üzerinde ribozom proteini, 20'nin üzerinde amino asit habercisi olan moleküller, bir düzinenin üzerinde yardımcı enzim, 100'ün üzerinde son işlemleri gerçekleştiren enzimler, 40'ın üzerinde RNA molekülü olmak üzere yaklaşık 300 makro molekül, bir koordinasyon halinde protein sentezinde rol alır. Büyük bir mühendis kadrosunun bile zorlukla koordine edebileceği bu kusursuz üretim, milimetrenin binde biri kadar küçük bir alanda, bundan çok daha küçük yüzlerce molekülün

yoğun faaliyetiyle yaşamın devam edebilmesini sağlar. Bu üretimde görev alan moleküllerden tek bir tanesinin olmaması durumunda tüm üretim zinciri aksar. Bu da protein üretiminin canlılardaki "indirgenemez kompleks" yapılardan biri olduğunun bir göstergesidir. Yani böyle bir sistemin içinden tek bir parça dahi çıkarılsa tüm yapı bozulmaya uğrar. Örneğin sadece üretimi sonlandırma ve üretilen yeni proteini serbest bırakma elemanının (proteininin) olmaması, protein üretiminin dengesini bozmaya yeterlidir.

Protein Sentezi için İlk Sinyal :

Vücutta herhangi bir proteine ihtiyaç duyulduğu zaman bu ihtiyacı ifade eden bir mesaj, üretimi gerçekleştirecek olan hücrelerin çekirdeklerinde bulunan DNA molekülüne ulaştırılır. Burada dikkat edilmesi gereken çok önemli bir nokta bulunmaktadır; vücutta herhangi bir protein ihtiyacı olduğunda yine protein olan bazı haberciler, nereye başvurmaları gerektiğini bilerek, tüm vücutta ilgili yeri bulabilmekte, ihtiyaç mesajını doğru yere doğru şekilde iletebilmektedirler. Bu iletişimi sağlayan protein, kendisine göre karanlık bir gezinti olan vücudun içinde kaybolmadan yolunu bularak, taşıdığı mesajı kaybetmeden yada herhangi bir parçasına zarar vermeden oraya ulaştırmaktadır. Yani her bir parçada çok büyük bir görev bilinci bulunmaktadır.

Hücre çekirdeğine gelen mesaj, bir dizi kompleks ve son derece organize işlemten sonra proteine dönüşür. Protein talebinin, vücuttaki 100 trilyon hücreden doğru hücrelere ulaşması, mesajı alan hücrenin kendisinden ne istendiğini anlayarak hemen işe koyulması ve kusursuz bir sonuç elde etmesi insanda hayranlık uyandıran olaylardır. Çünkü burada söz edilen bilinç, akıl, bilgi ve irade sahibi insanların oluşturduğu bir topluluk değil, fosfor, karbon, yağ gibi maddelerden oluşmuş şüursuz ve gözle dahi görülemeyecek kadar küçük varlıklardır. Talimatın alınmasından sonra ilk işlem, üretilmesi istenen proteinle ilgili bilgilerin DNA'dan alınmasıdır.

DNA'ya sipariş verilir :

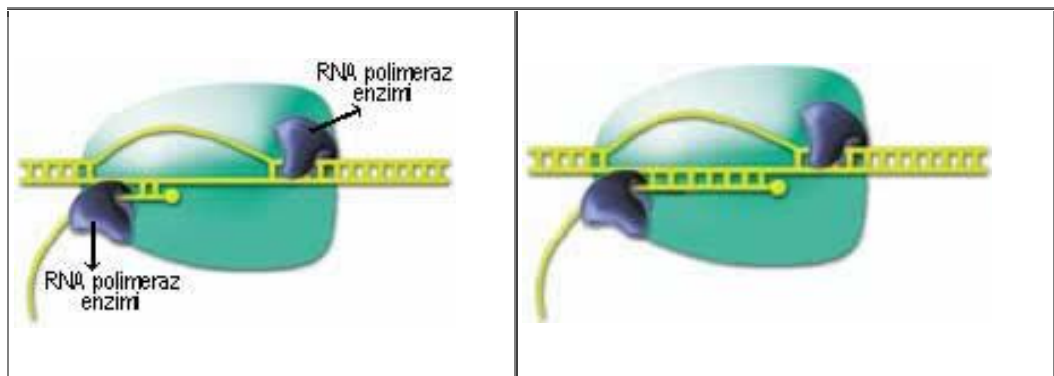
Vücudumuzda görev yapan bütün proteinlere ait bilgiler hücre çekirdeğinde yer alan DNA molekülünde depolanır. Yani bir protein üretileceği zaman, bu proteinle ilgili bilgiler DNA'dan alınır. Ancak bunun için DNA'nın, ihtiyaç olan protein hakkındaki bilgiyi tam ve doğru olarak anlaması ve doğru bilgiyi vermesi gerekir. Tıpkı üretim yapacak olan bir kimyagerin, bu üretim sırasında kendisine gerekli olacak hammaddeleri ve üretimi yapmak için ihtiyaç duyduğu tüm teknik bilgileri yetkili bir yere başvurarak talep etmesi gibi... Bir kimyager bunu karşısındaki kişi veya kurumdan yazılı veya sözlü olarak talep eder; işte DNA'dan da bir proteinin formülünü talep etmek için özel bir lisan kullanılır. Bu lisan 4 harften oluşan bir alfabeye sahiptir.

DNA molekülü 4 farklı nükleotidin farklı sıralamalarla artarda gelerek dizilmesinden oluşur. Bu dört farklı nükleotid sahip oldukları bazı moleküllerinin adlarıyla anılırlar; A (Adenin), G (Guanin), C (Citosin) ve T (Timin). Bu moleküllerin sıralamaları canlının kullanacağı bütün proteinlerin yapısının nasıl olması gerektiğine dair bilgiyi oluşturur. Yani her insanın hücrelerindeki DNA'da kendisine ait her özelliği meydana getiren proteinlerin bilgisi, 4 harfli özel bir alfabe ile yazılmıştır ve bu bilgiler bir kütüphane dolusu ansiklopediye sığacak kadar çoktur.

Milimetrenin binde birinden daha küçük bir alanda böyle ciltlerce ansiklopediye sığacak bilginin şifrelenmesi olağanüstü bir durumdur. Bu bilgi, yazılı hale getirildiğinde, 500'er sayfalık 1000 ansiklopedi uzunluğunda olacaktır, ki bu büyüklükte bir eser henüz yazılmamıştır. Günümüzde bilgi saklanması için çok yüksek kapasiteli bilgisayar çipleri tasarlanmıştır. Halen farklı şifreleme sistemleriyle bu kapasiteyi artırmak için çok yüksek maliyetli çalışmalar yapılmaktadır.

Ancak DNA molekülünde protein bilgilerinin şifrelenmesi, yeryüzünde üretilmiş hiçbir teknoloji ile kıyaslanmayacak kadar üstün bir kapasiteye sahiptir. Öyle ki, kapladığı alanda maksimum bilgiyi şifreleme kapasitesine sahiptir.

Hücre içindeki işlerin aksamaması, ihtiyacın doğru karşılanması, kısacası hücre yaşamının devam edebilmesi için doğru proteinin üretilmesi çok önemlidir. Bu yüzden hangi proteinin üretilmesi gerektiği ile ilgili mesaj alındıktan sonra DNA'dan doğru bilginin seçilerek alınması gereklidir. Peki bu seçimi kim yapacaktır ?



Şekil 1. 2. DNA'dan polimeraz enzimi ile bilginin alınması

Bir protein üretileceği zaman, RNA polimeraz isimli enzim, DNA'dan, üretilecek protein için gerekli olan bilgileri seçer ve kopyalar.

Hayati bir öneme sahip bu üretimi yapmak için gereken hammadde, yıllar süren eğitim hayatından sonra, yıllar süren bir bilimsel geçmişe sahip, deneyimli, akıllı, tecrübeli, görebilen, duyabilen bir bilim adamı değil, şüursuz atomların birleşiminden oluşmuş bir moleküldür. Bu hayati önem taşıyan seçme işlemini yapan, yine mükemmel bir yapıya sahip bir protein olan RNA polimeraz enzimidir. Bu enzimin yaptığı iş son derece zordur. Her şeyden önce, 3 milyar harften oluşan DNA molekülünün içinden, üretilecek proteinle ilgili gerekli harfleri seçip alması gerekmektedir. Polimeraz enziminin 3 milyar harften oluşan DNA molekülünün içinden, birkaç satırlık bir bilgiyi bulup çıkarması, 1000 ciltlik bir ansiklopedinin herhangi bir sayfasına saklanmış, birkaç satırlık özel bir yazıyı hiçbir tarif olmadan o anda bulmaya benzer.

Sipariş kopyalanır :

Sipariş fişinin, yani DNA'dan alınan bilginin doğru olarak kopyalanması çok önemlidir. Çünkü üretim boyunca kullanılacak bütün bilgiler bu sipariş fişi üzerinden okunur. Bu fişin kopyalanması sırasında meydana gelebilecek tek bir hata dahi canlı için ölümcül olabilir. Örneğin kanda dokulara oksijen taşımakla görevli olan hemoglobin proteininin amino asit diziliminde bulunan 600 amino asit içinden tek bir tanesinin yerinin değişmesi, hemoglobinin apayrı bir yapıya sahip olmasına ve görevini yapamamasına neden olur. Bu şekildeki bozuk hemoglobinler oksijen taşıyamadıkları için Akdeniz anemisi olarak bildiğimiz ölümcül hastalık ortaya çıkar.

Yanlış Kopyalama Kanseri Sebebi :

Son zamanlarda kanser araştırmacıları kanser oluşumunda hücre içinde yanlış üretilen proteinlerin önemli rol oynadığını ortaya çıkardılar. DNA'nın kopyalanması sırasında yanlış kopyalanan genler yanlış proteinlerin üretilmesine neden olurlar. Bu buluş ilk defa mesane kanseri araştırmalarında ortaya çıktı. Bu bölgedeki hücrelerde DNA'da bulunan 5000 basamaklı özel bir genin tek bir basamağında yanlış kodlanmasının hücreyi bozduğu anlaşıldı. Bu hatalı genlerin üreteceği proteinler yanlış bilgilerle üretildikleri için, hücreye yarar sağlayacağına zarar vermektedirler.

Kopyalama işleminin başlaması için çok önemli bir engel aşılmalıdır. DNA molekülünün merdiven gibi birbirine dolanmış kollarının kopyalama işlemi için ayrılması gerekir. Bu ayrılma işleminde yine RNA polimeraz enzimi iş başındadır. RNA polimeraz, kodlanacak genin

başlangıcından 35 harf öncesine bağlanarak, sarılmış merdiven gibi olan DNA'nın basamaklarını bir fermuarı açar gibi açar. Bu açılma çok hızlı yapılır. Öyle ki, bu hızdan dolayı DNA'nın ısınıp yanma tehlikesi oluşur. Ama sistem öylesine mükemmel düzenlenmiştir ki, bu tehlike de düşünülmüştür. Önceden alınan bir dizi tedbir sayesinde yanma tehlikesi ortadan kaldırılır; özel bir enzim sanki oluşabilecek tehlikenin farkındaymış gibi, gidip DNA'nın açılmış olan sarmalının iki ucunu tutarak bu sürtünmeye izin vermez. Ve yine özel enzimler DNA'nın açılması sırasında birbirine dolaşmasını önlerler. Bu enzimler olmasa "mesajcı RNA" olarak adlandırılan sipariş fişinin kopyalanması mümkün olmaz. Çünkü fermuar gibi açılan DNA sarmalının kolları kopyalama işlemi başlamadan tekrar birbirine dolanır ve sürtünmeden dolayı DNA'nın yapısı bozulur.

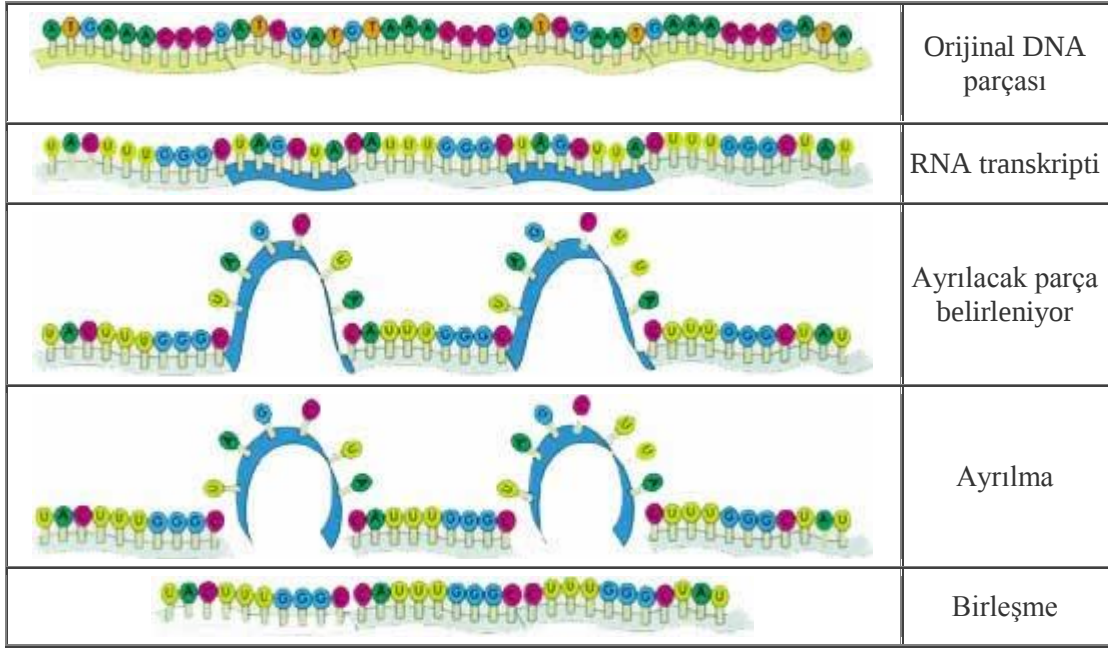
Burada bir kez daha hatırlatmakta yarar vardır: Bu bilgileri okurken, bunları yapanların belirli sayıda atomun birleşmesinden meydana gelmiş olan şuursuz moleküller olduğunu unutmamak gerekir. Enzim de protein de bu moleküllerdir.

Alınan bu özel tedbirlerden sonra aşılması gereken birkaç engel daha vardır. Örneğin istenilen proteinin amino asit dizilimini içeren bilgi büyük DNA molekülünün herhangi bir bölgesinde bulunabilir. Bu durumda farklı yerlerde bulunan bilgileri, yani amino asit dizilimini işaret eden şifreleri kopyalamak için polimeraz enzimi ne yapacaktır ? DNA'yı koparamaz, istemediği şifrelerin üzerinden atlayamaz. Doğrudan aynı hat üzerinde devam ettiğinde gereksiz bilgileri de kopyalayacak ve istenilen protein oluşmayacaktır.

Bu sorunun çözümü için olağanüstü şuurlu bir olay daha gerçekleşir ve DNA kopyalama işlemine yardım etmesi gerektiğini düşünmüş gibi, bükülerek, istenmeyen şifre dizisinin olduğu bölümü dışarı doğru kıvrır. Böylece ardı ardına okunması gereken, ama arada başka şifreler de olduğu için birbirlerinden uzak kalan şifre dizilerinin uçları birbirleri ile birleşir. Böylece kopyalanması gereken şifreler tek bir hat üzerine gelmiş olur. Bu şekilde polimeraz enzimi sipariş fişini üretilcek protein için kolayca kopyalayabilir.

Kimi zaman istenmeyen şifrelerin elenmesi için bir başka yöntem daha kullanılır. Bu yöntemde ise, RNA polimeraz enzimi, gereksiz şifreler de olmak üzere geni baştan sona kopyalar. Daha sonra, olay yerine gelen "spliceosome" enzimleri, gereksiz şifreleri bir halka şeklinde "büker ve atarlar". Bunun gerçekleşmesi için bu enzimlerin, ellerindeki reçete ile DNA'dan kopyalanan bilgileri kıyaslayıp, gerekmeyenleri tespit etmeleri gerekir. Sizin elinize harflerle dolu upuzun iki tane liste verilse ve bunların arasında gereksizleri tespit edip atmanız istense, bunun için ikisini dikkatlice inceleyip, satır satır kontrol etmeniz gerekir. Bunun için hem harfleri

tanımanız, hem istenen bilginin ne olduğunu bilmeniz, hem de neyi ne için yaptığının bilincinde olmanız gerekir. O nedenle herhangi bir biyoloji kitabında yada belgeselde "seçer, bükür ve atar" şeklinde bir cümle görmek insanları aldatmamalıdır. Çünkü burada kıyas yapan, tespit eden, inceleyen, ayıran, seçen, bükten ve atan karbon, asit ve fosfat gibi cansız ve şuursuz maddelerin bir araya gelmesiyle oluşan beyni olmayan, görmeyen, duymayan maddelerdir.



Şekil 1. 3. Gereksiz bilginin bükülüp atılması

Bazen bir proteine ait bilgi DNA'nın farklı yerlerinde olabilir. Böyle bir durumda RNA polimeraz enzimi geni baştan sona kopyalar. Daha sonra "spliceosome" isimli enzimler gelerek istenmeyen bölgenin iki ucunu birbirine değdirecek şekilde kopyalanan zinciri bükerekler. Bu işlemin sonucunda istenmeyen ve bükülen bölge zincirden kopartılıp atılır. Bu işlemlerin yerine getirilmesi için enzimlerin büyük bir şuur ve akıl göstermeleri gerekir. Ellerindeki reçetede ki milyonlarca harf içinden gerekli olanları hatasızca tespit edip ayıklayabilecek kadar dikkatli olmalıdırlar.

DNA'dan sipariş fişinin kopyalanması sırasında gerçekleşen şaşırtıcı ve olağanüstü olaylar bunlarla da bitmez. Kopyalamayı birilerinin durdurması gerekmektedir, aksi takdirde polimeraz enzimi, geni baştan sona kopyalar. Proteini kodlayan genin sonunda, o genin bittiğini gösteren bir kodon vardır. (DNA'daki şifreyi oluşturan nükleotidlerin her üçlü grubuna kodon denir.) RNA polimeraz durdurucu kodona geldiğinde, kopyalama işlemini durdurması gerektiğini anlar ve üzerinde protein için gerekli mesajı taşıyan mesajcı RNA ile DNA'dan ayrılır. Ancak bu noktada yine çok dikkatli davranılır. Çünkü mesajcı RNA hücre çekirdeğinden çıkıp, üretimin

yapılacağı ribozoma gidene kadar bir hayli yol kat edecektir. Bu esnada üzerindeki mesajın hiçbir zarar görmemesi gerekir. Bu nedenle, hücre çekirdeğinden bazı özel enzimlerin koruması altında çıkar.

Kopyalanan Bilgi Üretim Merkezine Ulaştırılır :

Hücrede protein üretimi için gerekli olan bilginin DNA'da bulunmasından ve kopyalanmasından sonra şimdi de bu bilginin proteinin üretileceği fabrika olan ribozomlara ulaştırılması gereklidir. Her hücrede bulunan bu organeller çekirdekteki DNA'dan oldukça uzakta ve hücrenin bütün sitoplazmasına (hücre içi sıvısına) dağılmış haldedirler. Bu fabrikalara üretim siparişleri eksiksiz bir biçimde süratle ulaştırılmalıdır. Mesajcı RNA (mRNA), yolunu şaşırmadan ve hücrenin içinde bulunan birçok organel ve molekül arasında hiç tereddüt etmeden ribozomu bulur. mRNA ribozomu bulduğunda onun dış kısmına bir hat şeklinde yerleşir. Bu şekilde artık üretilmek istenilen proteinin amino asit dizilimine ait bilgi üretim merkezine doğru biçimde ulaşmıştır. Bir proteinin üretilmesi için kopyalanan mRNA ne yapmaları gerektiği, ne zaman başlayıp ne zaman bitmesi gerektiği bilgilerini de taşır. Dolayısıyla bu talimat ribozoma ulaştığında, üretilcek protein için gerekli olan amino asitlerin ribozoma getirilmesi için hücrenin diğer bölgelerine mesajlar gönderilmeye başlanır.

Hammaddeler Üretim Merkezine Doğru Yola Çıkar :

İşte protein üretimindeki mükemmel organizasyon mucizelerinden biri de bu noktada gerçekleşir. Proteinin bilgilerini taşıyan mesajcı RNA ribozoma yerleştikten sonra başka bir RNA türü olan "taşıyıcı RNA" (tRNA) devreye girer. Bu RNA molekülü de DNA'daki bilgilere göre özel olarak üretilir. Bu RNA'lar sadece protein üretiminde hammadde olarak kullanılacak amino asitleri ribozoma taşımak üzere görevlendirildikleri için taşıyıcı olarak adlandırılırlar. Bu RNA'lar, bir fabrikadaki üretime hammadde taşıyan özel nakliye araçları gibidir. Ama bu özel taşıyıcı RNA'ların nakliye sisteminde çok farklı bir özellik vardır.

Daha önce de bahsedildiği gibi, her canlı hücresinde 20 çeşit amino asit vardır. İşte bu 20 çeşit amino asidin, yani hammaddenin her biri kendisine özel bir nakliye aracı tarafından taşınır. Amino asitlerin kendilerini taşıyacak olan tRNA'ya bağlanmaları da bir seri karmaşık işlem sonucunda gerçekleşir. Her amino asit çeşidini aktive eden özel bir enzim vardır. Aynı enzim hem amino asidi aktifleştirir hem de amino asidin tRNA'ya bağlanmasını sağlar. Buna göre enzimin (amino asit sentetaz) hem amino aside, hem de tRNA'ya bağlanabilecek yapısının olması gerekir. Görüldüğü gibi, her aşamada iç içe geçmiş birçok işlem ve görevi olan birçok parça bulunmaktadır. Bunlardan tek bir tanesi dahi olmadığında, canlının yaşamı devam

edemeyecek kadar hasar görebilmektedir. Örneğin amino asitleri aktive eden ve tRNA'ya bağlayan bu özel enzimler olmadığı takdirde protein sentezi için gerekli olan amino asitler ribozomlara ulaştırılamayacaktır.

Ribozoma tRNA tarafından getirilen her amino asit, mRNA'nın belirlediği üretim hattında belirli yerlerde işlenmelidir. Üretim boyunca tek bir amino asidin bile yanlış bir birimde işlenmesi proteini işe yaramaz bir molekül haline getirmeye yeterlidir. Oysa bu işlem bütün canlı hücrelerde kusursuz bir biçimde işler. Nakliye görevini yapan her tRNA, getirdiği her amino asidi üretim talimatında belirtilen yere götürür ve üretimdeki işleyişin bozulmamasını sağlar. Üretim talimatı ise bilindiği gibi mRNA'da kayıtlıdır.

Üretimden Önce Yapılması Gereken Tercüme :

Artık sipariş, yani üretilecek proteine ait bilgi ve gerekli hammaddeler hazırdır. Sipariş fişi üretimde bir hat boyunca yer alacak bütün makinelere iletilmiştir. Ortada aşılması gereken bir problem daha vardır. Üretim bilgisi, yani sipariş, daha önce bahsettiğimiz şekilde DNA'da özel bir dilde yazılmıştır. Ve üretim özel bir dilde yazılan bu bilgiye göre yapılmalıdır. Fakat hammadde olarak kullanılan amino asitlerin dizilimleri başka bir dildedir. Bu problemi şöyle ifade edebiliriz. Sipariş fişindeki yazılı emir, DNA'yı oluşturan şifrenin dilidir, yani 4 harfli bir alfabeden oluşan özel bir dilde yazılmıştır. Üretilecek olan proteinlerin dili de 20 harfli bir alfabeden oluşan bir başka dildir. (proteinleri oluşturan amino asitler 20 çeşit olduğu için) İşte bu lisanın farklılığı gibi, DNA'dan gelen üretim bilgisi amino asitlerin anlayacağı dilden değildir. Sonuç olarak, DNA'dan gelen bilgiye hangi amino asidin denk geldiğini anlayabilmek için, DNA'daki dilin diğerine tercüme edilmesi gerekir.

Ribozom fabrikası yaşamın sağlıklı biçimde devam etmesi için bu problemi en mükemmel şekilde çözen bir mekanizmayla donatılmıştır. Çözüm olarak üretim sırasında fabrikada yani ribozomda farklı iki dil arasındaki tercümeyi yapan bir tercüme sistemi yaratılmıştır. Kodon-antikodon metodu olarak adlandırılan bu tercüme sistemi şu andaki en gelişmiş bilgisayar merkezlerinden çok daha üstün bir şekilde, adeta bu iki dilde uzmanlaşmış bir tercüman gibi çalışır. DNA'nın özel lisanı ile yazılmış olan dört harfli protein bilgilerini 20 harften oluşan protein diline çevirir. Böylece hangi amino asitlerin yan yana dizileceğini ifade etmiş olur.

Sonuçta da istenilen proteinin doğru bir şekilde üretilmesini sağlar. İlerleyen satırlarda detaylarına yer vereceğimiz bu çeviri işlemindeki hatasızlık kuşkusuz çok dikkate değerdir. Bir hücrenin, dolayısıyla canlıların yaşaması için gerekli binlerce proteinin üretilmesinde ancak bir veya iki yanlışlığa yer olabilir.

Kodon-Antikodon Yani Anahtar-Kilit Metodu :

Bu metod sayesinde tercüme sistemi amino asitleri birleştiren üretim merkezinin hiç hata yapmamasını sağlar. Ribozomdaki birleştirme merkezine önceden yerleşip sipariş bilgisini taşıyan mRNA ile bir ucunda amino asit taşıyan tRNA anahtar-kilit gibi karşı karşıya gelirler. mRNA'daki her üç harf bir "kodon" yani bir kilit sayılır. tRNA'nın üç boyutlu şekli artı işaretine benzer. Bu artı şeklindeki yapının üst ucuna taşıdığı amino asit bağlıdır. Taşıyıcı RNA'nın bu kilidini açabilecek özellikte olan alt ucu da bir antikodon yani bir anahtar olarak tam karşısına geçer. Ribozomun üretim için kullandığı bu özel tercüme sistemi sayesinde proteinler kusursuz biçimde bir zincir gibi üretilir. Tercüme sisteminin bu metotla beraber en iyi şekilde çalışabilmesi için ribozom, her biri beraberce uyum içinde çalışan yüzün üzerinde yardımcı molekül kullanır. Bu moleküller üretim yerine gönderilen özel RNA'lardır ve bunların çoğu özelleşmiş proteinlerdir. Bu RNA'lardan en önemlileri mesajcı RNA'nın ribozoma getirdiği üretim bilgisinin taşıyıcı RNA tarafından anlaşılmasını ve diğer bir dilde okunmasını sağlayan "ribozomal RNA"dır. Hazırlanan bu mekanizmaların her biri tercüme işleminin hatasız yapılması ve sonucunda doğru proteinin üretilmesi için kusursuz bir biçimde çalışırlar.

Kalite Kontrol :

Daha önce de belirttiğimiz gibi, hücrelerin ihtiyaç duyduğu proteinlerde en ufak bir hata olduğunda hücre içindeki birçok mekanizma çalışamaz hale gelir. Bu şekilde de hücre yaşamını devam ettiremez ve hatta birçok durumda canlının kendisinde ciddi hastalıklar meydana gelir. Bugün birçok hastalığın kalıtsal nedenlerden kaynaklandığı bilinmektedir ve bunun nedeni bu aşamalardan birinde meydana gelen hatalardır. Hücre ve proteinler ise, sanki bu işlemlerin canlı için önemini biliyorlarmış gibi, son derece titiz davranırlar ve sentez sırasında belirli aşamalarda onları tekrar tekrar kontrol ederler.

Tek bir proteinin üretimi sırasında yapılması gereken kontrol işlemi için birçok enzim çalışır. Bu enzimler bir fabrikanın kalite kontrol departmanı gibidir. Çünkü her enzim, ürün hakkında çok detaylı bilgiye sahip olmalı ve üretimin her aşamasından haberdar olmalıdır. Aksi takdirde ortaya çıkan ürünü gereği gibi kontrol edemez. İlginç olan ise, üretilen proteinin kalitesini kontrol edenlerin yine proteinler olmasıdır.

Sipariş Yerine Teslim Edilir :

Bütün bu kontroller tamamlandıktan sonra artık protein kullanıma hazırdır. Proteinler kullanılacağı yere doğru yola çıkacaktır.

Üretimin bu aşamasına kadar olan tasarım mühendisliği protein kullanım yerine ulaşana kadar devam eder. Üretilen çok kıymetli protein molekülleri hiç zarar görmeden kullanım yerine kadar götürülmelidir. Ama nasıl ?

Bu sorunun cevabı hala tam olarak anlaşılmamıştır. Fakat bilindiği kadarıyla bu süreç insanı hayrete düşürecek derecede komplekstir

Hücre içinde üretilen proteinler, üretilip oldukları yerde bırakılmazlar. Aksi takdirde sürekli üretim yapan, ancak ürettikleri atıl kalan bir sistem oluşurdu. Ancak canlılıktaki tüm diğer sistemlerde olduğu gibi protein üretiminde de eksiksizlik ve kusursuzluk vardır.

Sonuç olarak üretilen her protein, kullanılacağı veya kullanılacağı zamana kadar depolanacağı ilgili yere yine çok özel yöntemlerle taşınır. Örneğin hücre dışına gönderilecek proteinler, enerji üretmekten sorumlu organel olan mitokondride kullanılacak proteinler, çekirdekte kullanılacak proteinler hep farklı mekanizmalar kullanılarak yerlerine gönderilirler. Proteinlerin kullanım yerlerine taşınmalarında devreye giren bu özel mekanizmalar ve yollar proteinlerin "hedeflenme sistemleri" olarak anılır. Hangi proteinin nereye gideceğini bilmesi başlı başına bir mucize iken, gideceği yere göre ulaşım aracı belirlenmesi, paketlenmesi, ulaşım sırasında zarar görmemesi için enzimlerle desteklenmesi daha da şaşkınlık yaratan bir olaydır.

Zigottan itibaren bir canlının oluşması, organların,, sistemlerin yapılması ve çalışması proteinler sayesinde olmaktadır. Protein, alyuvarlarda O₂ ve CO₂ bağlayan hemoglobin ve mikroplara karşı vücudu koruyan antikorların yapısını oluşturur. Ayrıca kanda, hücre dışı sıvı ve hücre içi sıvıda belli bir yoğunluk oluşturarak ozmotik dengeyi sağlar. Kanda ve dokularda protein miktarının dengesi bozulursa vücudun dengesi bozulur. Canlılığın temeli olan ve biyolojik reaksiyonlarda katalizör olarak iş gören enzimlerin yapısını ve yine vücut organlarının düzenli çalışmasını sağlayan hormonların bir kısmının yapısını proteinler oluşturur. Bütün bunların yanında, zorunlu durumlarda proteinden enerji temin edilir. Eğer sürekli kullanılırsa üremi hastalığı ve birçok bozukluklar ortaya çıkar. Bir gr. Protein 5 kalori enerji verir.

Görüldüğü üzere proteinler canlılığın devamı için çok önemli bir yere sahiptir. Bu sebeple son yıllarda gelişen bilgisayar teknikleri ile beraber bu konuya olan eğilim artmıştır. Bu alanda bir çok çalışma artan bir hızla devam etmektedir.

Tezin Amacı : Bu tezde yaşam için bu kadar önemli yere sahip olan protein makromoleküllerin sekonder yapıları incelenmiştir. Sekonder yapılar incelenirken, bu yapıdaki merkez karbon atomları arası bağ açıları esas alınmıştır. Bu açılardan yola çıkılarak, proteinin sekonder açıları sınıflara ayrılmış ve bu açı sınıfları tahmin edilmeye çalışılmıştır. Bunun için yumuşak hesaplama tekniklerinden olan yapay sinir ağları kullanılmıştır. Bu amaçla 2. bölümde biyoinformatik bilimi ve çalışmaları tanımlanmıştır.

Üçüncü bölümde proteinlerin genel yapısı, işlevleri, çeşitleri, yapılarının çeşitleri ve peptit bağlarından bahsedilmiştir.

Dördüncü bölümde yapay sinir ağlarının genel yapısı ve özellikleri, yapay sinir ağlarının çeşitleri , öğrenme algoritmalarından bahsedilmiştir.

Beşinci bölümde proteinlerin sekonder yapılarının tahmini üzerine geliştirilen diğer yöntemlere göz atılmıştır.

Altıncı bölümde ise geliştirilen yöntem ve sonuçları sunulmuştur.

Yedinci bölüm sonuç ve yorumlara ait olan bölümdür.

2. BİYOİNFORMATİK

2.1. Bioinformatik

1960'larda başlayan bilgisayar uygulamalarının biyolojide kullanılması girişimi, her iki alandaki teknolojik gelişime paralel olarak hızla ilerlemiş ve böylelikle ortaya çıkan “BİYOİNFORMATİK” dalı bugün en popüler akademik ve endüstriyel sektörlerin başına geçmiştir.

21 .yüzyılda biyoloji sadece laboratuarda yapılan bir bilim olmaktan çıkmış, aynı zamanda bilgi teknolojisine de dayanan bir bilim dalı haline gelmiştir. Bilgisayarların ve bilgisayar yöntemlerinin biyolojinin her alanında kullanılmaya başlanması günümüzde gitgide kaçınılmaz olmuş, biyoinformatik olarak adlandırılan yeni bir bilim dalı ortaya çıkmıştır.

Bilgisayarların moleküler biyolojide kullanımı üç boyutlu moleküler yapıların grafik temsili, moleküler dizilimler ve üç boyutlu moleküler yapı veritabanları oluşturulması ile başlamıştır. Kısa sürede çok yüksek miktarlarda 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 sonucunda, bu alandaki bilişim uygulamaları neredeyse takip edilemez bir hızda gelişmiştir. Biyoinformatik dalının ayrı bir bilim dalı olarak tanınması da son 10-15 yılda gerçekleşmiştir.

Biyoinformatik genel olarak biyolojik problemlerin çözümünde bilişim teknolojilerinin kullanılması olarak tanımlanabilir. En dar tanımı ile : 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 organizmanın DNA'sı 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 bilgiler, 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, ileride 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.

Bu alanı kısaca tanımlamanın bir yolu da, biyoinformatik araçların kullanıldığı genel araştırma konularını özetlemek olabilir :

Metodolojik çalışmalar :

- 1) DNA sıra ve dizilimi araştırmaları
- 2) Protein sıra ve dizilimi araştırmaları
- 3) Makromoleküler yapıların (DNA, RNA, protein) üç boyutlu dizilim araştırmaları
- 4) Küçük moleküllerin ligandlarıyla etkileşiminin araştırılması
- 5) Heterojen biyolojik veri tabanlarının entegrasyonu
- 6) Biyolojik enformasyonun paylaşımının kolaylaştırılması
- 7) Bilgisayar ile otomize edilmiş veri analizi ve iletimi
- 8) Etkileşimde bulunan gen ürünleri için bilgi ağları oluşturulması
- 9) Kimyasal reaksiyonlardan hücreler arası iletişime kadar pek çok biyolojik faaliyet sürecinin simülasyonu
- 10) Büyük çaplı biyolojik deneylerden (GENOM PROJESİ gibi) çıkan sonuçların analizi

Biyolojik çalışmalar :

- 1) Protein yapı ve fonksiyonunun belirlenmesi
- 2) Her hangi bir biyolojik fonksiyonu arttıran yada engelleyen küçük moleküllerin tasarlanması
- 3) Karmaşık genetik fonksiyon yada regülasyon (düzenleme) faaliyetlerinin tanımlanması
- 4) Tıbbi yada endüstriyel amaçlı yeni makromoleküller üretmek
- 5) Genetik faktörlerin, hastalık yatkınlığına etkilerini ortaya çıkarmak

Bu konudaki biyolojik çalışmalar moleküler biyolojinin ana paradigması üzerine kurulmuştur. DNA, nükleotidler, üçlü kodonlar, genler, amino asitler, proteinler gibi konular üzerine temellendirilmiştir. [19]

Bugün biyoinformatik alanında yapılan çalışmalar sadece yaygın kullanım bulan veritabanları ve algoritmalar geliştirilmesi ile sınırlı kalmamış, kısa denebilecek bir sürede hem bu alandaki araştırmalar hem de alansal eğitime yöne veren bazı temel prensiplerinde ortaya çıkmasına imkan vermiştir.

Bu prensipleri genel olarak özetlemek gerekirse :

- 1) Mutasyon, delesyon, insersiyon gibi evrimsel değişiklikler hep gerçekleşse de protein moleküllerinin yapısı neredeyse değişmez biçimde korunur. Amino asit dizileri çok farklı olsa da, yapısal benzerlik gösteren pek çok makromolekül bulunsa da, dizilim benzerliği taşımasına rağmen yapısal benzerlik göstermeyen iki molekül neredeyse yoktur.
- 2) Dinamik programlama temelli algoritmalar , biyolojik dizilimlerde en optimal sonuçları vermektedir. Matematiksel optimizasyon iki dizin arasında yüksek benzerlik durumunda daha güvenlidir.
- 3) Moleküller arası fiziksel etkileşimler, enerji denklemleri ile tanımlanabilir. Atomlar arasındaki temel fiziksel çekim ve itme kuvvetlerini göz önünde bulunduran metotlar sayesinde, sadece temel enerji denklemleri ile moleküllerin dinamik hareketlerinin modellenmesi mümkündür.
- 4) Proteinlerin biyolojik fonksiyonlarının pek çoğu amino asit dizini içinde rastlanan lineer parçalar üzerine kodludur. Toplam dizin analizinde büyük farklılıklar gösterse de benzer fonksiyonlara sahip tüm proteinlerde aynı motifi bulmak zordur. Diğer fonksiyonlar biyofiziksel ve biyokimyasal özellikler bağlı olarak oluşturulan üç boyutlu katlanmalar ile sağlanır. [5]
- 5) Dizin modellemede olasılıksal yaklaşımlar, birbirleriyle ilişkili dizinlerin sınıflandırılmasında ve bu sınıfların yeni üyelerinin belirlenmesinde çok etkilidir.
- 6) Üç boyutlu biyolojik yapıların diziliminde geometrik algoritmaların kullanılması etkin sonuçlar vermektedir. Bu metotta atomlar arasındaki mesafelerin yanı sıra, incelenen molekülün bilinen fonksiyonu da atomsal iletişiminin tarifi için göz önünde bulundurulmaktadır.
- 7) Biyolojik moleküllerin üç boyutlu yapısı belirlenirken, gerekli bilginin bir bölümü yaklaşık %70 doğruluk oranı ile lokal dizinden (20 amino asitlik sıralamalardan) elde edilebilir. Üç boyutlu yapının tamamen ortaya konabilmesi için gerekli bilgi dizin içinde mevcut olmakla birlikte molekül içinde uzun mesafeli etkileşimlerde göz önünde bulundurulmalıdır.
- 8) Yeni bir dizinin , bilinen bir yapıyı alabilme ihtimalinin araştırılması, hangi yapıyı aldığını araştırmaktan daha kolaydır.
- 9) Konak ve patojen DNA bilgilerindeki varyasyonlar ve varyasyon korrelansı, bir hastalıkta hangi proteinlerin rol aldığının ve nasıl bir etkileşime geçtiklerinin belirlenmesinde kullanılabilir.

3. BİYOKİMYA VE PROTEİNLER

3.1. Biyokimya

Biyokimya biyolojik materyallerin moleküler seviyede ve kimyasal prensipler çerçevesinde incelendiği bir temel bilim dalıdır. Esasında biyoloji ve kimya temel bilimlerinin bir çalışma alanı olan biyokimya, günümüzde başta tıp olmak üzere veterinerlik, eczacılık, diş hekimliği, ziraat, gıda teknolojisi ve mühendislik gibi uygulamalı sahalarda başvurulmuş bir bilim dalı haline gelmiştir. [2] Biyokimyada kendi içerisinde alt dallara ayrılmıştır. Moleküler biyoloji, biyokimya mühendisliği, klinik biyokimya, fizyolojik biyokimya, zirai biyokimya, gıda enzimolojisi, gen mühendisliği v.s. gibi alt dallar mevcuttur.

Bu bilim dalına olan ilgi canlılığın sırrını çözmek ile alakalı olmuştur. Günümüzde artık canlılığın devamını sağlayan bir takım kimyasal bileşikler tanımak, özelliklerini çözmeye çalışmak ve bunların birbirleriyle olan etkileşimlerinin sebeplerini ortaya koymaktır. Esasen bizi oluşturan cansız varlıklardır. Fakat bunlar incelenirse bütün fiziksel ve kimyasal kanunlara uydukları görülür.

Canlı varlıklar en basit olanından en karmaşığına doğru son derece karmaşık fakat buna rağmen bir o kadarda düzenli yapıya sahiptir. Canlılarda biyomoleküller olağanüstü bir çeşitlilik ve komplekslik arz eder. En küçük hücreler olan bakteriler bile binlerce farklı organik bileşik ihtiva ederler. Örneğin üzerinde biyokimyacıların en çok çalıştığı ve biyokimyasal olayların en fazla açıklığı kavuşturulan *Escherichia coli* bakterisi , yaklaşık 5000 farklı organik bileşiğe sahiptir. [2]

Günümüzde biyokimyacılar artık biyoloji ve tıp alanında en zor işlere el atmışlardır hatta o kadar ki insanın gen haritası çıkarılmaya ve bu yolla daha insan dünyaya gelmeden hastalıklarına müdahale edilme yoluna gidilmeye çalışılmaktadır.

3.2. Proteinler

Protein “ilk element” (Yunanca’da “proteios”) anlamına gelir. Proteinler tüm insanlar, bitkiler ve hayvanlardan en basit bakterilere kadar canlı hücrelerinin temel parçasıdır. Proteinler hücrelerdeki bütün biyokimyasal reaksiyonları yönlendiren olayların yapıtaşlarıdır. Hücreler içerisinde gerçekleşen olaylar, çok sayıda farklı proteinin vazifelerini yerine getirmeleri ile devam eder.

Mesela kanda bulunan hemoglobin proteini oksijen taşımacılığında vazife görmekte, antikor denilen proteinler vücudun savunma sisteminin temelini oluşturmakta, insülin hücreye glikoz alımını sağlamakta, keratin saç ve tırnak yapısını meydana getirmekte, enzim adı verilen

proteinler hücre içi kimyasal reaksiyonları mükemmel bir hız ve doğrulukta yerine getirmektedir. Proteinler vücudumuzun kuru ağırlığının yaklaşık % 50 sini teşkil etmektedirler.

Proteinler hücrenin yapıtaşlarıdır. Eğer hücreyi dev bir gökdelene benzetirsek, proteinler de bu gökdelenin tuğlaları sayılabilirler. Ancak tuğlalar gibi standart şekil ve yapıda değildirler. En basit hücrelerde bile 2000'e yakın farklı türde protein bulunur. Hücre bu çok farklı proteinlerin hepsinin olağanüstü bir uyum içinde çalışması sayesinde yaşar. Proteinleri gördükleri biyolojik fonksiyonlarına göre sınıflamaya tabi tutabiliriz.

3.2.1. Enzimler

Çok çeşitlilik gösteren ve en fazla özelleşmiş olan katalitik aktiviteye sahip protein molekülüdür. Hücre içindeki organik biyomoleküllerin kimyasal reaksiyonlarının hemen hemen hepsi enzimler tarafından katalize edilmektedir. Enzimler protein yapısındadırlar. Her organizma kendi enzimlerini aldığı amino asitlerden kendisi sentez eder.

Enzimler, tüm vücudumuzdaki her türlü işlevleri yapan organik maddelerdir. Protein yapısında olan enzimler, reaksiyonlara yardımcıdırlar. Ancak, aktif olabilmeleri için belirli bir ısı, bazı mineraller ve özgül pH isterler. Enzimler aktif duruma geldiklerinde iki grup oluştururlar. Apoenzim ve yardımcı enzim (koenzim = CoE). Enzimler etkin oldukları yer ve etki çeşitlerine göre iki ana grupta sınıflandırılabilir. Enzimlerin aktivasyonunu birçok etmen etkiler. Başlıcaları; ürün (feedback mekanizması = fazla yapımda geriye dönük kontrol mekanizması) ve enzim konsantrasyonu, reaksiyon süresi, ısı, pH, iyonlar ve ilaçlar ek olarak sayılabilir.

3.2.2. Transport Proteinleri

Kan plazmasında bulunan transport proteinleri bazı özel molekül ve iyonlara bağlanarak bunları bir organdan diğerine taşımaktadır. Örneğin hemoglobin akciğerden geçerken oksijeni geçici olarak yüzeyine bağlar ve dolaşımla periferel dokulara taşır. Bu dokuda serbest bırakılan oksijen ise besin oksidasyonunda kullanılarak enerji elde edilir. Serum proteinleri, pek çok moleküle ve özellikle serbest yağ asitlerine bağlanarak bunları adipoz dokulara ve vücudun diğer kısımlarına taşımaktadır. İlaçların hedefleri bazı istisnalar dışında, resptörler, enzimler ve transport proteinleridir. [1]

3.2.3. Lipoproteinler

Lipoproteinlerin genel işlevi, suda çözünmeyen lipidlerin, çözünür lipid ve protein kompleksi şeklinde kanda taşınmalarını sağlamaktır. Lipidler trigliserid, kolesterol esterleri serbest kolesterol ve fosfolipidleri içerir. Apolipoprotein denilen yaklaşık on değişik protein partikülleri çeşitli lipoproteinlerle ilişkilidir ve harflerle "A" , "B" , "C" vb. adlandırılırlar.

Lipoproteinler ayrıca yağda eriyen vitaminler (A,D ve E), ilaçlar (probukol, sklosporin vb.), bazı virüsler ve bazı antioksidan enzimleri (paraoksonaz, trombosit kökenli aktive edici faktör hidrolaz) gibi bir çok maddeyi taşırlar.

Lipoproteinler, genellikle çekirdeği çoğunlukla hidrofobik lipidler, (trigliserid ve kolesterol esterleri) içeren protein yüzey tabakası, serbest kolesterol ve fosfolipidler (daha hidrofilik yapı taşları) taşıyan küresel parçacıklardır. Altı majör lipoprotein sınıfı lipid taşınmasında farklı roller oynarlar. Yüzeydeki özgün apolipoproteinler farklı lipoproteinlerin kaderini belirler. Bu nedenle, lipoprotein metabolizması ve lipid anormallikleri ile ilgili hastalıkların anlaşılması için, lipid metabolizmasını düzenleyen her bir apolipoproteinin rolünü incelemek gerekir.

3.2.4. Besin ve Depo Proteinleri

Bitkilerin tohumları, bitki embriyosunun büyümesi için besin proteinleri depo etmektedir. Tohumlarında besin proteini depolayan bitkilere örnek olarak buğday, mısır ve pirinci gösterebiliriz. Yumurta akında bulunan başlıca protein ovalbumin, buğday ununda bulunan gliadin, mısır ununda bulunan zein, sütte bulunan ise kazeindir. Bazı proteinler ise diğer maddelerle birleşerek onların depo edilmesine yardımcı olmaktadır. Örneğin demir, dalakta ferritin adı verilen bir protein tarafından depo edilmektedir.

3.2.5. Kontraktil Sistemde Bulunan Proteinler

Bazı hücre ve organizmalarda yer alan proteinler kontraksiyon yaparak şekil değiştirmeye ve hareket etmeye yardımcı olmaktadır. Omurgalılarda iskelet kası proteininin büyük bir kısmını kontraktil özelliğe sahip olan aktin ve miyozin oluşturmaktadır. Kas kontraksiyonu enerji gerektiren bir olay olup bu iki filamentli yapıya sahip proteinin birbiri üzerinde kayması ile meydana gelmektedir.

3.2.6. Yapısal Proteinler

Yapısal proteinler genellikle fibriller yapıda olup organizmaya desteklik yapmakta ve dayanıklılık sağlamaktadır. Örneğin tendon ve kıkırdak dokusunun en önemli yapısal proteini kolajen olup çok kuvvetli liflerden meydana gelmiştir. Deri, hemen hemen saf kalojenden oluşmuştur. Saçlar, tırnaklar ve tüyler hemen hemen hiç çözünürlüğü olmayan keratin proteininden meydana gelmiştir.

3.2.7. Savunma Proteinleri

Pek çok protein, organizmayı istila etmek isteyen diğer canlılara ve yaralanmalara karşı korumaktadır. Limfositler tarafından meydana getirilen ve omurgalıların özelleşmiş proteinleri olan immunoglobulinler yada antibodiler, canlıları istila etmek isteyen bakterileri, virüsleri veya diğer canlılara ait olan yabancı proteinleri tanıyarak çöktürmekte veya nötralize etmektedir. Fibrinojen ve thrombin kanı pıhtılaştıran proteinler olup yaralanma sonucu kan kaybını önlemektedir.

3.2.8. Regülasyonu Sağlayan Proteinler

Bazı proteinler hücrel ve fizyolojik aktiviteleri düzenlemektedir. Bunlar arasında hormonları sayabiliriz. Örneğin insülin hormonu şeker metabolizmasını düzenlemektedir. Eksikliği halinde şeker hastalığı ortaya çıkmaktadır.

Adenohipofizden salgılanan büyüme hormonu ve parathroidlerden salgılanan parathroid hormonu Ca^{+2} ve fosfat (P) transportunu düzenlemektedir. Diğer bir düzenleyici protein olan repressörler bakteri hücrelerinde enzimlerin sentezini düzenlemektedir.

3.3. Amino Asitler

Proteinler de kendilerinden çok daha küçük parçalardan oluşur. Bu parçalar, "amino asit" adı verilen ve karbon, azot, hidrojen gibi atomların farklı şekillerde birleşmesiyle oluşan moleküllerdir. Ortalama bir proteinde 500-1000 kadar amino asit vardır. Bazı proteinler çok daha büyüktür.

İşin en önemli yanı ise, amino asitlerin bir proteini oluşturmak için mutlaka belirli bir sıra içinde dizilmeleri zorunluluğudur. Canlı bedenlerinde kullanılan 20 farklı türde amino asit vardır. Bu amino asitler protein oluşturmak için birbirlerine gelişigüzel bağlanmazlar. Aksine, her proteinin belirli bir amino asit dizilimi vardır ve bu dizilimin harfiyen tutturulması gerekir. Protein yapısındaki tek bir amino asidin bile eksilmesi veya yerinin değişmesi, o proteini işe yaramaz bir molekül yığını haline getirir. Bu nedenle her amino asit, tam gereken yerde, tam gereken sırada yer almalıdır. [3]

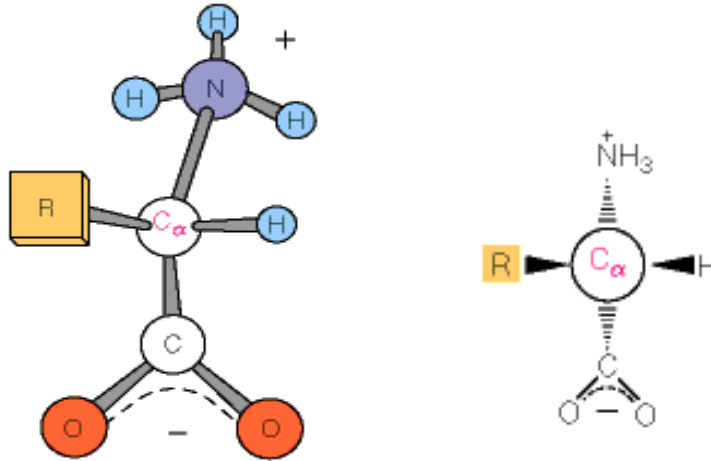
Canlı hücresinde bu dizilimin bilgisi DNA'da saklanır ve proteinler de DNA'daki bu bilgi okunarak üretilir. Genetik bilginin doğru bir şekilde okunması ve tercüme edilmesi ile amino asitlerin sırası tayin edilmekte ve sonunda proteinlerin arzu edilen biyolojik aktiviteleri ortaya çıkmaktadır.

Genetik bir bozukluk nedeni ile enzimlerde veya diğer proteinlerdeki amino asitlerden bir tanesinin yerine bir başka amino asit girdiği zaman ortaya çıkan protein molekülü biyolojik bakımdan “defektli” bir molekül olup fonksiyonunu tam olarak yapamaz.

Amino asitler proteinlere göre daha küçük moleküller olmalarına rağmen, oldukça kompleks bir yapıları vardır. Amino asitleri oluşturan atomlar üç ayrı grup halinde bulunurlar; amino grubu, karboksil grubu ve yan zincir grubu (yada radikal grup). Bütün amino asitlerde, amino ve karboksil grupları aynıdır. Bir amino asidi diğerlerinden farklı kılan tek özellik, moleküle bir ucundan bağlanan yan zincir grubudur. Bu yan zincir gruplarının her amino asitte farklı olması sayesinde her amino asit birbirinden çok farklı özelliklere sahip olur. [3]

Nasıl ki bir makinenin yapısında çeşitli malzemeler kullanılmaktaysa, vücudumuzdaki çok karmaşık görevleri yerine getirebilmesi için protein makinelerinde de farklı özelliklere sahip malzemeler bulunmalıdır.

İşte amino asitlerin yan zincir gruplarındaki atomların şekli, sayısı ve sıralamaları, elektrik yükleri, hidrojen bağı kapasitelerinin farklı farklı olması, amino asitlere çeşitlilik kazandırır ve bu çok çeşitli malzemeden de yine çok çeşitli protein makineleri üretilir. Örneğin yan zincir gruplarının (+) veya (-) elektrik yükünün olması veya yüksüz olması amino asit molekülünün suda eriyip erimemesini sağlar. Bu şekilde farklı özelliklere sahip olan amino asitlerin farklı dizilimlerle yan yana gelmeleri, proteinlerin vücut içinde hayret verecek derecede çeşitli görevleri yerine getirebilmelerini sağlar. Aşağıdaki şekilde bir amino asidin genel yapısı verilmiştir.



Şekil 3.1. Amino asit genel yapısı

Yukarıdaki şekilde R olarak isimlendirilen yan zincir kısmının farklı olması sayesinde amino asitler birbirinden farklı özelliklere sahip olurlar.

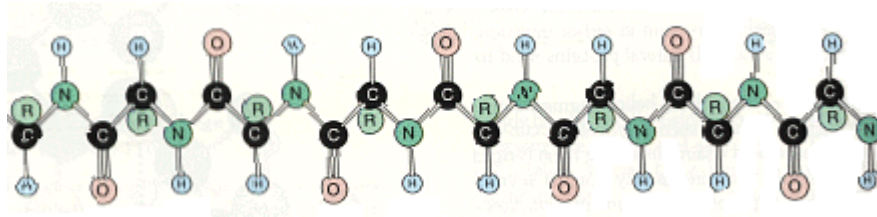
3.4. Proteinlerin Yapıları

Proteinler yapıları itibariyle 4 farklı düzeyde ele alınabilirler.

- 1- Birincil Yapı (Primer)
- 2- İkincil Yapı (Sekonder)
- 3- Üçüncül yapı (Tersiyer)
- 4- Dördüncül yapı (Kuaterner)

3.4.1. Primer Yapı

Bir proteinin primer yapısı yalnızca amino asit moleküllerinin yan yana gelip zincir oluşturmalarından ibarettir. Amino asitler arası peptit bağları mevcuttur. İki amino asidin birleşmesi ile dipeptit ikiden fazla amino asidin birleşmesinden oluşan yapıya ise polipeptit adı verilmektedir. Primer yapıyı oluşturan amino asitlerin cinsi ve sıralanış şekli, polipeptidin daha sonra kazanacağı üst yapısal organizasyonları belirlemektedir. Bu nedenle primer yapıdaki bozukluk, uygun konformasyonu ve dolayısıyla uygun biyolojik fonksiyonu kazanmamış bir proteinin ortaya çıkmasına neden olmaktadır. Primer yapı genetik olarak kontrol edilmekte olan tek yapı olup, bir proteinin sekonder, tersiyer ve kuaterner yapılarını belirlemektedir. [4]



Şekil 3.2. Primer yapı

3.4.2. Sekonder Yapı

Proteinlerin üç boyutlu yapılarına bakıldığında düz amino asit zincirleri olmadığı görülmektedir. Bu üç boyutlu konformasyonu sağlayan katlanmalardır. Bu katlanmalarda etkili olan bağlar kovalent bağlardan çok hidrojen bağları, tuz köprüleri, hidrofobik etkileşimlerdir. [4] primer yapıdaki bir proteinin katlanıp bu üç boyutlu oluşturduğu yapısına sekonder yapı denilir. Proteinlerin sekonder yapılarını kararlı hale getiren bir çok kuvvet vardır. Zayıf olmalarına rağmen çok fazla sayıda olan bu bağlar, proteinin yapısına önemli ölçüde esneklik ve aynı zamanda karalılık kazandırmaktadır.[4] Bu bağlar : hidrojen bağları, hidrofobik etkileşimler, elektrostatik etkileşimler, vander waals etkileşimleridir.

Hidrojen bağları : Ortaklanmamış elektron çifti taşıyan iki elektronegatif atom arasında oluşan bir hidrojen köprüsü hidrojen bağı olarak tanımlanabilir.[4] Dolayısıyla proteinlerdeki hidrojen bağları H atomları ile O atomları arasındaki çekimden kaynaklanır.

Hidrofobik Etkileşimler : Proteinin yapısında bulunan amino asit kalıntılarının polar olmayan R gruplarının birbirlerini çekmeleri sonucu oluşan zayıf bağlanmalardır. Daha çok protein molekülünün iç kısımlarında meydana gelen bu tür bağlanmalar molekülün yüzey alanını azaltarak ilgili grupların hidrojen bağı yapma şansını artırmaktadır. [4]

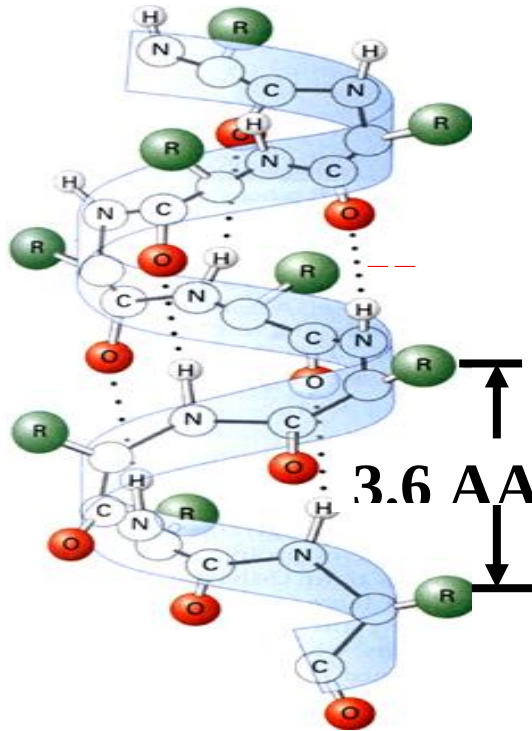
Elektrostatik Etkileşimler : Farklı yüklü iki grup arasındaki çekme kuvveti veya tuz bağları ile oluşan bağlardır. Aynı zamanda aynı yüklü grupların itme kuvvetleri de bu grup kuvvetler içine girmektedir.

Vander Waals Etkileşimler: Çok zayıf etkileşimlerdir. Birbirine yakın durumda bulunan atomların yörüngelerinde bulunan aynı yüke sahip dipollerin birbirini itmesi yada ters yüklü olan dipollerin birbirlerini çekmesi ile ortaya çıkan bağıdır.

Sekonder yapının 4 farklı çeşidi vardır.

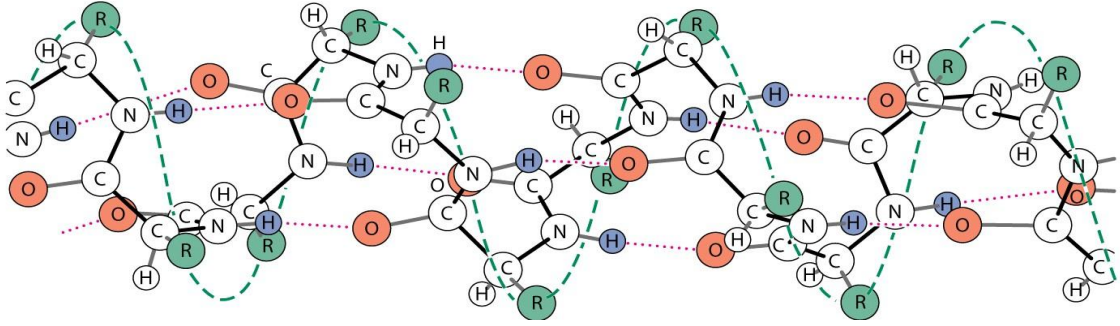
3.4.2.1. α -Heliks Yapı

Polipeptit zincirinin bir eksen etrafında sağa sola kıvrılmasıyla oluşan sarmal yapıya denir. Bu yapıda en çok etkili olan bağlar hidrojen ve oksijenler arası meydana gelen hidrojen bağlarıdır. [6]



Şekil 3.3. α - heliks yapı

Her amino asit ,heliks eksenini boyunca birbirinden yaklaşık 1.5 Å uzaklıkta bulunur. Her amino asit zincir üzerinde dönüş yaparken 100° açı yapar. Bu da yaklaşık 3.6 amino asit uzunluğunda 360° dönüş yani tam bir tur anlamına gelir. Heliksin tam bir dönüş yapmasıyla gidilen uzaklık ise $3.6 \times 1.5 = 5.4$ Å demektir. Primer yapıda aralarında 3-4 amino asitlik uzaklık bulunan amino asitler, α-heliks ekseninde birbirine en yakındır.



Şekil 3.4. Hidrojen bağları

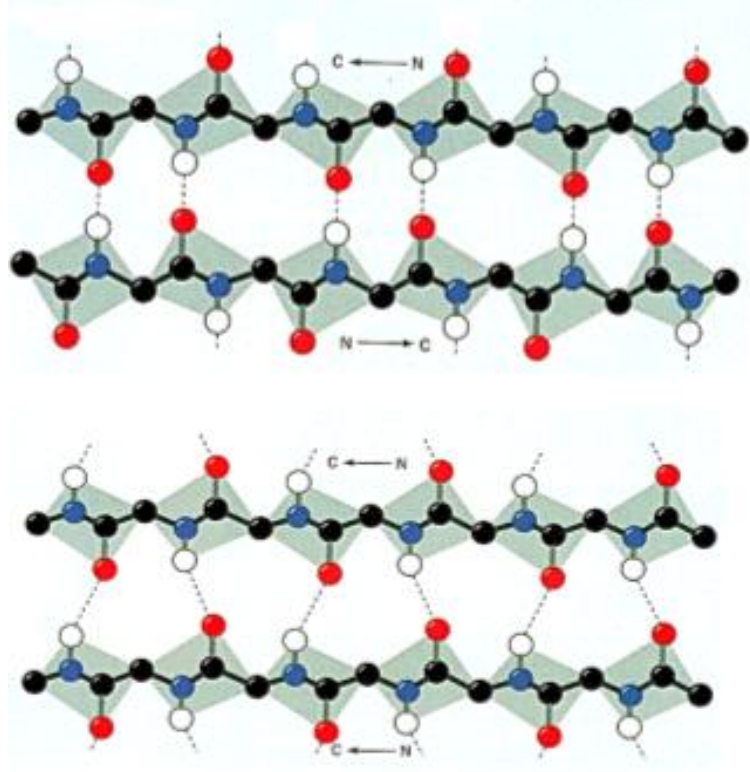
Bu yapı hidrojen bağlarının sayısının çokluğu nedeniyle en dayanıklı, en kararlı ve en düşük enerjiye sahip yapıdır. Enerjinin düşüklüğünün yapının kararlı olması ile alakasını ise şu şekilde izah edebiliriz: Proteinin yapısıyla ilişkili olan kararlılık terimi, doğal konformasyonu sürdürme eğilimi olarak tanımlanabilir.

Belirli bir polipeptit zincirinin teorik olarak sayısız farklı konformasyonu olduğu düşünülebilir ve bir proteinin katlanmamış hali, konformasyonel entropisinin en yüksek olduğu durum olarak nitelendirilebilir. Bu entropi ve polipeptit zincirindeki bir çok grubun çözücü (su) ile yaptığı hidrojen bağı etkileşimleri, katlanmamış yapıyı sürdürme eğilimindedir. Genelde en düşük serbest enerjili protein konformasyonu yani en kararlı konformasyon, en fazla sayıda zayıf etkileşimi içeren konformasyonlardan biridir. [5]

Yukarıdaki şekilde 1-4, 2-5, 3-6, Amino asit çiftlerinin H ve O atomları arasında bağ oluşur ve bu bağların çokluğundan dolayı bu sarmalın kararlılığı yüksektir.

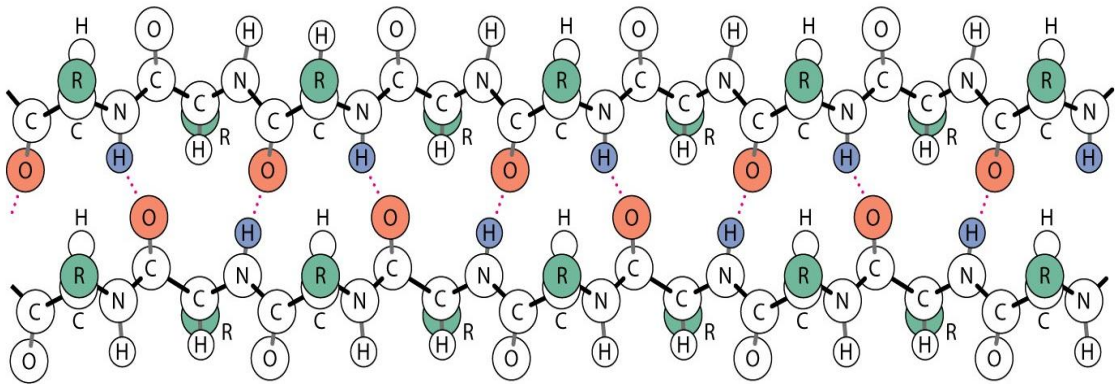
3.4.2.2. Beta Kırmalı Tabaka

Bu yapıda iki veya daha fazla sayıda polipeptit zinciri paralel yada anti paralel bir şekilde birleşerek bir tabaka oluşturur. Bunun neticesinde meydana gelen yapıda zincirler tabaka veya levha halindedir. R gurupları tabaka düzleminin altında yada üstünde yer alırlar. Polipeptit zinciri gergin durumdadır. Her amino asit, β-kırmalı tabaka boyunca birbirinden 3.5 Å uzaklıkta bulunur.



Şekil 3.5. Anti paralel ve paralel beta kırmalı yapı

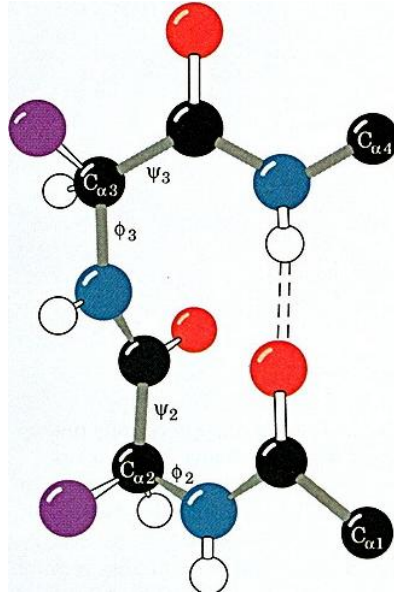
Beta kırmalı yapıda her amino asit, β -kırmalı tabaka boyunca birbirinden $3.5 \text{ \AA}$ uzaklıkta bulunur. Bu yapıda hidrojen bağları aşağıdaki şekilde olduğu gibi oluşmaktadır.



Şekil 3.6. Beta kırmalı yapıda hidrojen bağları

3.4.2.3. Beta Bendler Yapı

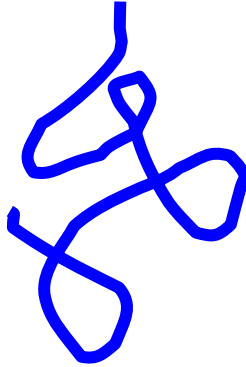
α heliks ve beta kırmalı yapıların birbirine bağlandığı yapıdır. β -bendler (menteşe bölgeleri), zincirin yönünü değiştirir. β -bendlerin varlığı, polipeptidlerin globüler kütleler oluşturmasını sağlar. β -bendler bölgelerindeki 1- 4 amino asit artıkları (residue) arasında H bağları oluşur. Pro ve Gly sıklıkla bulunur.



Şekil 3.7. Beta bendler yapı

3.4.2.4. Tesadüfi Kıvrılmalar

Proteinlerin, heliks, kırmalı tabaka veya β -bendler yapmayan bölgeleri, gelişi güzel helezonlar, kıvrılmalar şeklindedir. Düzlemler arasında belirli bir ilişki ve H bağları yoktur. Biyolojik fonksiyon bakımından, diğer sekonder yapılarla aynı öneme sahiptir.



Şekil 3.8. Tesadüfi kıvrılma yapısı

3.4.3. Tersiyer Yapı

Proteinlerin doğal şekillerinin boyutları bize sekonder katlanmalardan daha ileri katlanmalar olduğunu söylemektedir. Sekonder yapıyı oluşturan bölgelerin sayısı ve uzayda birbirlerine göre bulundukları konumlar proteinlerin daha ileri derecede katlanmalarını tanımlayan tersiyer yapıyı belirlemektedir. Biyolojik fonksiyonunu yerine getirebilmesi için polipeptidin almış olduğu son yapısal organizasyon tersiyer yapıdır. [4] Termodinamik olarak en dayanıklı şekil olmasına rağmen esnek yapıdadır. Tersiyer yapı, proteinin fonksiyonel karakterini belirler. Mevcut proteinlerin çok büyük kısmı, tersiyer yapıya sahiptir. Primer yapıda birbirinden uzakta bulunan amino asitler, tersiyer yapıda komşu olabilirler.

Bu yapıyı oluşturan bağlar ise :

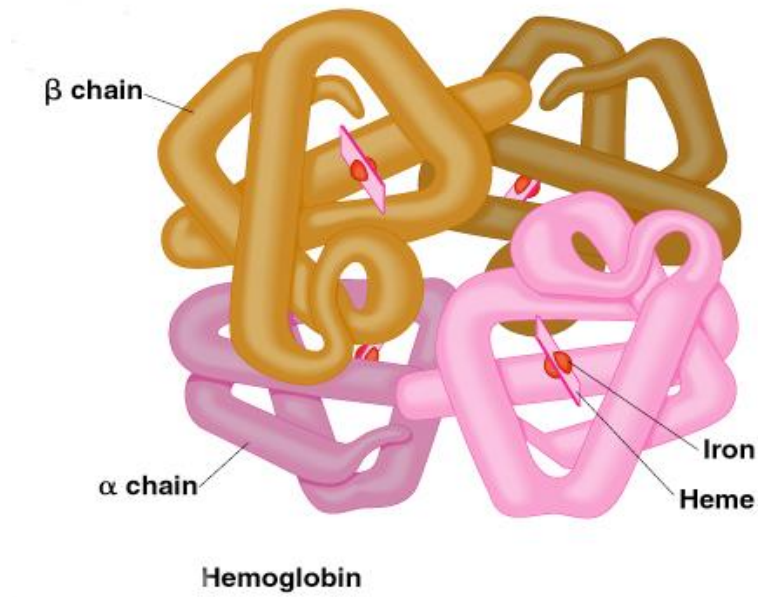
- Hidrojen bağları
- Disülfid bağları
- İyonik (tuz) bağlar (elektrostatik etkileşimler)
- Nonpolar etkileşimler
- Van der Waals bağları



Şekil 3.9. Tersiyer yapı

3.4.4. Kuaterner Yapı

Alt birimleri olan yada iki veya daha fazla polipeptitten oluşmuş proteinlerin göstermiş oldukları yapısal organizasyon kuaterner yapı olarak tanımlanır. Bir başka ifadeyle primer, sekonder ve tersiyer yapıları bulunan polipeptit zincirlerinin non-kovalent bağlarla bir arada tutulmasıdır. Bu yapıda alt birimler birbirlerine kovalent olmayan zayıf bağlarla (hidrojen bağları, elektrostatik çekimler gibi) bağlanmışlardır. Alt birimler aynı veya birbirinden farklı olabilirler.



Şekil 3.10. Kuaterner yapı

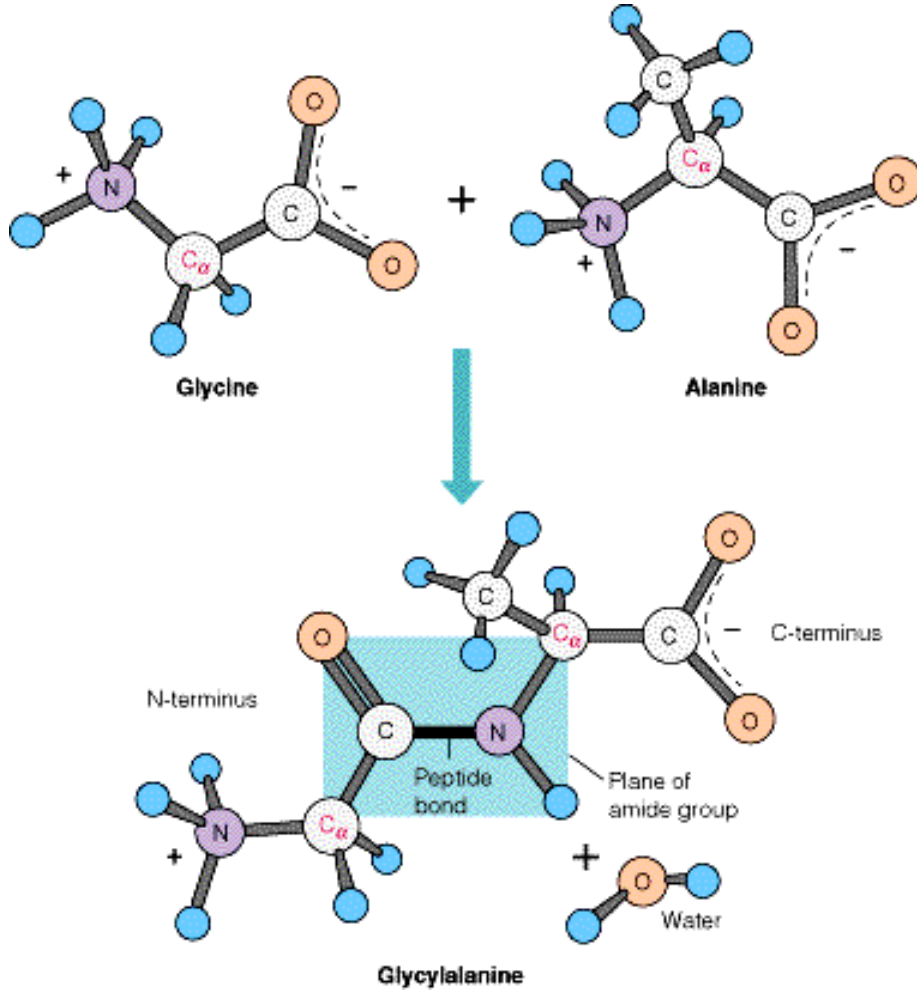
3.5. Peptit Bağı

İki amino asidin birbirine bağlanması ile oluşan yapıya peptit bağı denilir. Bu bağı yapımına, bir amino asidin karboksil ($-\text{COOH}$) grubu ile, diğer amino asidin amino grubu ($-\text{NH}_2$) katılmaktadır. Karboksil grubundan bir $-\text{OH}$ ile amino grubundan bir H birleşmekte ve su molekülü açığa çıkmaktadır.

İki amino asit bir araya gelirse dipeptit, üç amino asit bir araya gelirse tripeptit, daha uzun zincirler için ise polipeptit isimlendirmesi yapılır.

Polipeptitler isimlendirilirken N yani amino terminaldeki amino asidin adı ilk söylenir ve C terminal yani karboksil uç kısmı en son söylenecek şekilde isimlendirme yapılır.

Aşağıdaki şekil bir dipeptitin yapısını göstermektedir.



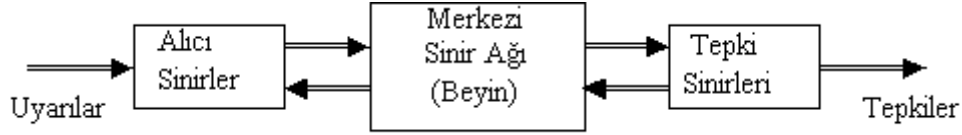
Şekil 3.11. Peptit bağı oluşumu

Bu tezde proteinlerin ikincil yapıları üzerinde çalışılırken peptitlerin merkezî karbonları olan C_{α} atomları arasında oluşan bağ açıları üzerinde durulmuştur.

4. YAPAY SİNİR AĞLARI

4.1. Biyolojik Sinir Sistemi

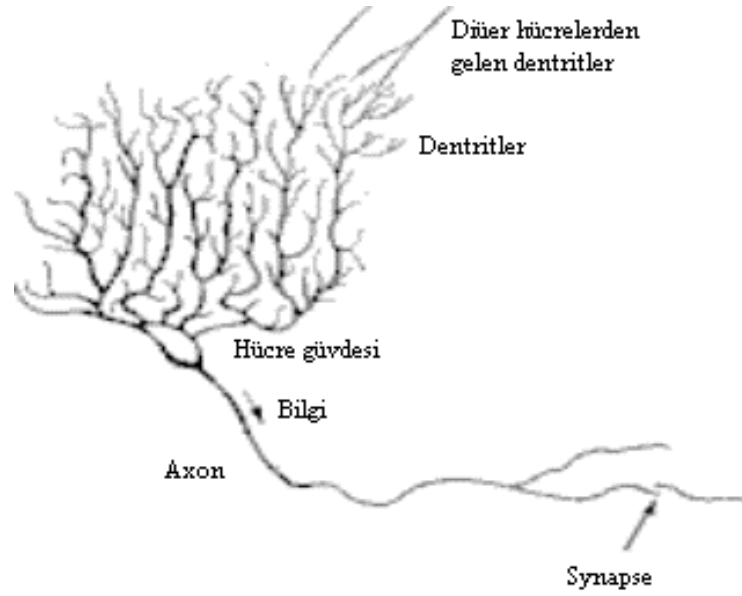
Biyolojik sinir sistemi, merkezinde sürekli olarak bilgiyi alan, yorumlayan ve uygun bir karar üreten beyin (merkezi sinir ağı) bulunduğu 3 katmanlı bir sistem olarak açıklanır. Alıcı sinirler (receptor) organizma içerisinde yada dış ortamlardan algıladıkları uyarıları, beyne bilgi ileten elektriksel sinyallere dönüştürür. Tepki sinirleri (effector) ise, beyinin ürettiği elektriksel darbeleri organizma çıktısı olarak uygun tepkilere dönüştürür. Şekil 4.1. de bir sinir sisteminin blok gösterimi verilmiştir. [8]



Şekil 4.1. Biyolojik sinir sistemi blok diyagramı

Merkezi sinir ağında bilgiler, alıcı ve tepki sinirleri arasında ileri ve geri besleme yönünde değerlendirilerek uygun tepkiler üretilir. Bu yönüyle biyolojik sinir sistemi, kapalı çevrim denetim sisteminin karakteristiklerini taşır. Merkezi sinir sisteminin temel işlem elemanı, sinir hücresidir (nöron) ve insan beyinde yaklaşık 10 milyar sinir hücresi olduğu tahmin edilmektedir. Sinir hücresi; hücre gövdesi, dendritler ve axonlar olmak üzere 3 bileşenden meydana gelir.

Dendritler, diğer hücrelerden aldığı bilgileri hücre gövdesine bir ağaç yapısı şeklinde ince yollarla iletir. Axonlar ise elektriksel darbeler şeklindeki bilgiyi hücreden dışarı taşıyan daha uzun bir yoldur. Axonların bitimi, ince yollara ayrılabilir ve bu yollar, diğer hücreler için dendritleri oluşturur. Şekil 4.2 de görüldüğü gibi axon-dendrit bağlantı elemanı synapse'dir.



Şekil 4.2. Biyolojik sinir hücre yapısı

Synapse'a gelen ve dendritler tarafından alınan bilgiler genellikle elektriksel darbelerdir. Ancak, synapse'daki kimyasal ileticilerden etkilenir. Belirli bir sürede bir hücreye gelen girişlerin değeri, belirli bir eşik değerine ulaştığında hücre bir tepki üretir. Hücrenin tepkisini artırıcı yöndeki girişler uyarıcı, azaltıcı yöndeki girişler ise önleyici girişler olarak söylenir ve bu etkiyi synapse belirler.

İnsan beyninin 10 milyar sinir hücresinden ve 60 trilyon synapse bağlantısından oluştuğu düşünülürse son derece karmaşık ve etkin bir yapı olduğu anlaşılır. Diğer taraftan bir sinir hücresinin tepki hızı, günümüz bilgisayarlarına göre oldukça yavaş olmakla birlikte duyuşal bilgileri son derecede hızlı değerdendirebilmektedir. Bu nedenle insan beyni; öğrenme, birleştirme, uyarılma ve genelleştirme yeteneğı nedeniyle son derece karmaşık, doğrusal olmayan ve paralel dağılmış bir bilgi işleme sistemi olarak tanımlanabilir.

4.2. Yapay Sinir Ağı

Günümüzde artık bilgisayarlar vazgeçilmez bir ihtiyaç haline gelmiştir. Evlerimizdeki bir çok eşyadan tutunda tıp, askeri alanlar, uzay bilimleri v.s. alanlarda kısacası hayatın her alanında bilgisayarlar ve bilgisayarlı sistemler hayatımızın ayrılmaz birer parçası olmuşlardır. İlk zamanlar sadece veri transferi yapmak ve bir takım karmaşık hesapları gerçekleştirmek amaçlı kullanılan bilgisayarlar zamanla büyük veri kümelerini inceleyip bunlardan sonuçlar çıkarabilecek kadar aşama kat etmiştir.

İnsanın davranışlarının bilgisayarlar tarafından modellenmesi çalışmalarının sonunda, yapay zeka bilimi başta olmak üzere bir çok önemli gelişme ortaya çıkmıştır. Artık bilgisayarlar hem veri toplayabilmekte hem de elde ettikleri bu verileri kullanarak olaylar hakkında yorumlar yaparak bir takım kararlara varabilmektedir. Bilinen bir takım klasik matematiksel metotlarla çözüm getirilemeyen problemlere, sezgisel bir takım yöntemlerle bilgisayarlar çözüm getirebilecek hale gelmiştir. Bu çalışmalar genel olarak yapay zeka bilimi çalışmaları olarak adlandırılmaktadır. Yapay sinir ağları ise yapay zeka biliminin bir alt dalı olarak çok ilgi gören bir alt daldır.

Yapay sinir ağları, insan beyninin özelliklerinden olan öğrenme yolu ile yeni bilgiler türetebilme, yeni bilgiler oluşturabilme ve keşfedebilme gibi yetenekleri herhangi bir yardım almadan otomatik olarak gerçekleştirme amacı ile geliştirilen bilgisayar sistemleridir.[7] Yapay sinir ağlarının olayları öğrenebilen , çevreden gelen olaylara karşı nasıl tepki verebileceğini bilen bir yapısı vardır. Bu yapısı öğrenebilen bir sistem olmasından dolayıdır.

İnsan beyninin aşağıdaki özelliklerini yapay sinir ağları taklit edebilmektedir:

- 1- Öğrenme
- 2- İlişkilendirebilme
- 3- Genelleme
- 4- Sınıflandırma
- 5- Özellik belirleme
- 6- Optimizasyon

Bir yapay sinir ağı,örneklerden oluşturduğu deneyim ile daha sonra benzer durumlarda kararlar alabilecek hale adapte olur.

Bir yapay sinir ağının en temel yaptığı iş, gelen örneklere karşılık gelebilecek bir çıkış elde etmektir. Tabi ki bu vazifeyi yerine getirebilmesi için daha öncesi için bir takım eğitim örnekleri ile ağın eğitilmiş olması gerekmektedir. Aynı şey insanoğlu içinde geçerlidir zaten. Eğitim belli bir seviyeye geldiğinde artık genelleme yapılabilir. Bu olaya öğrenme denir. Yapay zeka veya yapay sinir ağları açısından baktığımızda ise bu olaya makine öğrenmesi denir. Esasında temelde gerçekleşen şey sistemin örneklerden genellemeler yapabilir hale gelmesidir. Bu genellemeleri yapabilir hale geldikten sonra sistemin öğrenmesinin tamamlandığı düşünülür.

4.2.1. Yapay Sinir Ağının Özellikleri

Yapay sinir ağlarının hesaplama ve bilgi işleme gücünü, paralel dağılmış yapısından, öğrenebilme ve genelleme yeteneğinden aldığı söylenebilir. Genelleme, eğitim yada öğrenme sürecinde karşılaşılmayan girişler için de Yapay sinir ağının uygun tepkileri üretmesi olarak tanımlanır. Bu üstün özellikleri, yapay sinir ağlarının karmaşık problemleri çözebilme yeteneğini gösterir. Esasında yapay sinir ağlarının modellerine göre karakteristik özellikleri farklılıklar gösterebilmektedir. Fakat burada sıralayacağımız özellikler genel olup tüm modellerde bulunan özelliklerdir.

4.2.1.1. Makine Öğrenmesi Gerçekleştirme

Yapay sinir ağlarının temel işlevi sistemin öğrenmesini sağlamaktır. Olayları öğrenerek benzer olaylar karşısında benzer kararlar vermeye çalışırlar. [7] Esasında yapılan öğrenme karşılaşılan örnekler ile elde edilen tecrübeyi yeni olaylar karşısında yansıtmaktır.

4.2.1.2. Bilgi Saklama

Yapay sinir ağında bilgi nereye saklanacak ? Bu sorunun cevabı : ağırlıklar olacaktır. Klasik programlama yöntemlerinde olduğu gibi bilgi programın içinde gömülü yada bir bilgi tabanında saklı değil , sistem ağırlıkları üzerinde saklıdır.

4.2.1.3 Klasik Yöntemlere Benzememe

Klasik programlama yöntemlerinden veya diğer yapay zeka yöntemlerinden farklı bir işleyiş yönü vardır. Bilginin işlenişi diğer yöntemlerde olduğu gibi bir takım matematiksel yada mantıksal kurallara bağlı değildir.

4.2.1.4. Öğrenme

Yapay sinir ağları örnekleri kullanarak olay hakkında genellemeler yapabilecek yeteneğe kavuşur. Örnek mevcut değilse öğrenme olması mümkün değildir. Elde olan bütün örnekler olayın tamamen tüm yönleri ile gösterilmesi için önemlidir. Yapay sinir ağının arzu edilen davranışı gösterebilmesi için amaca uygun olarak ayarlanması gerekir. Bu, hücreler arasında doğru bağlantıların yapılması ve bağlantıların uygun ağırlıklara sahip olması gerektiğini ifade eder. Yapay sinir ağının karmaşık yapısı nedeniyle bağlantılar ve ağırlıklar önceden ayarlı olarak verilemez yada tasarlanamaz.

4.2.1.5. Doğrusal Olmama

Yapay sinir ağının temel işlem elemanı olan hücre doğrusal değildir. Dolayısıyla hücrelerin birleşmesinden meydana gelen yapay sinir ağıda doğrusal değildir ve bu özellik bütün ağa yayılmış durumdadır. Bu özelliği ile yapay sinir ağı, doğrusal olmayan karmaşık problemlerin çözümünde en önemli araç olmuştur.

4.2.1.6. Genelleme

Yapay sinir ağı, ilgilendiği problemi öğrendikten sonra eğitim sırasında karşılaşmadığı test örnekleri için de arzu edilen tepkiyi üretebilir. Örneğin, karakter tanıma amacıyla eğitilmiş bir yapay sinir ağı, bozuk karakter girişlerinde de doğru karakterleri verebilir yada bir sistemin eğitilmiş yapay sinir ağı modeli, eğitim sürecinde verilmeyen giriş sinyalleri için de sistemle aynı davranışı gösterebilir.

4.2.1.7. Uyarlanabilirlik

Yapay sinir ağı, ilgilendiği problemdeki değişikliklere göre ağırlıklarını ayarlar. Yani, belirli bir problemi çözmek amacıyla eğitilen yapay sinir ağı, problemdeki değişimlere göre tekrar eğitilebilir, değişimler devamlı ise gerçek zamanda da eğitime devam edilebilir. Bu özelliği ile yapay sinir ağı, uyarlamalı örnek tanıma, sinyal işleme, sistem tanılama ve denetim gibi alanlarda etkin olarak kullanılır.

4.2.1.8. Hata Toleransı

Yapay sinir ağı, çok sayıda hücrenin çeşitli şekillerde bağlanmasından oluştuğundan paralel dağılmış bir yapıya sahiptir ve ağın sahip olduğu bilgi, ağdaki bütün bağlantılar üzerine dağılmış durumdadır. Bu nedenle, eğitilmiş bir yapay sinir ağının bazı bağlantılarının hatta bazı hücrelerinin etkisiz hale gelmesi, ağın doğru bilgi üretmesini önemli ölçüde etkilemez. Bu nedenle, geleneksel yöntemlere göre hatayı tolere etme yetenekleri son derece yüksektir.

4.2.1.9. Eksik Bilgi ile Çalışabilme

Yapay sinir ağları eğitimini tamamladıktan sonra gelen örnekte eksik bilgi olsa da sonuç üretebilirler. Eksik bilgi ile çalışmak ağın performansının düşmesi demek değildir. Eksik olan bilginin önemli olup olmadığı ağın eğitimi esnasında anlaşılacaktır. Eğitim esnasında eksik olan bilgi performans düşüklüğüne sebep oluyorsa zaten eğ için önemlidir. Bunun dışında zaten ağ onu tolere edecektir.

4.2.1.10. Donanım ve Hız

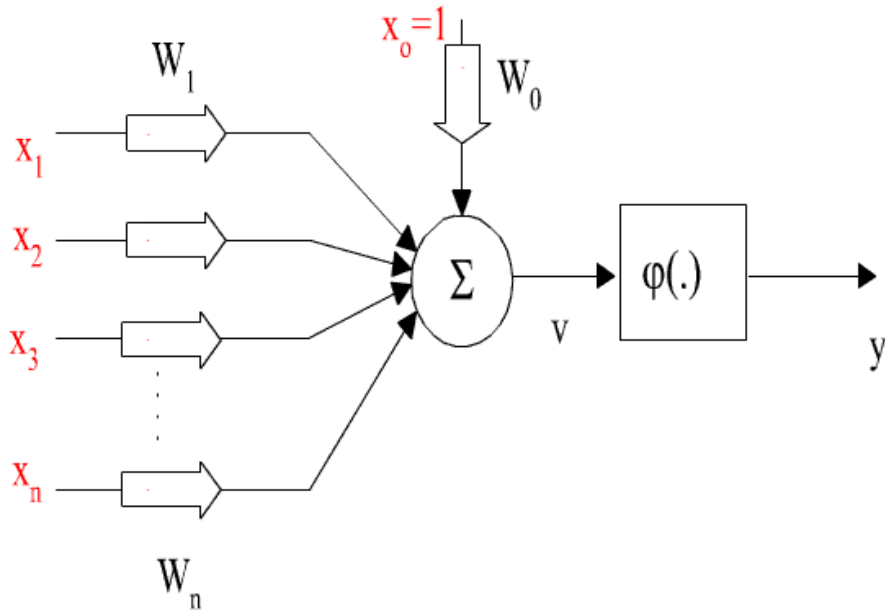
Yapay sinir ağı, paralel yapısı nedeniyle büyük ölçekli entegre devre (VLSI) teknolojisi ile gerçekleştirilebilir. Bu özellik, yapay sinir ağının hızlı bilgi işleme yeteneğini artırır ve gerçek zamanlı uygulamalarda arzu edilir.

4.2.1.11. Analiz ve Tasarım Kolaylığı

Yapay sinir ağının temel işlem elemanı olan hücrenin yapısı ve modeli, bölüm 1.3'de açıklandığı gibi bütün yapay sinir ağı yapılarında yaklaşık aynıdır. Dolayısıyla, yapay sinir ağının farklı uygulama alanlarındaki yapıları da standart yapıdaki bu hücrelerden oluşacaktır. Bu nedenle, farklı uygulama alanlarında kullanılan yapay sinir ağları benzer öğrenme algoritmalarını ve teorilerini paylaşabilirler. Bu özellik, problemlerin yapay sinir ağları ile çözümünde önemli bir kolaylık getirecektir.

4.2.2. Yapay Sinir Hücresi

Biyolojik sinir sisteminde olduğu gibi yapay sinir ağlarında da sistemi oluşturan temel eleman yapay sinir hücresidir. Bu sinir hücreleri farklı modellerde farklı şekillerde bir araya gelerek yapay sinir ağlarını oluştururlar. Her yapay sinir hücresinin 5 temel elemanı vardır. Şekil 4.3. bir yapay sinir hücresinin yapısını göstermektedir.



Şekil 4.3. Yapay sinir hücresi modeli

4.2.2.1. Girdiler

Sisteme dışarıdan gelen bilgilerdir. Bunlar ağı öğrenmesi istenilen örnekler tarafından belirlenir. Girdiler dış dünyadan gelen bilgiler olabileceği gibi bir başka hücreden gelen bir sinyalde olabilir.

4.2.2.2. Ağırlıklar

Bir yapay hücreye gelen önemini ve üzerindeki etkisini gösteren değerdir. Şekil 4.3. 'de $W_1, \dots, W_n$ ile ifade edilen değerler hücrelerin ağırlık değerleridir. Ağırlıkların büyüklüğü yada küçüklüğü önemini göstermez. Hatta bir ağırlık değerinin sıfır olması belki de en önemli olay olabilir.

4.2.2.3. Toplama Fonksiyonu

Hücreye gelen net girdiyi hesaplayan fonksiyondur. Değişik toplama fonksiyonları kullanılmaktadır. En yaygın olanı ağırlıklı toplamdır. Her gelen girdi ağırlık değeri ile çarpılarak toplanır. Böylece ağı gelen net girdi bulunmuş olur.

$$\text{Net Girdi} = \sum_i^n X_i \cdot W_i$$

Burada X girdileri W ise ağırlıkları temsil etmektedir. n bir hücreye gelen girdi sayısını göstermektedir. Genellikle deneme yanılma yoluyla bir problem için en uygun toplama fonksiyonu bulunur. [7] Aşağıdaki tabloda toplama fonksiyon örnekleri verilmiştir.

Tablo 4.1. Toplama fonksiyonu örnekleri

$$\text{Çarpım Net Girdi} = \prod_i^n X_i W_i$$

Maksimum

$$\text{Net Girdi} = \text{Max} (X_i \cdot W_i) \quad i=1, \dots, N$$

Minimum

$$\text{Net Girdi} = \text{Min} (X_i \cdot W_i) \quad i=1, \dots, N$$

Çoğunluk

$$\text{Net Girdi} = \sum_i^n \text{sgn}(X_i W_i)$$

Kümülatif toplam

$$\text{Net Girdi} = \text{Net (eski)} + \sum (X_i W_i)$$

Ağırlık değerleri girdiler ile çarpılır ve daha sonra bulunan değerler çarpılıp hesap yapılır.

N tane net girdi içinden ağırlıklar ile çarpıldıktan sonra en büyük olan kabul edilir.

N tane net girdi içinden ağırlıklar ile çarpıldıktan sonra en küçük olan kabul edilir.

N adet girdiden ağırlıklar ile çarpıldıktan sonra pozitif ve negatiflerin sayısı bulunur .Büyük olan net girdi kabul edilir.

Ağırlıklı olan toplam eski bilgilerle toplanarak net girdi hesaplanır.

4.2.2.4. Aktivasyon Fonksiyonu

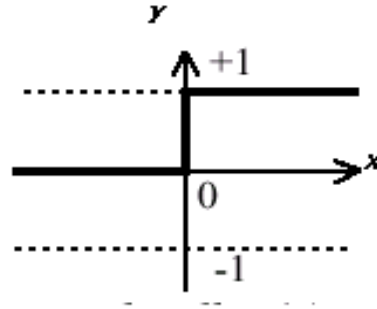
Hücreye gelen net girdiyi işleyerek hücrenin bu girdiye karşılık bir çıktı üretmesini sağlar. Aktivasyon fonksiyonunun da bir çok farklı çeşidi vardır. Tüm hücrelerin aynı aktivasyon fonksiyonunu kullanması da gerekmez. Farklı aktivasyon hücreleri kullanılması mümkündür. Yine bu probleme özel seçilebilir. En yaygın kullanılan aktivasyon fonksiyonu sigmoid fonksiyonudur.

$$F(\text{Net}) = 1 / (1 + e^{-\text{NET}})$$

Burada NET hücreye gelen net girdiyi göstermektedir. yukarıda belirtildiği gibi bu NET değeri toplama fonksiyonundan hesaplanarak gelmektedir. Aktivasyon fonksiyonu olarak da değişik fonksiyonlar kullanılmaktadır. Bunlardan bazıları aşağıdaki gibidir:

Eşik Aktivasyon fonksiyonu

Eşik aktivasyon fonksiyonu eğer net değeri sıfırdan küçükse sıfır, sıfırdan daha büyük bir değer ise net çıkışında +1 değeri verir. Eşik aktivasyon fonksiyonunun -1 ile +1 arasında değişeni ise signum aktivasyon fonksiyonu olarak adlandırılır. Signum aktivasyon fonksiyonu, net giriş değeri sıfırdan büyükse +1, sıfırdan küçükse -1, sıfıra eşitse sıfır değerini verir.

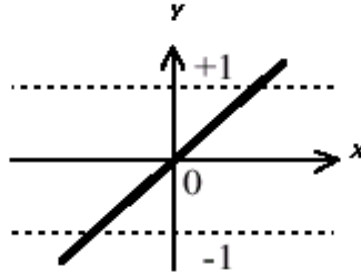


Şekil 4.4.. Eşik aktivasyon fonksiyonu

Doğrusal Aktivasyon fonksiyonu

Lineer aktivasyon fonksiyonunun çıkışı girişine eşittir. Sürekli çıkışlar gerektiği zaman çıkış katmanındaki aktivasyon fonksiyonunun lineer aktivasyon fonksiyonu olabildiğine dikkat edilmelidir.

$F(x) = x$ şeklinde ifade edilir.

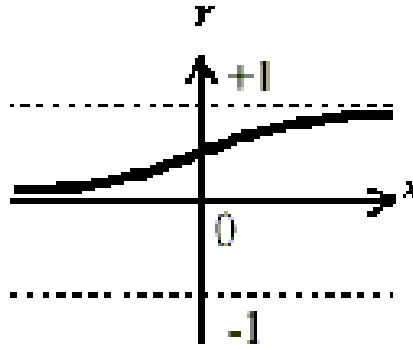


Şekil 4.5. Doğrusal aktivasyon fonksiyonu

Logaritma Sigmoid Aktivasyon Fonksiyonu

Lojistik fonksiyon olarak da adlandırılmaktadır. Bu fonksiyonunun lineer olmamasından dolayı türevi alınabilmektedir böylece daha sonraki bölümlerde görülecek olan geri yayımlı ağırlarda kullanmak mümkün olabilmektedir.

$f(x) = \text{lojistik}(x) = \frac{1}{1 + \exp(-\beta x)}$ şeklinde ifade edilir. Buradaki β eğim sabiti olup genelde bir olarak seçilmektedir.

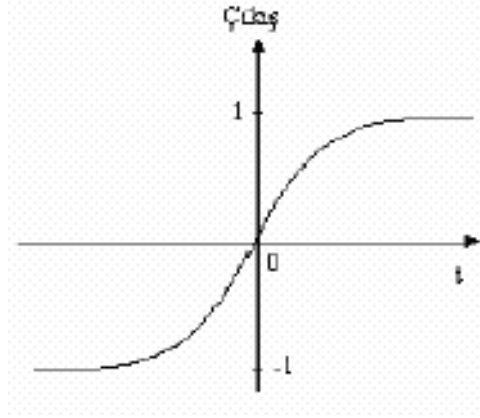


Şekil 4.6. Logaritma sigmoid aktivasyon fonksiyonu

Hiperbolik Tanjant Aktivasyon Fonksiyonu

Hiperbolik tanjant aktivasyon fonksiyonu da lineer olmayan türevi alınabilir bir fonksiyondur. +1 ile -1 arasında çıkış değerleri üreten bu fonksiyon lojistik fonksiyona benzemektedir.

$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ şeklinde ifade edilir.



Şekil 4.7. Tanjant Hiperbolik Aktivasyon Fonksiyonu

Bu aktivasyon fonksiyonları dışında sinüs fonksiyonu, basamak fonksiyonu, kutuplamalı basamak fonksiyonu, parçalı doğrusal fonksiyon v.b. aktivasyon fonksiyonları da mevcuttur. Probleme göre tasarımcı en uygun olanı seçecektir. Seçim için bir kriter yoktur deneme yolu ile bulunur.

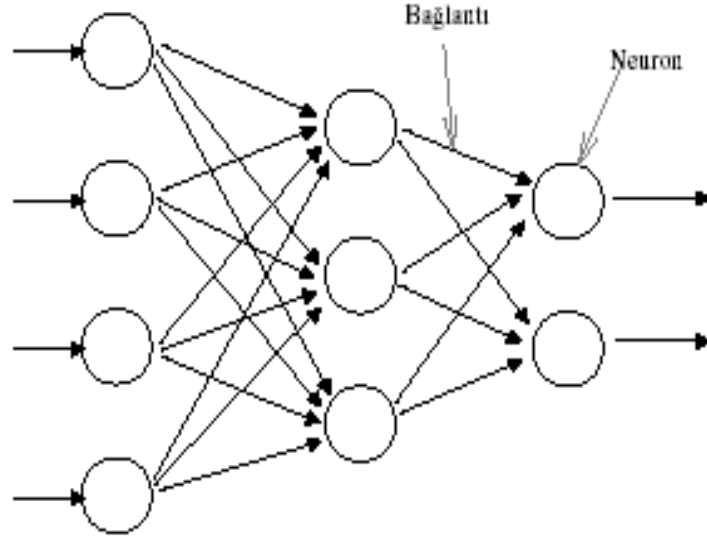
4.2.2.5. Çıktı

Aktivasyon fonksiyonu ile belirlenen çıkış değeridir. Çıkış sistemin son çıkışı olabileceği gibi bir başka hücreye giriş olabilecek bir çıkış da olabilir. Yapay sinir ağlarının yapısına bakıldığında zaman bir hücrenin birden fazla çıktısı varmış gibi görünmesine rağmen aslında her hücrenin bir tek çıktısı vardır. Sadece her hücrenin çıkışı birden fazla elemana girdi olarak işlenebileceğinden gösterim olarak o şekilde temsil edilmektedir.

4.2.3. Yapay Sinir Ağının Yapısı

Yapay sinir hücrelerinin bir araya gelmesiyle yapay sinir ağı oluşur. Bu işlem tabi ki belirli bir düzende olur. Genelde yapay sinir ağlarında 3 katmanlı bir yapı vardır.

- 1- Giriş katmanı
- 2- Ara katman
- 3- Çıkış katmanı



Şekil 4.8. Yapay sinir ağı modeli

4.2.3.1. Giriş Katmanı

Bu katman dış dünyadan gelen bilgileri alarak yapay sinir ağına iletmekle sorumlu katmandır. Genelde girdi katmanında bilgi işlenmeden sisteme iletilir.

4.2.3.2. Ara Katman

Giriş katmanından gelen bilgi bu katman veya katmanlarda işlenerek çıktı katmanına iletilir. Bilginin işlenmesi bu katmanda gerçekleştirilir. Tek katman olması şart değildir. Birden fazla katmandan oluşabilir. Yine bu sistemi tasarlayan kimsenin problemine özel seçtiği duruma bağlıdır. Bazen tek katman çözüm olabilirken bazen de bir tek katmanla probleme yakınsanmayabilir.

4.2.3.3. Çıkış Katmanı

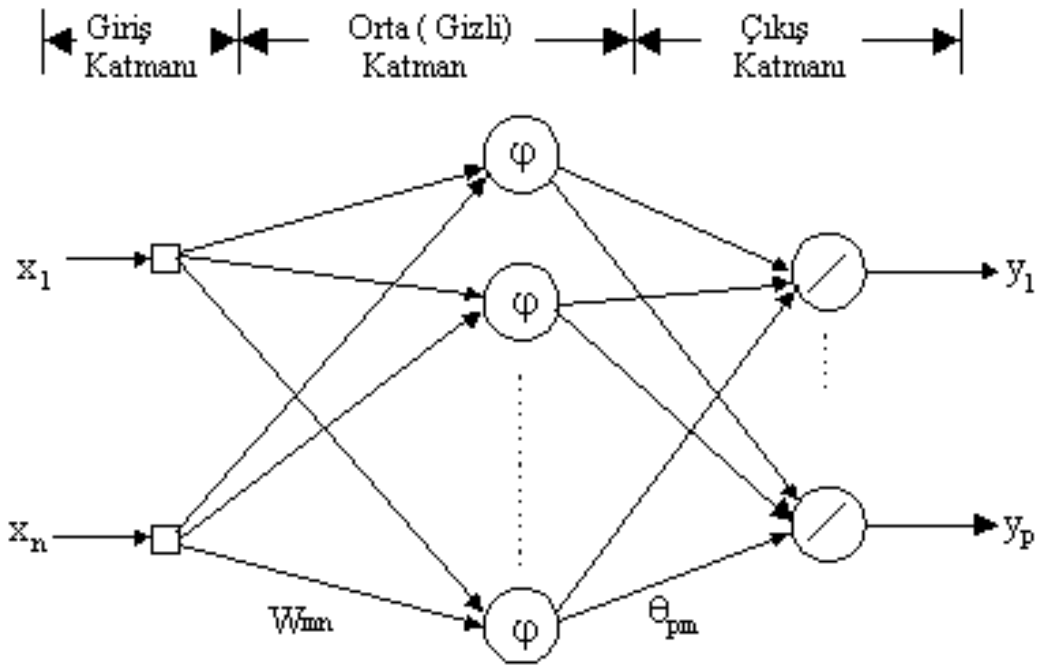
Ara katmandan gelen bilgileri işleyerek girişten gelen örnekler için uygun çıktı değerini sistemin çıkış değerleri olarak sunan katmandır. Başka bir ifade ile sistemin girişe olan cevabının sunulduğu katmandır.

4.2.4. Yapay Sinir Ağı Yapıları

Yapay sinir ağları, hücrelerin birbirleri ile çeşitli şekillerde bağlanmalarından oluşur. Hücre çıkışları, ağırlıklar üzerinden diğer hücelere yada kendisine giriş olarak bağlanabilir ve bağlantılarda gecikme birimi de kullanılabilir. Hücrelerin bağlantı şekillerine, öğrenme kurallarına ve aktivasyon fonksiyonlarına göre çeşitli yapay sinir ağları yapıları geliştirilmiştir.

4.2.4.1. İleri Beslemeli Yapay Sinir Ağları

İleri beslemeli yapay sinir ağlarında, hücreler katmanlar şeklinde düzenlenir ve bir katmandaki hücrelerin çıkışları bir sonraki katmana ağırlıklar üzerinden giriş olarak verilir. Giriş katmanı, dış ortamlardan aldığı bilgileri hiçbir değişikliğe uğratmadan ara (gizli) katmandaki hücrelere iletir. Bilgi, orta ve çıkış katmanında işlenerek ağ çıkışı belirlenir. Bu yapısı ile ileri beslemeli ağlar doğrusal olmayan statik bir işlevi gerçekleştirir. İleri beslemeli 3 katmanlı yapay sinir ağının, orta katmanında yeterli sayıda hücre olmak kaydıyla, herhangi bir sürekli fonksiyonu istenilen doğrulukta yaklaştırabileceği gösterilmiştir. En çok bilinen geriye yayılım öğrenme algoritması, bu tip yapay sinir ağlarının eğitiminde etkin olarak kullanılmakta ve bazen bu ağlara geriye yayılım ağları da denmektedir. Şekil 4.9 da giriş, orta ve çıkış katmanı olmak üzere 3 katmanlı ileri beslemeli yapay sinir ağı yapısı verilmiştir.



Şeki
1

4.9. İleri beslemeli yapay sinir ağı modeli

İleri beslemeli 3 katmanlı ve çıkış katmanı doğrusal olan yapay sinir ağının matematiksel modeli, x - giriş vektörünü, o - orta katman çıkış vektörünü, y - ağ çıkış vektörünü göstermek üzere aşağıdaki formülde olduğu gibi yazılabilir. x_0 ve o_0 girişleri, polarma girişleri olarak alınmıştır.

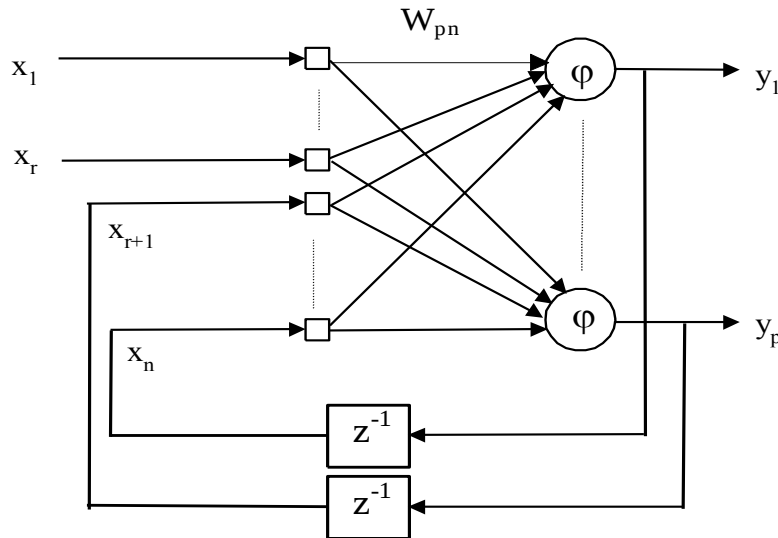
$$v_j = \sum_{i=0}^n W_{ji} \cdot x_i, \quad o_j = \varphi(v_j) \quad j = 1, 2, \dots, m$$

$$y_\ell = \sum_{j=0}^m \theta_{\ell j} \cdot o_j \quad \ell = 1, 2, \dots, p$$

Herhangi bir problemi çözmek amacıyla kullanılan yapay sinir ağlarında, katman sayısı ve orta katmandaki hücre sayısı gibi kesin belirlenememiş bilgilere rağmen nesne tanıma ve sinyal işleme gibi alanların yanı sıra ileri beslemeli yapay sinir ağı, sistemlerin tanılanması ve denetiminde de yaygın olarak kullanılmaktadır.

4.2.4.2. Geri Beslemeli Yapay Sinir Ağları

Geri beslemeli yapay sinir ağlarında, en az bir hücrenin çıkışı kendisine yada diğer hücrelere giriş olarak verilir ve genellikle geri besleme bir geciktirme elemanı üzerinden yapılır. Geri besleme, bir katmandaki hücreler arasında olduğu gibi katmanlar arasındaki hücreler arasında da olabilir. Bu yapısı ile geri beslemeli yapay sinir ağı, doğrusal olmayan dinamik bir davranış gösterir. Dolayısıyla, geri beslemenin yapılış şekline göre farklı yapıda ve davranışta geri beslemeli yapay sinir ağı yapıları elde edilebilir.. Şekil 4.10 da iki katmanlı ve çıkışlarından giriş katmanına geri beslemeli bir yapay sinir ağı yapısı görülmektedir.



Şekil 4.10. Geri beslemeli yapay sinir ağı modeli

Şekil 4.12 de verilen geri beslemeli yapay sinir ağında giriş vektörü, r adet dış giriş ve p adet gecikmiş ağ çıkışlarından oluşmaktadır. Buna göre yapay sinir ağının giriş vektörü aşağıdaki gibi yazılabilir.

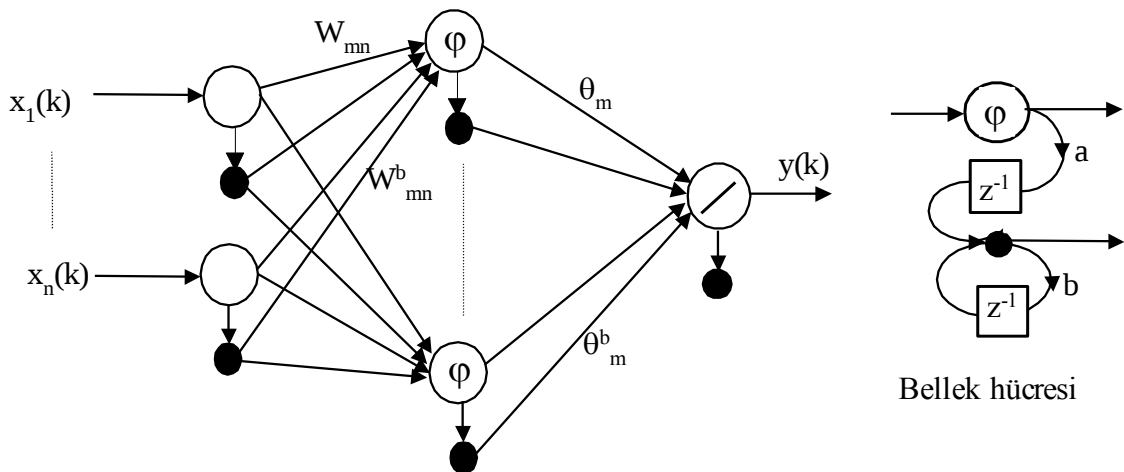
$$x = (x_1, x_2, \dots, x_r, y_1(-1), y_2(-1), \dots, y_p(-1))$$

Yukarıdaki denklem ile tanımlanan toplam $n=r+p$ adet ağ girişine göre yapay sinir ağının çıkışları aşağıdaki gibi yazılabilir.

$$y_j = \varphi \left(\sum_{i=0}^n W_{ji} \cdot x_i \right) \quad j=0,1,2,\dots,p$$

4.2.4.3. Bellek Hücreli Yapay Sinir Ağları

Doğrusal olmayan sistemlerin tanınması ve denetiminde, katmanlı yapay sinir ağı yapıları etkin olarak kullanılmaktadır. Yapay sinir ağı ile sistem tanılamada, doğru model yapısının seçilebilmesi ve model girişlerinin belirlenebilmesi için sistemin giriş ve çıkışının gecikme derecelerinin bilinmesi gerekir. Sistemin derecesinin doğru belirlenememesi, modelde temsil edilemeyen dinamikler nedeniyle kararlı ve değişen dinamik şartlarda doğru bir model elde edilmesini etkiler. Bu nedenle, geri beslemeli yapay sinir ağı yapıları kullanılarak sistemin derecesine ihtiyaç duymayan tanı modelleri geliştirilmiştir. Şekil 4.13’de Bellek Hücreli Yapay Sinir Ağları olarak söylenen ve ağdaki her bir hücre için bir bellek hücresinin kullanıldığı katmanlı-geri beslemeli bir ağ yapısı verilmiştir.



Şekil 4.11. Bellek hücreli yapay sinir ağı modeli

Bellek hücreli yapay sinir ağlarında her bir ağ hücresine ait olan bellek hücresi, bir ağırlık (b_j) üzerinden öz geri besleme girişine ve başka bir ağırlık üzerinden (a_j) ait olduğu hücrenin gecikmiş girişine göre bir çıkış üretir. Çıkış katmanında ise genellikle sadece öz geri besleme kullanılır. Buna göre Bellek hücreli yapay sinir ağlarının matematiksel ifadesi aşağıdaki gibi yazılabilir.

$$x_i^b(k) = w_j^a \cdot x_i(k-1) + w_j^b \cdot x_i^b(k-1)$$

$$v_j(k) = \sum_{i=1}^n \{W_{ji} \cdot x_i(k) + W_{ji}^b \cdot x_i^b(k)\}$$

$$o_j(k) = \varphi\{v_j(k)\}$$

$$o_j^b(k) = \theta_j^a \cdot o_j(k-1) + \theta_j^b \cdot o_j^b(k-1)$$

$$y(k) = \sum_{j=1}^m \{\theta_j^c \cdot o_j(k) + \theta_j^d \cdot o_j^b(k)\} + \theta^c \cdot y(k-1)$$

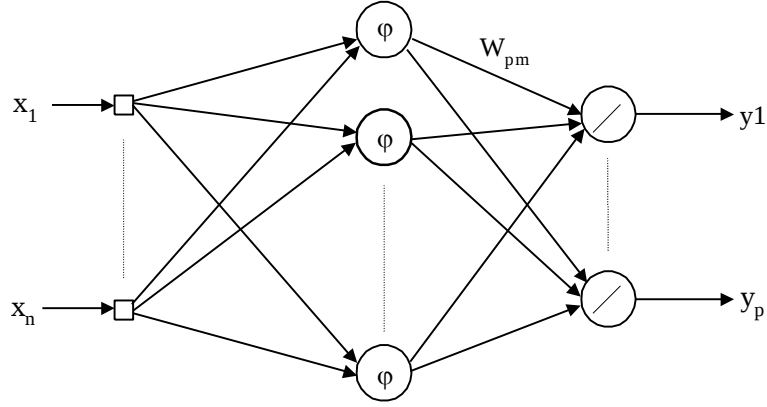
Bellek hücreli yapay sinir ağlarının doğrusal olmayan bir sistemi modelleme ve denetim yeteneği, sadece sistemin o an ki giriş ve bir önceki çıkış verileri model girişi alınarak incelenmiş ve tatmin edici sonuçlar alındığı gösterilmiştir. İleri beslemeli katmanlı yapay sinir ağlarının sadece gizli katmanında bellek hücreleri kullanılarak bellek hücresinin, ait olduğu hücre çıkışının geçmişteki örneklerini giriş olarak aldığı ve zaman gecikmeli yapay sinir ağı olarak söylenen geri beslemeli ağ yapıları da incelenmiştir.

4.2.4.4. Radyal Tabanlı Fonksiyon Ağları

Katmanlı yapay sinir ağlarının tasarımında eğiticili geriye yayılım öğrenme algoritması bir en iyileme uygulamasıdır. Radyal tabanlı fonksiyon ağı tasarımı ise çok boyutlu uzayda eğri uydurma yaklaşımıdır ve bu nedenle radyal tabanlı fonksiyon ağının eğitimi, çok boyutlu uzayda eğitim verilerine en uygun bir yüzeyi bulma problemine dönüşür. Radyal tabanlı fonksiyon ağının genellemesi ise test verilerini interpolate etmek amacıyla, eğitim sırasında bulunan çok boyutlu yüzeyin kullanılmasına eşdeğerdir. Radyal tabanlı fonksiyonlar, sayısal analizde çok değişkenli interpolasyon problemlerinin çözümünde kullanılmış ve yapay sinir ağının gelişmesi ile birlikte bu fonksiyonlardan yapay sinir ağı tasarımında yararlanılmıştır.[8]

Radyal tabanlı fonksiyon ağları, ileri beslemeli yapay sinir ağı yapılarına benzer şekilde giriş, orta ve çıkış katmanından oluşur ancak, giriş katmanından orta katmana dönüşüm, radyal

tabanlı aktivasyon fonksiyonları ile doğrusal olmayan sabit bir dönüşümdür. Orta katmandan çıkış katmanına ise uyarlamalı ve doğrusal bir dönüşüm gerçekleştirilir.



Şekil 4.12. Radyal tabanlı fonksiyon ağırları modeli

Eğri uydurma teorisi, herhangi bir çok değişkenli ve sürekli $f(x)$ fonksiyonunu yaklaştırma yada interpolate etme problemi ile ilgilidir. İnterpolasyon problemi, $k = 1, 2, \dots, N$ için x_k – veri noktası ve d_k -gerçek değerler olmak üzere $F(x_k) = d_k$ interpolasyon koşulunu sağlayan $F(\cdot)$ fonksiyonunun bulunması olarak tanımlanır. Radyal tabanlı fonksiyonlarla doğrusal $F(\cdot)$; aşağıdaki gibi tanımlanır. [8]

$$F(x) = \sum_{k=1}^N W_k \cdot \varphi_k(\|x - x_k\|)$$

Burada $\varphi(\cdot)$ - doğrusal olmayan radyal tabanlı fonksiyonu, $\|\cdot\|$ - genellikle standart öklit uzaklığıdır. Bilinen veri noktaları, radyal tabanlı fonksiyonların merkezleri olarak söylenir. RTFA'nın genel yapısında veri örneği kadar orta katman hücrelerine ve ağırlığa gerek duyulduğundan genel çözümü yakınsayan en iyi çözüm aranır. $M < N$ olmak üzere RTFA çıkışı,

$$F^*(x) = \sum_{i=1}^M W_i \cdot \varphi_i(\|x - c_i\|)$$

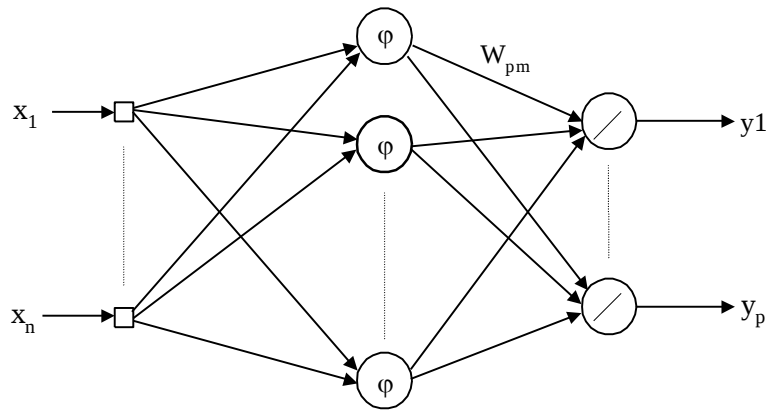
Burada c_i – veri noktalarından belirlenecek olan radyal tabanlı fonksiyonların merkez vektörleridir. RTFA'da orta katman aktivasyon fonksiyonu genellikle standart öklit uzaklıklarını üstsel fonksiyondan geçiren gaussian fonksiyonudur ve aşağıdaki gibi tanımlanır.[8]

$$\varphi_i(\|x - c_i\|) = e^{-\frac{1}{2\sigma^2}\|x - c_i\|^2}$$

RTFA’da uyarlanabilecek serbest parametreler; merkez vektörleri, radyal fonksiyonların genişliği ve çıkış katman ağırlıklarıdır. Çıkış katmanı doğrusal olduğundan ağırlıklar, eğitim düşme yada doğrusal en iyileme yöntemleri ile kolayca bulunabilir. Merkezler, girişler arasından rasgele ve sabit olarak seçilebilmekle birlikte RTFA’nın performansını iyileştirmek amacıyla merkez vektörlerinin ve genişliğin uyarlanması için çeşitli yöntemler geliştirilmiştir. Merkez vektörleri, eğitim düşme yöntemine göre eğitici öğrenme algoritması ile uyarlanarak, dik en küçük kareler yöntemi ile, yada kendiliğinden düzenlemeli yöntemle giriş örneklerinden öbekleme yapılarak belirlenebilir.

4.2.4.5. Fonksiyonel Link Ağları

Katmanlı YSA, orta katmandaki doğrusal olmayan aktivasyon fonksiyonları nedeniyle doğrusal olmayan öğrenme algoritmaları ile eğitilmelidir. Bu durum, öğrenme hızını yavaşlatır ve fonksiyon yaklaşımında yöresel en aza ulaşılabilir. Bu sorunlar, ağ girişlerini önce doğrusal olmayan bir dönüşümle genişlettikten sonra doğrusal çıkış katmanlı ağ yapıları ile giderilebilir. Şekil 1.16 da verilen ve Fonksiyonel Link Ağları olarak bilinen ağlar iyi bir fonksiyon yaklaştırma performansına sahiptir. Bu yönüyle fonksiyonel link ağları, merkezleri ve genişliği sabit tutulan radyal tabanlı fonksiyon ağlarına benzer ancak, fonksiyonel link ağlarında orta katmanın görevi ve aktivasyon fonksiyonları farklıdır.



Şekil 4.13. Fonksiyonel link ağları modeli

Fonksiyonel link ağlarında ağ çıkışı, orta katmanda genişletilmiş ağ girişlerinin doğrusal toplamıdır ve fonksiyonel link ağlarının eğitimi yalnızca çıkış katman ağırlıklarının ayarlanmasına gereksinim duyar.

$$y_l = \sum_{j=1}^m W_{lj} \cdot \varphi_j(x) \quad l=1,2,\dots,p$$

Ağ girişlerini genişleten orta katman aktivasyon fonksiyonları;

$\sin(k.\pi.x_i), \cos(k.\pi.x_i), e^{-x_i^2/x_j}$ gibi genellikle üstel, trigonometrik yada polinomsal açılım fonksiyonlarıdır.

$$\{x_i, 1 \leq i \leq n\} \Rightarrow \{x_i, x_i.x_j, x_i.x_j.x_k \quad 1 \leq i < j < k \leq n\}$$

Fonksiyonel link ağlarının fonksiyon yaklaştırma performansı, ağ girişlerinin yeterince genişletilmesine bağlıdır. Yeterince yüksek dereceden aktivasyon fonksiyonu içeren fonksiyonel link ağlarının, herhangi bir doğrusal olmayan sürekli fonksiyonu arzu edilen doğrulukta yaklaştırabileceği belirlenmiştir. Ancak, ağ girişleri arttıkça orta katman aktivasyon fonksiyonları büyük boyutlara ulaşacak ve ağın gerçekleştirilmesi güç olacaktır. Örneğin, 20 girişli bir ağ için 3.dereceden polinomsal açılımla 1270 adet genişletilmiş giriş oluşur. Bu nedenle, arzu edilen doğruluğu sağlayacak şekilde en iyiye yakın bir alt model seçimine gereksinim duyulur. Radyal tabanlı fonksiyon ağlarının benzer şekilde çeşitli yöntemlerin yanı sıra dik en küçük kareler yöntemi Fonksiyonel link ağları da alt model seçiminde etkin olarak kullanılır.

Bunların dışında çağrışimli bellek ağları, modül yapay sinir ağları gibi farklı yapay sinir ağları modelleri de mevcuttur.

4.2.5. Yapay Sinir Ağlarında Öğrenme

Yapay sinir ağlarının en önemli özelliklerinden biri, ilgilendiği problemde aldığı örneklerle problemi öğrenerek öz yeteneğini düzeltebilmesidir. Yapay sinir ağları, tanımlanan bir ölçüte göre ağı serbest parametrelerinin ardışıl olarak uyarlanmasıyla problemi öğrenir. Yapay sinir ağları açısından öğrenme aşağıdaki gibi tanımlanır.

Öğrenme, yapay sinir ağının ilgilendiği ortam tarafından ağı belirli bir süre uyarılmasıyla yapay sinir ağının serbest parametrelerinin, arzu edilen öz yeteneği sağlayacak şekilde ayarlanması işlemidir ve öğrenmenin tipi, parametre değişikliklerinin yapılış şekline bağlıdır. [8]

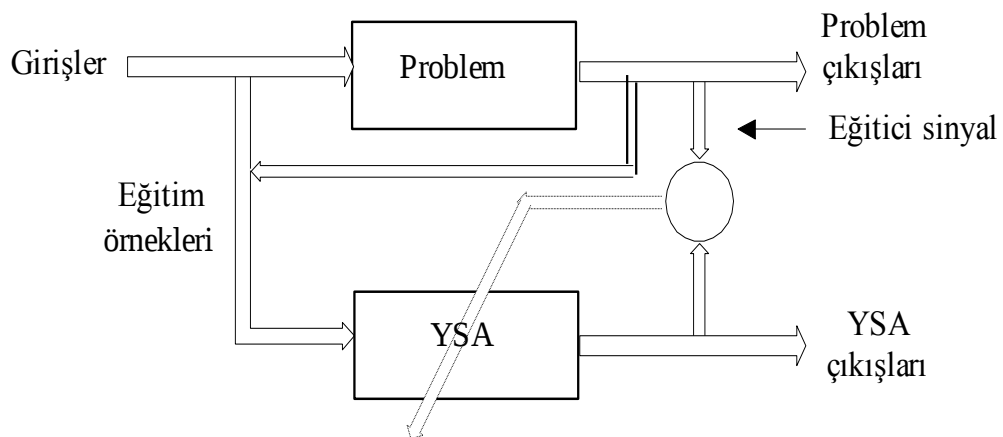
Bu tanıma göre öğrenme (yada eğitim) sürecinde; ağın uyarılması, bu uyarının sonucu olarak ağın parametrelerinin uyarlanması ve ağın probleme yeni tepkisini vermesi gerekir. Yapay sinir ağındaki herhangi bir ağırlık ele alınırsa öğrenme, matematiksel olarak aşağıdaki gibi ifade edilir.

$$W_{yeni} = W_{eski} \pm \Delta W$$

Burada ΔW , belirli bir kurala göre hesaplanarak o anki ağırlıklara uygulanacak düzeltme miktarını verir ve ΔW yi belirlemek için tanımlanmış kurallara öğrenme algoritmaları yada kuralları denir. Yapay sinir ağına ilgilendiği ortam tarafından uyarılma yöntemi de öğrenmede önemli bir etkidir ve bu tanımlara göre öğrenme çeşitli sınıflara ayrılır.

4.2.5.1. Eğitici Öğrenme

Yapay sinir ağına öğrenbilmesi için ağına ilgilendiği ortamın (problemin) davranışlarını ihtiva eden bir eğitici sinyal kullanılır ve bu sinyal, Yapay sinir ağı için arzu edilen tepkiyi temsil eder. Eğitici sinyalle birlikte problemde alınan ağ girişleri eğitim örneklerini oluşturur. Eğitici çıkış ile gerçek ağ çıkışı arasındaki hata sinyaline göre Yapay sinir ağı ağırlıkları uyarılır. Sonuçta, eğitici sinyalde bulunan ilgili problemin davranışları Yapay sinir ağı tarafından bilinir ve eğitimsiz de aynı davranışı gösterebilir.[8] Bu özelliği ile eğitici öğrenme fonksiyon yaklaştırma, nesne tanıma, sistem tanılama ve uyarlamalı denetim alanlarında kullanılan yapay sinir ağları için etkin bir öğrenme yöntemidir. Eğitici öğrenme yöntemi, Şekil 4.16. da verilen blok şema ile gösterilebilir.

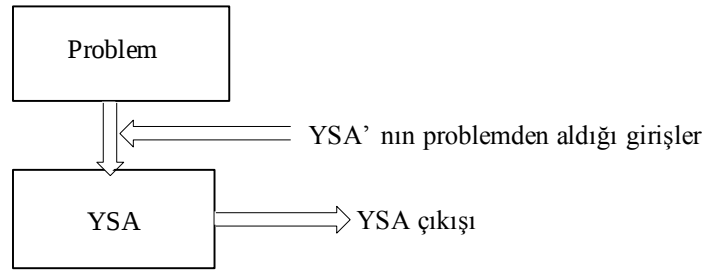


Şekil 4.14. Eğitici öğrenme blok şeması

4.2.5.2. Eđiticisiz Öğrenme

Eđiticisiz yada kendiliğinden düzenlemeli öğrenme yönteminde ađın öğreneceđi belirgin davranış örnekleri yani eđitici bir sinyal yoktur. Ađ girişleri üzerinden gerçekleştirilmek istenen bir amaca göre YSA ađırlıkları ayarlanır.

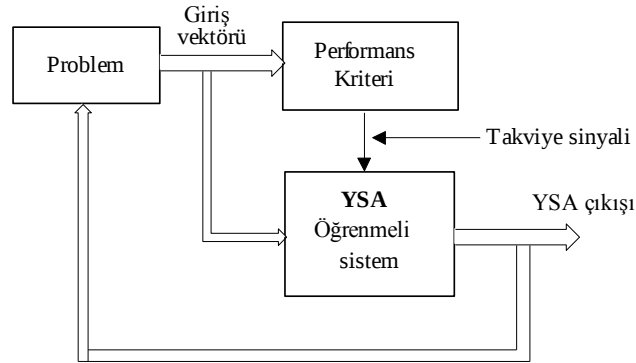
Eđiticisiz öğrenme genellikle,ađ giriş verilerini kümelemek ,sürekli zaman ađ girişlerini nitelendirmek, ađ giriş verilerini daha farklı boyutlu uzayda göstermek, ađ giriş sinyalini temsil eden özellikleri belirlemek gibi amaçlar için kullanılır. Radyal tabanlı fonksiyon ađlarında, yapay sinir ađının giriş verilerini kümeleyerek en iyi en iyi taban fonksiyonu merkezlerinin belirlenmesi , bir sinyalin öz niteliklerini çıkararak sinyal, nesne ve ses tanıma v.s gibi uygulamalarda eđiticisiz öğrenme yöntemleri kullanılmıştır.



Şekil 4.15. Eđiticisiz öğrenme blok şeması

4.2.5.3. Takviyeli Öğrenme

Takviyeli öğrenme yönteminde, yapay sinir ađının öğrenmesi için bir eđitici sinyal bulunmaz ancak, ađın davranışının uygun olup olmadığını bildiren bir öz yetenek bilgisine göre ađırlıklar ayarlanır. Bu nedenle genellikle gerçek zamanda öğrenme yöntemidir ve deneme yanılma esasına göre yapay sinir ađı öğrenir.



Şekil 4.16. Takviyeli öğrenme blok şeması

4.2.5.4. Öğrenme Algoritmaları

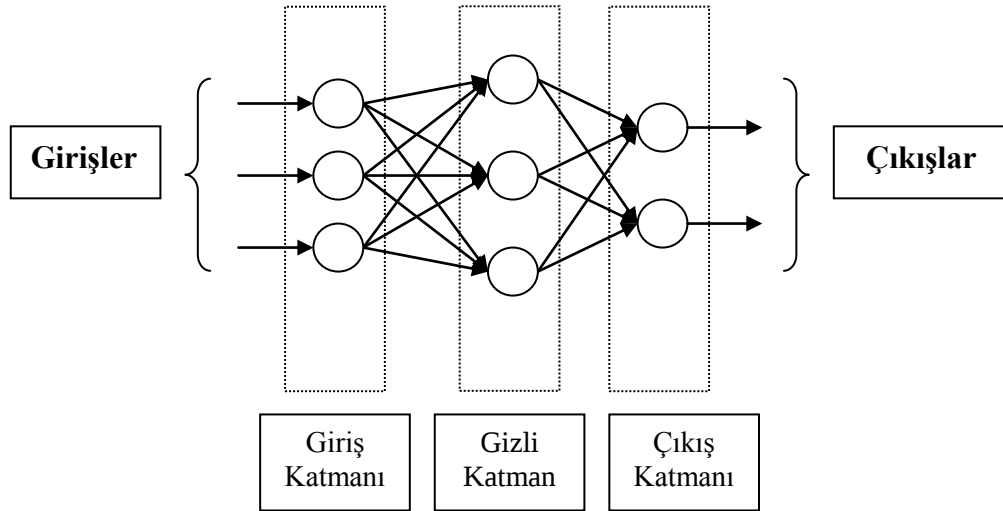
Yapay sinir ağlarında bir çok farklı öğrenme algoritması vardır. Bu bölümde geri yayılım algoritması ve bu tezde proteinlerin peptit bağ açılarının tahmininde kullanmış olduğumuz kullanmış olduğumuz Levenberg Marquadt öğrenme algoritmaları anlatılmıştır.

Geri Yayılım Algoritması

Geri yayılım algoritması, sinir ağının eğitici sınıfına giren genel bir algoritmadır. Daha öncede belirtildiği gibi girişlerle çıkışlar arasındaki hata sinyali bulunarak, ağırlıklar bu hata sinyaliyle güncellenmektedir. Hata yani $e(k)$, arzu edilen çıkış (gerçek çıkış - $y(k)$) ile sinir ağının çıkışı ($o(k)$) arasındaki farktır.

$$e(k) = y(k) - o(k)$$

Aşağıdaki şekilde birçok sinir hücresinin bir birine bağlandığı ileri yönlü çok katmanlı bir yapay sinir ağı görülmektedir. giriş katmanı ile çıkış katmanı arasındaki katman veya katmanlar gizli katman olarak adlandırılır. Sinir ağlarında kaç tane gizli katman kullanılacağı ve her bir gizli katmanda kaç nöron olacağı bugüne kadar belirlenememiştir, probleme göre değişen bu nitelikler deneme-yanılma yoluyla bulunur.



Şekil 4.17. İleri beslemeli çok katmanlı sinir ağı modeli

Şekilde çok katmanlı ağda $k+1$. katmanda i . birime net giriş;

$$n^{k+1}(i) = \sum_{j=1}^{S_k} w^{k+1}(i, j) o^k(j) + b^{k+1}(i)$$

Birim i 'nin çıkışı şöyle olacaktır.

$$o^{k+1}(i) = f^{k+1}(n^{k+1}(i))$$

M katmanlı bir ağ matris biçiminde ifade edilirse;

$$\underline{o}^0 = \underline{x}$$

$$\underline{o}^{k+1} = \underline{f}^{k+1}(\underline{W}^{k+1} \underline{o}^k + \underline{b}^{k+1})$$

$$k = 0, 1, \dots, M-1$$

Ağın temel görevi giriş-çıkış çiftleri arasındaki ilişkiyi öğrenmektir.

$$\{(\underline{x}_1, \underline{y}_1), (\underline{x}_2, \underline{y}_2), \dots, (\underline{x}_Q, \underline{y}_Q)\}$$

Ağın performansı şöyledir;

$$E = \frac{1}{2} \sum_{q=1}^Q (\underline{y}_q - \underline{o}_q^M)^T (\underline{y}_q - \underline{o}_q^M) = \frac{1}{2} \sum_{q=1}^Q \underline{e}_q^T \underline{e}_q$$

Buradaki $\underline{o}_q^M$, q . giriş ($\underline{x}_q$) olduğunda ağın çıkışıdır, $\underline{e}_q = \underline{y}_q - \underline{o}_q^M$ ise q .girişin hatasıdır.

Standart geri yayılım algoritması için yaklaşık adım düşümü algoritmasını kullanır. Performans indeksiyle şöyle yaklaşılr;

$$E = \frac{1}{2} \underline{e}_q^T \underline{e}_q$$

burada toplam karelerin toplamı tek giriş-çıkış çifti için karesel hatayla yer değiştirilir. Yaklaşık adım (eğim) düşümü algoritması bundan sonra şöyle olur;

$$\Delta w^k(i, j) = -\alpha \frac{\partial E}{\partial w^k(i, j)}$$

$$\Delta b^k(i) = -\alpha \frac{\partial E}{\partial b^k(i)}$$

burada α öğrenme oranıdır, şöyle tanımlanır;

$$\delta^k(i) \equiv \frac{\partial E}{\partial n^k(i)}$$

performans indeksinin duyarlılığı k.katmanda i. birimin net girişinde değiştirilir.

$$\frac{\partial E}{\partial w^k(i, j)} = \frac{\partial E}{\partial n^k(i)} \frac{\partial n^k(i)}{\partial w^k(i, j)} = \delta^k(i) o^{k-1}(j)$$

$$\frac{\partial E}{\partial b^k(i)} = \frac{\partial E}{\partial n^k(i)} \frac{\partial n^k(i)}{\partial b^k(i)} = \delta^k(i)$$

Hassaslık yeterliği aşağıdaki yinelemeli ilişkiyle de gösterilebilir.

$$\delta^k = \dot{F}^k(n^k) W^{k+1} \delta^{k+1}$$

burada;

$$\dot{F}^k(n^k) = \begin{bmatrix} \dot{f}^k(n^k(1)) & 0 & & 0 \\ 0 & \dot{f}^k(n^k(2)) & ... & 0 \\ . & .. & ... & .. \\ 0 & 0 & ... & \dot{f}^k(n^k(Sk)) \end{bmatrix}$$

ve

$$f^k(n) = \frac{df^k(n)}{dn}$$

$$\delta^M = -\dot{F}^M(n^M)(y_q - o_q)$$

Levenberg- Marquardt Algoritması

Geri yayılım algoritması çok kullanılmasına rağmen bazı dezavantajları bulunmaktadır. Geri yayılım sonuca çok yavaş olarak yaklaşmaktadır. Ayrıca lokal minimuma yakalanma riski de vardır. Geri yayılım, bir adım düşme algoritmasıyken, Levenberg -Marquardt algoritması Newton metoduna bir yaklaşımdır. Levenberg Marquardt algoritması, Newton metodunun hızıyla, adım düşme metodunun sağlamlığının bileşkesidir.

Levenberg Marquardt öğrenme algoritması minimumu araştırma metotlarının ikincisidir. Her bir iterasyon adımımda hata yüzeyine parabolik yaklaşımla yaklaşılr ve parabolün minimumu o adım için çözümü oluşturur. $E(x)$ fonksiyonuna sahip olduğumuzu ve x parametresine göre minimize etmek istediğimizi düşünelim. Newton metodunda şöyle olacaktır.

$$\Delta \underline{x} = -[\nabla^2 E(\underline{x})]^{-1} \nabla E(\underline{x})$$

burada $\nabla^2 E(\underline{x})$ ifadesi Hessian matrisidir ve $\nabla E(\underline{x})$ ise eğimdir. $E(\underline{x})$ 'in karelerin toplamı fonksiyonu olduğunu farz edelim.

$$E(\underline{x}) = \sum_{i=1}^N e_i^2(\underline{x})$$

bundan sonra şöyle gösterilebilir.

$$\nabla E(\underline{x}) = J^T(\underline{x}) e(\underline{x})$$

$$\nabla^2 E(\underline{x}) = J^T(\underline{x}) J(\underline{x}) + S(\underline{x})$$

burada $J(x)$ Jacobian matristir.

$$J(\underline{x}) = \begin{bmatrix} \frac{\partial e_1(\underline{x})}{\partial x_1} & \frac{\partial e_1(\underline{x})}{\partial x_2} & \dots & \frac{\partial e_1(\underline{x})}{\partial x_n} \\ \frac{\partial e_2(\underline{x})}{\partial x_1} & \frac{\partial e_2(\underline{x})}{\partial x_2} & \dots & \frac{\partial e_2(\underline{x})}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial e_N(\underline{x})}{\partial x_1} & \frac{\partial e_N(\underline{x})}{\partial x_2} & \dots & \frac{\partial e_N(\underline{x})}{\partial x_n} \end{bmatrix}$$

ve

$$S(\underline{x}) = \sum_{i=1}^N e_i(\underline{x}) \nabla^2 e_i(\underline{x})$$

Gauss-Newton metodu için $S(\underline{x}) \approx 0$ kabul edilir formül güncellenerek şu hale gelir.

$$\Delta \underline{x} = [J^T(\underline{x}) J(\underline{x})]^{-1} J^T(\underline{x}) e(\underline{x})$$

Levenberg-Marquardt modifikasyonu Gauss-Newton metodu şöyle olur.

$$\Delta \underline{x} = [J^T(\underline{x})J(\underline{x}) + \mu I]^{-1} J^T(\underline{x})e(\underline{x})$$

μ parametresi, bir adımda $E(\underline{x})$ artımı sonucunda bir faktörle (β) ile çarpılır, bir adım $E(\underline{x})$ azaltışında, μ parametresi β 'ya bölünür. μ büyükse algoritma adım düşümü ($1/\mu$ adımı ile) olur, küçük μ parametresi için algoritma Gauss-Newton olur.

Bu algoritmadaki anahtar adım Jacobian matrisinin hesaplanmasıdır. Yapay sinir ağı tasarım probleminde Jacobian matristeki terimler geri yayılım algoritmasının basitçe değiştirilmesiyle hesaplanabilir.

Burada $\underline{x} = [w^1(1,1)w^1(1,2)...w^1(S1,R)b^1(S1)w^2(1,1)...b^M(SM)]^T$ ve $N=Q \times SM$.

Standart geri yayılım algoritması terimleri şöyle hesaplanır;

$$\frac{\partial E}{\partial w^k(i,j)} = \frac{\partial \sum_{m=1}^{SM} e_q^2(m)}{\partial w^k(i,j)}$$

Levenberg-Marquardt algoritması için Jacobian matrisin elemanları şöyle hesaplanabilir.

$$\frac{\partial e_q(m)}{\partial w^k(i,j)}$$

bu terimler standart geri yayılım algoritması kullanılarak son katmanda bir değişiklik ile hesaplanabilir.

$$\Delta^M = -\dot{F}^M(n^M)$$

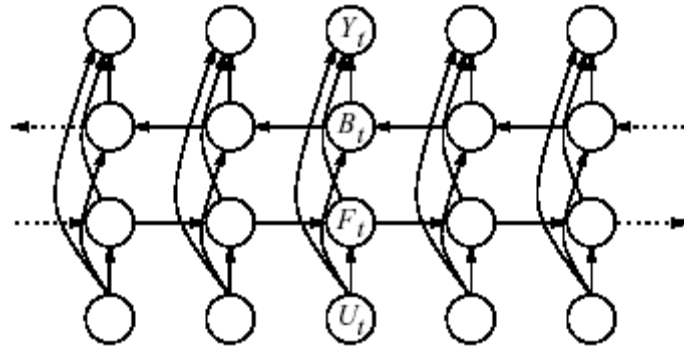
Sonuç olarak Levenberg-Marquardt algoritması çok hızlı olarak çözüme ulaşmasına rağmen çok fazla bellek gerektirmektedir. Geri yayılım algoritması ise sonuca yavaş ulaşmakta ve daha az bellek gerektirmektedir.

5. PROTEİN YAPI TAHMİNİ İÇİN UYGULANAN DİĞER YÖNTEMLER

Bu bölümde protein yapı tahmini için kullanılan diğer yöntemlerden bahsedilecektir. Şu ana kadar bu konuda yapay sinir ağları da dahil olmak üzere yapılmış olan bir çok çalışma vardır. Bunların bir çoğu katlanmanın şekli (coil/halka veya helix/sarmal) esas alınarak yapılmıştır. Esasında literatür incelenmesi yapılacak olursa amino asitlerin içermiş oldukları atomların arasındaki bağ açılarından yola çıkarak yaklaşım getirmeye çalışan yoktur. Bölüm 6'da da genişçe bahsedileceği gibi bu tezde özellikle atomlar arası bağ açılarından yola çıkılmıştır. Literatürde farklı yöntemler kullanılmıştır. Aşağıda kullanılan diğer yöntemlerden genel bir şekilde bahsedilmiştir.

Kullanılan yöntemleri ve elde edilen başarı oranlarını sıralayacak olursak:

1-Bidirectional Dynamic for Protein Secondary Structure Prediction [9] Bu çalışmada helix ve coil yapılar üzerinde çalışılmıştır. Bu yöntemde Bayesian Network'ler ve olasılık hesapları kullanılmıştır. Proteinlerin mevcut konformasyonları içerisinde birbiriyle komşuluk ve bağ ilişkisi olan taraflardan yola çıkılarak olasılık formülleri yardımıyla tahminler yapılmaya çalışılmıştır.



Şekil 5.1. Bayesian network modeli

BRNN ([Bi-Directional Recurrent Neural Networks](#)) 'den Rost ve Sender örnekleri üzerinde çalışılmıştır. Yaklaşık %75 başarı elde edilmiştir.

2-Use of Multiple Alignment in Protein Secondary Structure Prediction [10] Bu çalışmada da Helix , Coil ve Extended Strand yapılar üzerinde çalışılmıştır.

Kristolografik protein dimerleri üzerinde DSSP algoritması yardımıyla çözüme gidilmeye çalışılmıştır. Farklı ardışıl örnekleri sayesinde çoklu bir hizalama yardımı ile çözüme gidilmeye çalışılmıştır. Brook haven databank' tan alınan 20 protein örneği kullanılmıştır. Yaklaşık %69 başarı elde edilmiştir.

3- Two-Stage Multi-Class Support Vector Machines To Protein Secondary Structure Prediction [11] Bu çalışmada PSI-BLAST yöntemi ile RS126 ve CB396 örnekleri üzerinde Helix , Coil ve Extended Strand yapıları üzerinde çalışılmıştır. Birkaç birleştirilmiş binary sınıflandırıcı tarafından yapılandırılmış çok sınıflı destek vektör makineleri kullanılmış bununla çözüme yaklaşım getirilmeye çalışılmıştır. Yaklaşık %77 başarı elde edilmiştir.

4- HYPROSP: A hybrid protein secondary structure prediction algorithm—a knowledge-based approach [12] Bu çalışmada PSI-BLAST yöntemi ile Helix , Extended Strand ve diğer yapılar üzerinde çalışılmıştır. Üç adımda tamamlanır. Birinci adımda eğitim iterasyonları tamamlanır. İkinci adımda her örnek için benzer peptitler belirlenir. Son adımda ise skor matrisleri yardımıyla benzer yapıların yakınlıkları değerlendirilir. Sonuç tüm peptitlerin benzer bulunan örneklere oranı olarak ifade edilmiştir. Yaklaşık %80 başarı elde edilmiştir.

5- Context Sensitive Vocabulary And its Application in Protein Secondary Structure Prediction [13] Bu çalışmada BLAST algoritması yöntemiyle RS126 ve CB513 protein örnekleri üzerinde çalışılmıştır. Amino asitlerin ardışıl şekilde gelmeleri sonucu ortaya çıkan heliks ve coil yapılar bir tabloda tutularak numaralandırılmıştır. Daha sonrasında tablodan faydalanılarak gelen örnekler değerlendirilmiştir. Yaklaşık % 70 başarı elde edilmiştir.

	$N = 1$	$N = 2$
N-gram	E, C, P, V, N, C, I	EC, CP, PV, VN, NC, CI, IQ
Context Sensitive N-gram	E-3, C-2, P-1, V+0, N+1, C+2, I+2	EC-3, CP-2, PV-1, VN+0, NC+1, CI+2, IQ+3

Şekil 5.2. Yapı pozisyon tablosu

6- Prediction of Protein Secondary Structure Using Genetic Programming [14] Bu çalışmada Helix, Coil ,Extended Strand yapılar üzerinde çalışılmıştır.

Bu çalışmada amino asit ardışıkları bir kromozom; heliks ve coil yapılar ayrı bir kromozomda kodlanmıştır. İki farklı kromozom üzerine uygulanan genetik işlemler sonucunda çözüme gidilmiştir. Yaklaşık % 76 başarı elde edilmiştir.

7-Parallel Stochastic Search for Protein Secondary Structure Prediction [15] Bu çalışmada EDA algoritması yardımıyla CB513 data seti üzerinde çalışılmış ve ikincil yapının olasılık tahminleri üzerinde durulmuştur. Bunun için n adet amino asit içeren bir pencere tanımlanmış ve bayesian sınıflandırıcı fonksiyon tanımlanmıştır. Daha sonra ise her k. pozisyonda farklılık arz eden amino asit ardışıkları için olasılık değerleri tanımlanmıştır. Yaklaşık % 70 oranında başarı elde edilmiştir.

8-Protein Structure Prediction Using Decision Lists[16]

Bu çalışmada Greedy decision list learner algoritması CB513 data seti üzerine uygulanmıştır. Kural tabanları oluşturularak işlem yapılmıştır. Yaklaşık %69 civarı bir başarı elde edilmiştir.

9- Protein secondary structure prediction with a neural network [17] Bu çalışmada protein yapı tahmini için bir neural network modeli kullanılmıştır. 20 değişik amino asit ve birde sonlardaki null değer (üst üste binme durumlarında) ile beraber 21 giriş mevcut olup pencereleme metodu ile 17 pencerelik bir boyut seçilmiştir.

Yapay sinir ağına girişten rasgele amino asitler gelmektedir. Pencere boyutu 17 olarak seçildiğinden ilk 17 amino asitten 9. sıradaki merkezde olacak şekilde 0 ; ilk 8 tanesi -8 'e kadar sonraki 8 de +8'e kadar numaralandırılır. Oluşturulan tabloda tüm amino asitler pozisyonlarına göre sıralanmakta ve alabilecekleri değerler tabloya yerleştirilmektedir.

Yapay sinir ağının iki çıkışı mevcuttur. Çıkış =(0,0) ise yapı coil , Çıkış = (0,1) ise yapı sheet , Çıkış=(1,0) ise yapı helix durumundadır. Bu yöntem ile yaklaşık % 79 başarı elde edilmiştir.

10-Secondary Structure Prediction Of Hemoglobin By Using Combined Neural Networks [18]

Bu çalışmada hemoglobine proteininin yapısını tahmin için birleştirilmiş yapay sinir ağları metodu kullanılmıştır.

Proteinlerin yapısında bulunan 20 farklı amino asit 20 bitlik string ile yapılarının helix veya coil oluşu ise 2 bit ile kodlanmıştır.

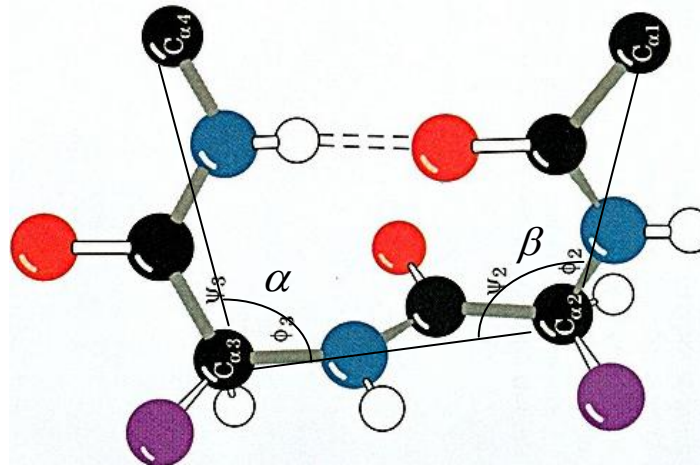
Alanine = 10000000000000000000 Helix(H) = 1,2

Giriřten yapay sinir aęlarına gelen amino asit ardışıkları pencereleme metoduna tabi tutulmaktadır. Pencerede ortadaki amino asit $W = 2*r+1$ formülü yardımıyla bulunarak r . sıraya denk gelen amino asit pencerenin ortası olup yapısına bakılıp sonra pencerede r . den sonra sıra gelen amino asit pencere ortası seçilip aynı işlem ardışıl sonuna kadar devam etmektedir.

İşlem sonunda gerçekte olması gereken helix ve coil sayısı ile yapay sinir aęının bulduęu helix ve coil sayısı oranlandığı zaman başarı oranı ortaya çıkmaktadır. Bu çalışmada farklı pencere boyutları için sonuçlar elde edilmiştir. 20 alfa-141 ve 20 beta-146 hemoglobin zincirleri için elde edilen sonuçlar yaklaşık % 90 civarındadır.

6. GELİŞTİRİLEN YÖNTEM VE SONUÇLARI

Bu tezde proteinlerin sekonder yapılarının tahmini için peptitlerin merkezil alfa karbonları arasında oluşan bağ açılarından yola çıkılmıştır. Bu amaçla değişik proteinlerden alınan 290 tane açı örneği tasarlanan yapay sinir ağına verilerek yapay sinir ağının eğitimi tamamlanmıştır. Daha sonra 1bbe, 1a89 ve 1a9a kolajen proteinleri üzerinde test uygulanmış ve sonuçlar elde edilmiştir.

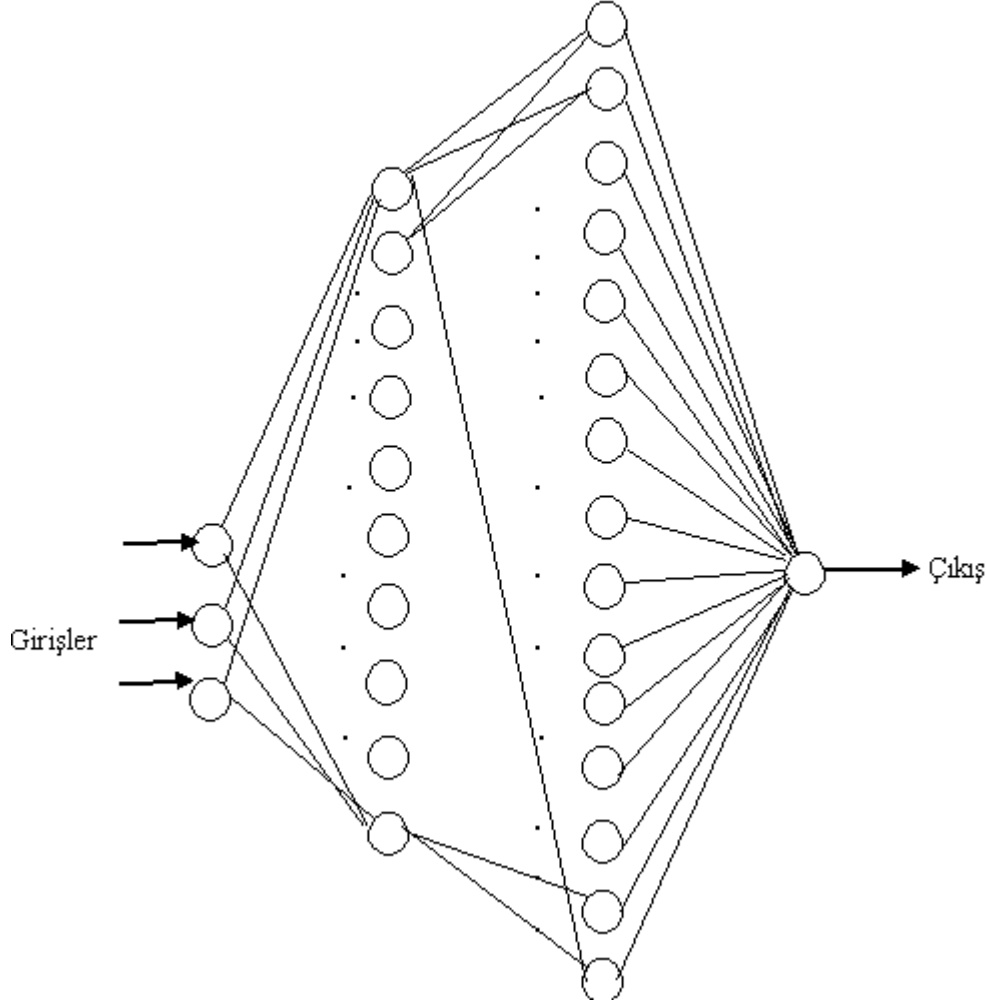


Şekil 6.1. Tripeptit yapıda kullanılan bağ açıları

Şekil 6.1. de görüldüğü gibi α ve β açıları sıralı amino asitlerin merkezil karbonları arasında oluşan açı olarak seçilmiştir. Burada α 1-2-3 , β ise 2-3-4 numaralı merkezil karbonlar arasında oluşan açılardır. Bu şekilde elde edilen tüm açılar değerlendirilerek bir proteinin tüm sıralı merkezil karbonlarının oluşturmuş oldukları bağ açıları tahmin edilmeye, bu sayede proteinin sekonder yapısı ortaya konulmaya çalışılmıştır. Şekil 6.1.'den de görüleceği gibi, esas sıralı açılar merkezil karbonlar arasında değildir. Sırayla bakılacak olursa, merkezil karbonlardan önce arada oluşan farklı açılarda mevcuttur. Fakat bu açılar her amino asit için sabit kabul edilip, esas ölçü merkezil karbonlar arası oluşan açı olarak kabul edilmiştir.

Bu tezde protein data banklardan alınan protein örnekleri “PDB” uzantılı olup bu dosyalar Swiss PDB Viewer V3.7. programı kullanılarak örnekler üzerinde çalışmalar yapılmıştır. Bu programla elde edilen açılar oluşturulan yapay sinir ağına eğitim seti olarak verilmiştir.

Yapay sinir ağımızın yapısı ise Şekil 6.2. de olduğu gibidir.



Şekil 6.2.Kullanılan yapay sinir ağı modeli

Yapay sinir ağının girişinden sıra ile amino asitler gelir. Bu sıra, proteinin birincil yapısında bulunan amino asit zincirinin kendisidir. Gelen sıra bozulmadan üçer üçer alınarak yapay sinir ağına giriş olarak verilecektir.

Yapay sinir ağı nümerik değerlerle çalıştığından amino asitleri nümerik kodlarla kodlanmıştır. Testimizi oluşturan proteinlerin ihtiva etmiş oldukları 11 değişik amino asit mevcuttur. Bu nedenle tüm amino asitleri kullanmak zorunda kalmadık. Sadece lazım olan 11 farklı amino asit için bir değişim tablosu hazırlanmıştır.

Tablo 6.1. Amino asitlerin nümerik değere 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 290 farklı açı örneğimiz ve eğitim sonucunda yapay sinir ağından elde edilen gerçek açı değerlerimiz 109.10 ile 135.10 arasındadır. Bu aralık 11 farklı amino asit sayısına bölünüp taban fonksiyonu alındığında 3 sayısı bulunur. 109.10 ile 135.10 arasını 3 derece aralıkla açı sınıflarına bölecek olursak 9 farklı açı sınıfı elde ederiz.

$$X = \lceil (135.10 - 109.10) / 11 \rceil = 3$$

Açı sınıflarımız ise tablo 6.2. de olduğu gibidir.

Tablo 6.2. 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

Yapay sinir ağının eğitiminde farklı proteinlerden elde edilmiş 41 değişik açılı tipi kullanılmıştır. Bu açılı tipleri ise aşağıdaki gibidir.

1-2-6 , 1-6-1, 2-6-4, 2-6-5, 2-6-7, 2-6-9, 2-9-6, 3-9-6, 4-2-6, 5-2-6, 5-9-6 , 6-1-2, 6-1-6, 6-2-9, 6-3-9, 6-4-2, 6-5-2, 6-5-9, 6-7-9, 6-7-10, 6-8-9, 6-9-2, 6-9-9, 6-9-10, 6-11-4, 7-9-6, 7-10-6, 8-9-6, 9-2-6 , 9-6-1, 9-6-5, 9-6-7, 9-6-9, 9-6-11, 9-9-2, 9-9-6, 9-10-6, 9-10-9, 10-6-3, 10-6-9, 10-9-9

41 adet farklı açılı çeşidi kullanılmakta olup bunların her biri farklı protein örneklerinden farklı değerlerde seçilerek yapay sinir ağının farklı değerlere de adapte olması sağlanmıştır. Örneğin 1-2-6 açılı için farklı proteinlerden 8 adet örnek alınmıştır.

Tablo 6.3. Eğitim örnekleri

1	2	6	117.9
1	2	6	142.7
1	2	6	125.31
1	2	6	142.23
1	2	6	142.23
1	2	6	142.23
1	2	6	142.22
1	2	6	123.67
1	6	1	112.69
1	6	1	115.62
1	6	1	113.76
1	6	1	117.04
1	6	1	113.82
1	6	1	117.22
1	6	1	114.16
1	6	1	117.02
1	6	1	115.3
1	6	1	117.76
2	6	4	117.74
2	6	4	119.04
2	6	4	117.8
2	6	4	118.2
2	6	4	118.73
2	6	5	119.15
2	6	5	118.37
2	6	5	120.24
2	6	5	119.3
2	6	5	118.44
2	6	7	116.86
2	6	7	108.12
2	6	7	108.86
2	6	7	112.81
2	6	7	110.11
2	6	9	120.81
2	6	9	117.59
2	6	9	119.84
2	6	9	119.48

2	6	9	118.95
2	6	9	121.82
2	6	9	119.36
2	6	9	117.96
2	6	9	119.12
2	6	9	118.03
2	9	6	117.41
2	9	6	123.02
2	9	6	123.04
2	9	6	123.79
2	9	6	124.96
2	9	6	111.08
2	9	6	112.07
3	9	6	116.98
3	9	6	108.82
3	9	6	110.98
3	9	6	111.7
3	9	6	110.82
3	9	6	113.33
3	9	6	116.08
4	2	6	116.9
4	2	6	118.13
4	2	6	115.3
4	2	6	116.74
4	2	6	117.65
5	2	6	116.81
5	2	6	117
5	2	6	116.92
5	2	6	117.14
5	2	6	117.04
5	9	6	117.57
5	9	6	115.95
5	9	6	116.71
5	9	6	115.32
5	9	6	115.6
5	9	6	112.85
5	9	6	113.3
5	9	6	113.69
5	9	6	113.78
5	9	6	115.46
6	1	2	119.72
6	1	2	140.23
6	1	2	123.27
6	1	2	139.99
6	1	2	141.1
6	1	2	91.16
6	1	2	90.89
6	1	6	119.76
6	1	6	117.88
6	1	6	118.57
6	1	6	116.88
6	1	6	118.54

6	1	6	117.3
6	1	6	118.26
6	1	6	117.13
6	1	6	118.25
6	1	6	119.15
6	2	9	117.55
6	2	9	129.52
6	2	9	125.33
6	2	9	125.54
6	2	9	113.82
6	2	9	113.8
6	3	9	118.81
6	3	9	107.04
6	3	9	108.9
6	3	9	125.67
6	3	9	125.08
6	3	9	125.72
6	3	9	125.23
6	4	2	116.38
6	4	2	118.12
6	4	2	118.29
6	4	2	118.31
6	4	2	117.7
6	5	2	117.45
6	5	2	118.56
6	5	2	118.58
6	5	2	119.49
6	5	2	115.97
6	5	2	116.5
6	5	2	126.08
6	5	2	100.3
6	5	2	115.19
6	5	9	119.98
6	5	9	110.97
6	5	9	116.8
6	5	9	115.82
6	5	9	115.8
6	7	9	118
6	7	9	136.72
6	7	9	136.71
6	7	9	104.46
6	7	9	116.21
6	7	9	122.13
6	7	10	117.66
6	7	10	123.46
6	7	10	129.73
6	7	10	129.68
6	7	10	96.84
6	7	10	101.11
6	8	9	119.47
6	8	9	120.37
6	8	9	119.55

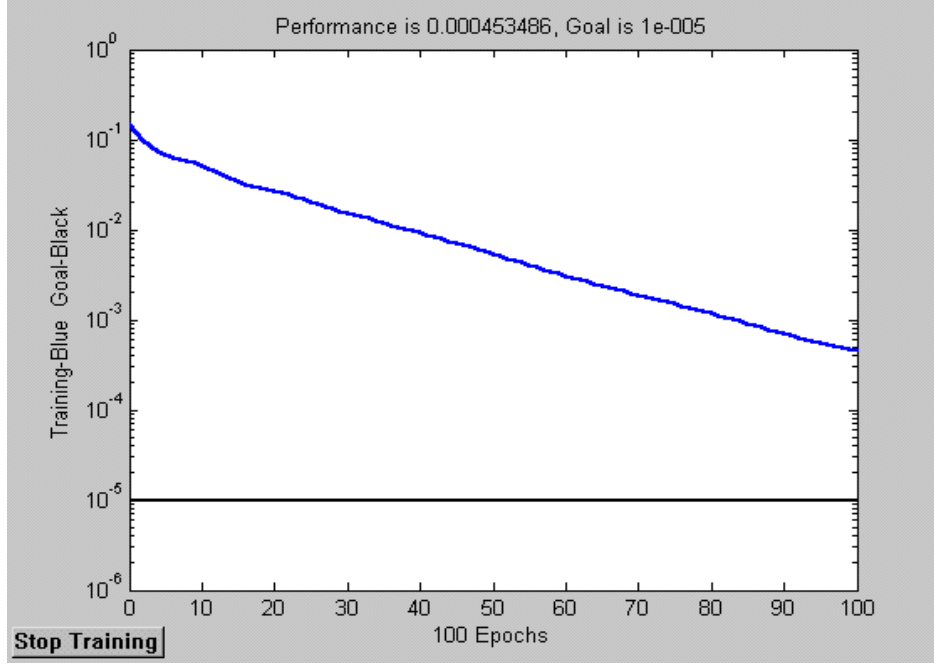
6	8	9	117.19
6	8	9	118.11
6	9	2	118.43
6	9	2	118.2
6	9	2	119.57
6	9	2	115.86
6	9	2	117.3
6	9	2	117.26
6	9	2	117.8
6	9	2	121.25
6	9	2	116.7
6	9	2	117.98
6	9	2	120.76
6	9	2	116.67
6	9	2	119.49
6	9	2	118.83
6	9	9	121.53
6	9	9	121.53
6	9	9	121.54
6	9	9	121.53
6	9	9	121.54
6	9	9	121.51
6	9	9	121.53
6	9	9	124.1
6	9	9	122.9
6	9	9	122.34
6	9	9	119.74
6	9	9	112.95
6	9	9	119.4
6	9	9	123.28
6	9	9	119.73
6	9	10	117.36
6	9	10	116.13
6	9	10	119.07
6	9	10	119.53
6	9	10	118.1
6	11	4	118.68
6	11	4	107.05
6	11	4	107.6
6	11	4	118.99
6	11	4	89.75
6	11	4	110.88
7	9	6	118.68
7	9	6	107.74
7	9	6	117.16
7	9	6	111.16
7	9	6	106.81
7	9	6	109.47
7	9	6	107.62
7	9	6	107.82
7	10	6	116.28
7	10	6	109.35

7	10	6	109.49
7	10	6	110.25
7	10	6	118.55
7	10	6	142.72
8	9	6	116.14
8	9	6	117.81
8	9	6	116.29
8	9	6	117.77
8	9	6	118.01
9	2	6	116.7
9	2	6	117.06
9	2	6	117.8
9	2	6	116.96
9	2	6	116.71
9	2	6	117
9	2	6	116.53
9	2	6	118.07
9	2	6	117.23
9	2	6	117.3
9	6	1	118.24
9	6	1	133.29
9	6	1	133.47
9	6	1	133.38
9	6	1	81.67
9	6	5	117.8
9	6	7	116.34
9	6	7	90.33
9	6	7	95.31
9	6	7	148.48
9	6	7	148.49
9	6	7	95.58
9	6	7	130.16
9	6	7	93
9	6	9	120.55
9	6	9	120.54
9	6	9	120.55
9	6	9	120.54
9	6	9	120.54
9	6	9	120.53
9	6	9	127.1
9	6	9	125.47
9	6	9	123.73
9	6	9	126.85
9	6	9	129.77
9	6	9	123.22
9	6	9	121.18
9	6	9	120.5
9	6	9	118.21
9	6	9	118.43
9	6	9	118.05
9	6	9	119.02
9	6	9	118.58

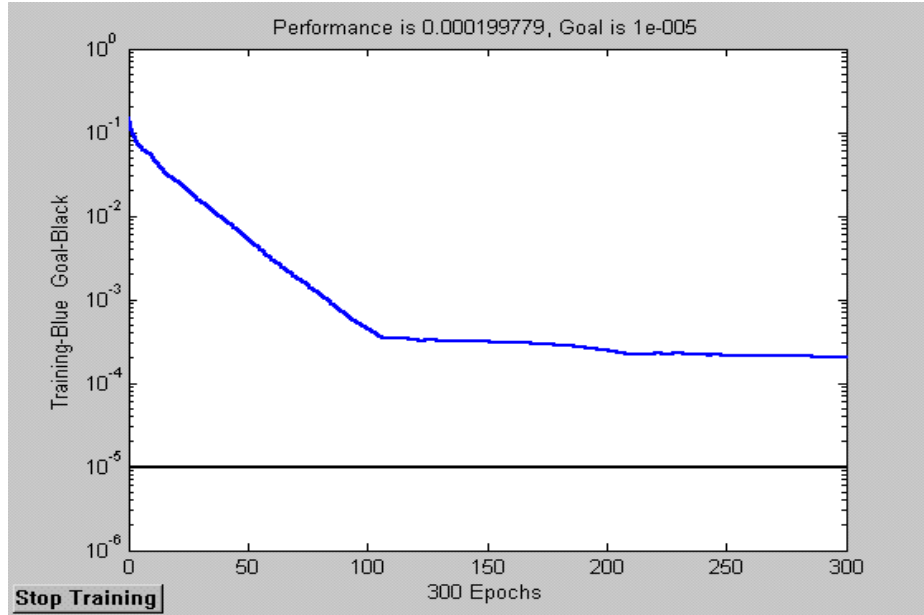
9	6	11	117.75
9	6	11	98.53
9	6	11	88.65
9	6	11	138.11
9	6	11	118.75
9	6	11	117.57
9	9	2	117.56
9	9	6	119.85
9	9	6	119.88
9	9	6	119.85
9	9	6	119.87
9	9	6	119.85
9	9	6	119.86
9	9	6	122.09
9	9	6	122.04
9	9	6	120.05
9	9	6	118.9
9	9	6	121.69
9	9	6	119.09
9	9	6	118.42
9	9	6	119.18
9	10	6	116.59
9	10	6	116.76
9	10	6	118.17
9	10	6	117.01
9	10	9	116.55
9	10	9	102.82
9	10	9	125.27
9	10	9	125.21
9	10	9	130.12
9	10	9	116.35
10	6	3	117.49
10	6	3	126.29
10	6	3	124.67
10	6	9	120.14
10	6	9	119.68
10	6	9	119.6
10	6	9	118.79
10	9	9	119.86
10	9	9	114.41
10	9	9	116.46
10	9	9	89.55
10	9	9	105.75
10	9	9	116.5
10	9	9	117.68

Yapay sinir ağı modelimiz Matlab 6.5 programının Neural Network toolbox kısmında oluşturulmuştur. Eğitim örnekleri verilerek eğitim tamamlanmış ve sonuçlar yine bu programda

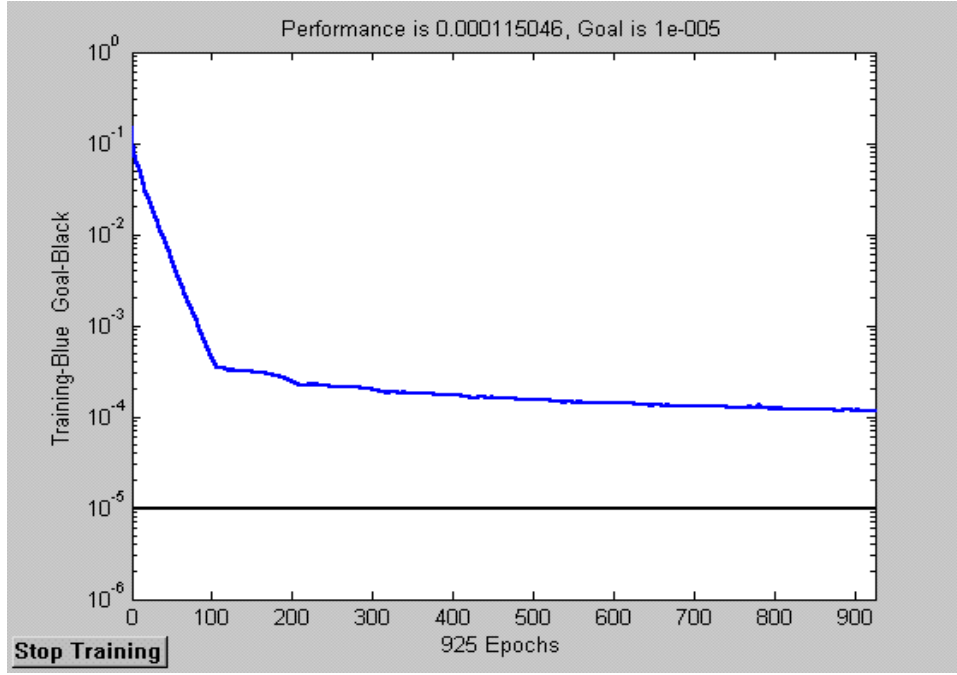
elde edilmiştir. Aşağıda eğitim esnasında yapay sinir ağının hata değerinin giderek nasıl düştüğü görülmektedir.



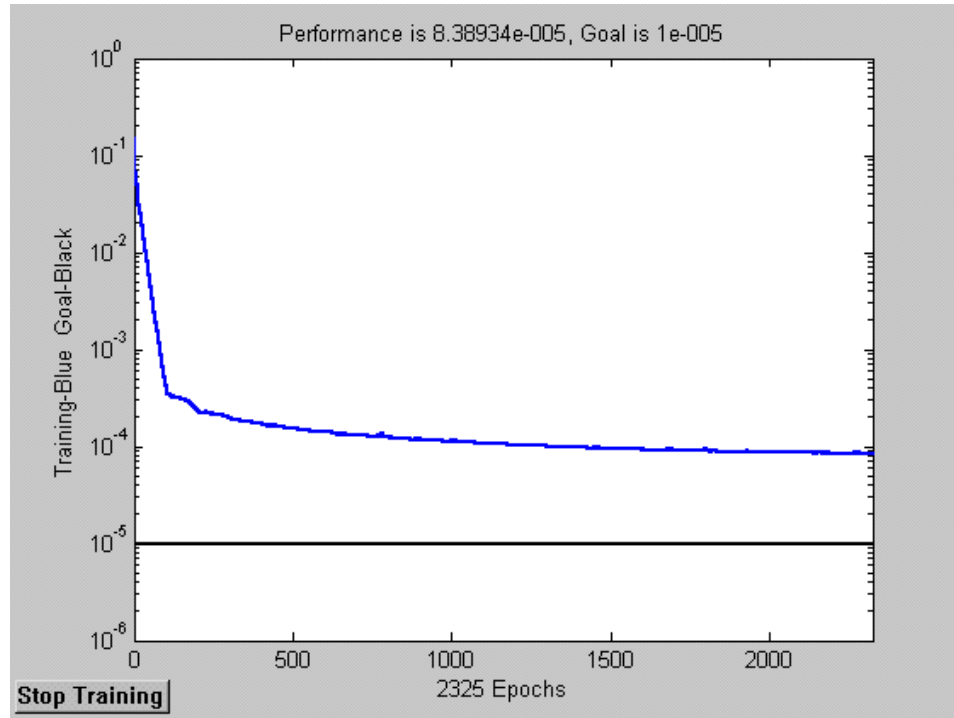
Şekil 6.3. Yapay sinir ağı eğitimi birinci örnek şekil



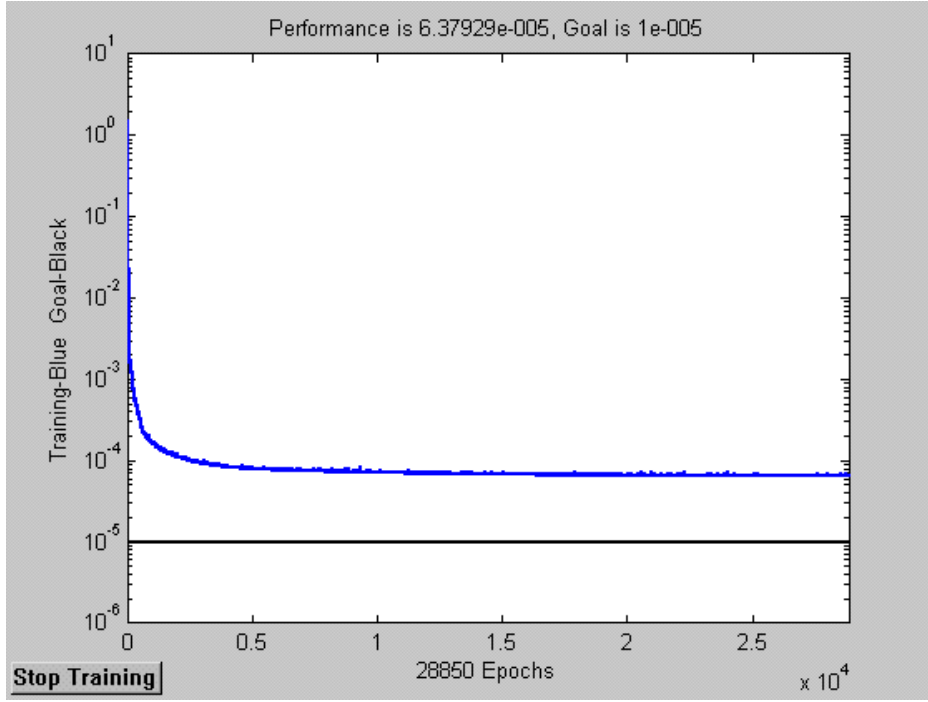
Şekil 6.4. Yapay sinir ağı eğitimi ikinci örnek



Şekil 6.5. Yapay sinir ağı eğitimi üçüncü örnek



Şekil 6.6. Yapay sinir ağı eğitimi dördüncü örnek



Şekil 6.7. Yapay sinir ağı eğitimi beşinci örnek

Yapay sinir ağının eğitimi tamamlandıktan sonra elde edilen değerler, gerçek açı değerleri, açı sınıfları ve hata değeri Tablo 6.4., 6.5., 6.6., 6.7., 6.8., 6.9. ile sunulmuştur.

Tablo 6.4. 1bbe proteini açı değerleri

Gerçek Açı	121.53	119.85	120.55	121.53	119.88	120.54	121.54	119.85	120.55	121.53
YSA Açısı	121.3	120.3	121.8	121.3	120.3	121.8	121.3	120.3	121.8	121.3
Hata	0.23	0.45	1.25	0.23	0.42	1.26	0.24	0.45	1.25	0.23
Gerçek Açı	121.54	119.87	120.54	121.51	119.85	120.54	121.53	119.86	120.53	121.53
YSA Açısı	121.3	120.3	121.8	121.3	120.3	121.8	121.3	120.3	121.8	121.3
Hata	0.24	0.43	1.26	0.21	0.45	1.26	0.23	0.44	1.27	0.23
Gerçek Açı	121.53	119.86	120.53	121.52	119.85	120.52	121.53	119.86	120.52	121.53
YSA Açısı	121.3	120.3	121.8	121.3	120.3	121.8	121.3	120.3	121.8	121.3
Hata	0.23	0.44	1.27	0.22	0.45	1.28	0.23	0.44	1.28	0.23

Ortalama açı hatası : 0.6

Tablo 6.5. 1bbe proteini açi sınıfları

Gerçek Açı Sınıfı	A5	A4	A4	A5	A4	A4	A5	A4	A4	A5	A5	A4	A4	A5	A4	A4	A5	A4
YSA Açısı	A5	A4	A5	A5	A4	A5	A5	A4	A5	A5	A5	A4	A5	A5	A4	A5	A5	A4
Bağıl hata			0.8			0.8			0.8				0.8			0.8		
Gerçek Açı	A4	A5	A5	A4	A4	A5	A4	A4	A5	A4	A4	A5						
YSA Açısı	A5	A5	A5	A4	A5	A5	A4	A5	A5	A4	A5	A5						
Bağıl hata	0.8				0.8			0.8			0.8							

Ortalama bağıl hata : 0.8**Tablo 6.6. 1a89 proteini açi değerleri**

Gerçek Açı	119.55	116.29	118.05	117.26	116.71	117.80	118.29	115.30	118.95	119.07	116.76	119.68
YSA Açısı	118.20	117.20	121.80	118.50	117.10	118.40	117.90	117.10	119.60	118.20	117.10	120.30
Hata	1.35	0.91	3.75	1.24	0.39	0.6	0.39	1.8	0.65	0.87	0.36	0.62
Gerçek Açı	117.80	117	120.24	118.58	116.92	121.82	121.25	117.19	117.77	119.02	116.70	116.53
YSA Açısı	118.50	117.10	119.80	116.60	117.40	119.60	118.50	118.20	117.20	121.80	118.50	117.10
Hata	0.7	0.1	0.44	1.98	0.48	2.22	2.75	1.01	0.57	2.78	1.8	0.57
Gerçek Açı	118.20	118.31	116.74	119.36	119.53	118.17	119.60	117.98	118.07	119.30	119.49	117.14
YSA Açısı	118.40	117.90	117.10	119.60	118.20	117.10	120.30	118.50	117.10	119.80	116.60	117.40
Hata	0.2	0.41	0.36	0.24	1.33	1.07	0.7	0.58	0.97	0.5	2.89	0.26
Gerçek Açı	117.96	120.76	118.11	118.01	118.58	116.67	117.23	118.73	117.70	117.65	119.12	118.10
YSA Açısı	119.60	118.50	118.20	117.20	121.80	118.50	117.10	118.40	117.90	117.10	119.60	118.20
Hata	1.64	2.26	0.09	0.81	3.22	1.88	0.13	0.33	0.2	0.55	0.48	0.1
Gerçek Açı	117.01	118.79	119.49	117.30	118.44	115.97	117.04	118.03	118.83			
YSA Açısı	117.10	120.30	118.50	117.10	119.80	116.60	117.40	119.60	118.50			
Hata	0.09	1.51	0.99	0.2	1.36	0.63	0.36	1.57	0.33			

Ortalama açi hatası : 0.99

Tablo 6.7. 1a89 proteini açısı sınıfları

Gerçek Açısı	A4	A3	A4	A3	A3	A3	A4	A3	A4	A4	A3	A4	A3	A3	A4	A3	A3	A5
YSA Açısı	A4	A3	A5	A4	A3	A4	A3	A3	A4	A4	A3	A4	A4	A3	A4	A3	A3	A4
Bağıl h.			0.8	0.5		0.4	0.1						0.5					1.4
Gerçek Açısı	A5	A3	A3	A4	A3	A3	A4	A4	A3	A4	A4	A4	A4	A3	A4	A4	A4	A3
YSA Açısı	A4	A4	A3	A5	A4	A3	A4	A3	A3	A4	A4	A3	A4	A4	A3	A4	A3	A3
Bağıl h.	2.5	0.2		0.8	0.5			0.1				0.9		0.5	0.9		1.4	
Gerçek Açısı	A3	A4	A4	A4	A4	A3	A3	A4	A3	A3	A4	A4	A3	A4	A4	A3	A4	A3
YSA Açısı	A4	A4	A4	A3	A5	A4	A3	A4	A3	A3	A4	A4	A3	A4	A4	A3	A4	A3
Bağıl h.	1.6			0.8	0.8	0.5												
Gerçek Açısı	A3	A4	A4															
YSA Açısı	A3	A4	A4															
Bağıl h.																		

Ortalama bağıl hata : 0.8**Tablo 6.8. 1a9a proteini açısı değerleri**

Gerçek Açısı	119.47	116.14	118.21	118.43	116.70	117.74	116.38	116.90	120.81	117.36	116.59	120.14
YSA Açısı	118.20	117.20	121.80	118.50	117.10	118.40	117.90	117.10	119.60	118.20	117.10	120.30
Hata	1.27	0.96	3.59	0.07	0.4	0.66	1.52	0.2	1.21	0.84	0.61	0.16
Gerçek Açısı	118.20	117.06	119.15	117.45	116.81	117.59	119.57	117.55	117.41	117.80	119.98	117.57
YSA Açısı	118.50	117.10	119.80	116.60	117.40	119.60	118.50	121.10	118.30	118.10	116.20	114.90
Hata	0.3	0.04	0.65	0.85	0.59	2.01	1.07	3.55	0.89	0.3	3.78	2.67
Gerçek Açısı	116.34	118	118.68	118.24	119.72	117.90	116.86	117.66	116.28	117.49	118.81	116.98
YSA Açısı	114.80	123.10	111.40	120.40	121.10	135.10	111.50	116.60	118.40	122.80	120	114.50
Hata	1.54	5.1	7.28	2.16	1.38	17.2	5.34	1.04	2.12	5.31	1.19	2.48
Gerçek Açısı	117.75	118.68	120.37	117.81	118.43	115.86	117.80	119.04	118.12	118.13	119.81	116.13
YSA Açısı	113.30	109.10	118.20	117.20	121.80	118.50	117.10	118.40	117.90	117.10	119.60	118.20
Hata	4.45	9.58	2.17	0.61	3.37	2.64	0.7	0.64	0.22	1.03	0.21	2.07
Gerçek Açısı	116.55	119.86	117.56	116.96	118.37	118.56	117	119.48	117.30			
YSA Açısı	119.80	117.70	117.20	117.10	119.80	116.60	117.40	119.60	118.50			
Hata	3.25	8.16	0.36	0.14	1.43	1.96	0.4	0.12	1.2			

Ortalama açısı hatası : 2.19

Tablo 6.9. 1a9a proteini aç ı sınıfları

Gerçek Aç ı s.	A4	A3	A4	A4	A3	A3	A3	A3	A4	A3	A3	A4	A4	A3	A4	A3	A3	A3
YSA	A4	A3	A5	A4	A3	A4	A3	A3	A4	A4	A3	A4	A4	A3	A4	A3	A3	A4
Bağı l hata			0.8			0.4				0.2								1.6
Gerçek Aç ı s.	A4	A3	A3	A3	A4	A3	A3	A4	A4	A4	A4	A3	A3	A3	A3	A3	A4	A3
YSA	A4	A5	A4	A4	A3	A2	A2	A5	A1	A4	A5	A9	A1	A3	A4	A5	A4	A2
Bağı l hata		3.1	0.3	0.1	1.8	0.1	0.2	2.1	6.0		0.1	17.1	3.0		0.4	4.8		0.5
Gerçek Aç ı s.	A3	A3	A4	A3	A4	A3	A3	A4	A4	A4	A4	A3	A3	A4	A3	A3	A4	A4
YSA	A2	A1	A4	A3	A5	A4	A3	A4	A3	A3	A4	A4	A4	A3	A3	A3	A4	A3
Bağı l hata	1.7	6.0			0.8	0.5			0.1	0.9		0.2	1.8	0.3				1.4
Gerçek Aç ı s.	A3	A4	A3															
YSA	A3	A4	A4															
Bağı l hata			0.5															

Ortalama bağı l hata : 1.96

Tablolardan elde edilen ortalama aç ı hatası değ eri ař ağı daki formülle elde edilmiř tir.

$$\text{Ortalama aç ı hatası} = \left(\sum_{i=1}^n (\text{Gerçek aç ı değ eri}) - (\text{yapay sinir ağı değ eri}) \right) / n$$

Burada n değ eri proteinin iç erdiğı aç ı sayısıdır.

Ortalama bağı l hata değ eri ise ař ağı daki ř ekilde hesaplanmıř tir.

$$\text{Ortalama bağı l hata} = \left(\sum_{i=1}^n (|(\text{Yapay sinir ağı değ eri}) - (\text{En yakın sınıfın sınır değ eri})|) \right) / n$$

Burada yapılmak istenen ř udur : Hesaplanan aç ı sınıfı gerçek sınıf değ eri ile farklı olduğı durumda en yakın aç ı değ erinden farkına bakılıp fark hesap edilir. Mutlak değ er alınmasının sebebi ise, bazen hesaplanan değ erin bir en yakın aç ı sınıfının sınır değ erinden küçük olmasındandır. O durumda değ er negatif çıkacaktır.

Sınıf değ erleri bu kadar keskin değ il de biraz hata payı ile kabul edilebilir olursa mutlaka doğı ruluk oranları değ iř ecektir. Bunun iç in 0.2, 0.5 ve 1 derecelik hata paylarını düş ünerek tablolarımızı yeniden gözden geç irecek olursak:

1bbe proteininde 0.2 ve 0.5 hata toleransında değişme olmayacak ve doğruluk oranı yine %70 olacaktır. Fakat 1 derece hata toleransı konulması durumunda doğruluk oranı % 100 olacaktır.

1a89 proteini için 0.2 hata payı durumunda tablo aşağıdaki gibi olacaktır.

Tablo 6.10. 1a89 proteini için 0,2 derece hata payı sınıf tablosu

Gerçek Aç. S.	A4	A3	A4	A3	A3	A3	A4	A3	A4	A4	A3	A4	A3	A3	A4	A3	A3	A5
YSA Açısı	A4	A3	A5	A4	A3	A4	A4	A3	A4	A4	A3	A4	A4	A3	A4	A3	A3	A4
Bağıl h.			0.8	0.5		0.4	0.1						0.5					1.4
Gerçek Aç.	A5	A3	A3	A4	A3	A3	A4	A4	A3	A4	A4	A4	A4	A3	A4	A4	A4	A3
YSA Açısı	A4	A3	A3	A5	A4	A3	A4	A4	A3	A4	A4	A3	A4	A4	A3	A4	A3	A3
Bağıl h.	2.5	0.2		0.8	0.5			0.1					0.9		0.5	0.9		1.4
Gerçek Aç.	A3	A4	A4	A4	A4	A3	A3	A4	A3	A3	A4	A4	A3	A4	A4	A3	A4	A3
YSA Açısı	A4	A4	A4	A3	A5	A4	A3	A4	A3	A3	A4	A4	A3	A4	A4	A3	A4	A3
Bağıl h.	1.6			0.8	0.8	0.5												
Gerçek Aç.	A3	A4	A4															
YSA Açısı	A3	A4	A4															
Bağıl h.																		

Bu durumda ise eşleşen açı sınıf sayımız 41 olmakta ve doğruluk oranı %71.93 olmaktadır.

Aynı protein için hata payı 0.5 olarak kabul edilirse tablo aşağıdaki gibi olacaktır:

Tablo 6.11. 1a89 proteini için 0,5 derece hata payı sınıf tablosu

Gerçek Aç. S.	A4	A3	A4	A3	A3	A3	A4	A3	A4	A4	A3	A4	A3	A3	A4	A3	A3	A5
YSA Açısı	A4	A3	A5	A3	A3	A3	A4	A3	A4	A4	A3	A4	A3	A3	A4	A3	A3	A4
Bağıl h.			0.8	0.5		0.4	0.1						0.5					1.4
Gerçek Aç.	A5	A3	A3	A4	A3	A3	A4	A4	A3	A4	A4	A4	A4	A3	A4	A4	A4	A3
YSA Açısı	A4	A3	A3	A5	A3	A3	A4	A4	A3	A4	A4	A3	A4	A3	A3	A4	A3	A3
Bağıl h.	2.5	0.2		0.8	0.5			0.1					0.9		0.5	0.9		1.4
Gerçek Aç.	A3	A4	A4	A4	A4	A3	A3	A4	A3	A3	A4	A4	A3	A4	A4	A3	A4	A3
YSA Açısı	A4	A4	A4	A3	A5	A3	A3	A4	A3	A3	A4	A4	A3	A4	A4	A3	A4	A3
Bağıl h.	1.6			0.8	0.8	0.5												
Gerçek Aç.	A3	A4	A4															
YSA Açısı	A3	A4	A4															
Bağıl h.																		

Bu durumda ise eşleşen 47 aç bulunmaktadır. Doğruluk oranı ise %82,46 olmaktadır.

Aynı protein için hata payı 1 derece olursa tablo aşağıdaki gibi olacaktır :

Tablo 6.12. 1a89 proteini için 1 derece hata payı sınıf tablosu

Gerçek Aç S.	A4	A3	A4	A3	A3	A3	A4	A3	A4	A4	A3	A4	A3	A3	A4	A3	A3	A5
YSA Açısı	A4	A3	A4	A3	A3	A3	A4	A3	A4	A4	A3	A4	A3	A3	A4	A3	A3	A4
Bağıl h.			0.8	0.5		0.4	0.1						0.5					1.4
Gerçek Aç	A5	A3	A3	A4	A3	A3	A4	A4	A3	A4	A4	A4	A4	A3	A4	A4	A4	A3
YSA a.	A4	A3	A3	A4	A3	A3	A4	A4	A3	A4	A4	A4	A4	A3	A4	A4	A3	A3
Bağıl h.	2.5	0.2		0.8	0.5			0.1				0.9		0.5	0.9		1.4	
Gerçek Aç	A3	A4	A4	A4	A4	A3	A3	A4	A3	A3	A4	A4	A3	A4	A4	A3	A4	A3
YSA a.	A4	A4	A4	A4	A4	A3	A3	A4	A3	A3	A4	A4	A3	A4	A4	A3	A4	A3
Bağıl h.	1.6			0.8	0.8	0.5												
Gerçek Aç	A3	A4	A4															
YSA a.	A3	A4	A4															
Bağıl h.																		

Bu durumda da eşleşen 53 adet aç olur ve doğruluk oranı %92,98 olur.

1a9a proteini için hata payı 0,2 olarak seçilirse tablo aşağıdaki gibi olacaktır :

Tablo 6.13. 1a9a proteini için 0,2 derece hata payı sınıf tablosu

Gerçek Aç s.	A4	A3	A4	A4	A3	A3	A3	A3	A4	A3	A3	A4	A4	A3	A4	A3	A3	A3
YSA	A4	A3	A5	A4	A3	A4	A3	A3	A4	A3	A3	A4	A4	A3	A4	A3	A3	A4
Bağıl hata			0.8			0.4				0.2								1.6
Gerçek Aç s.	A4	A3	A3	A3	A4	A3	A3	A4	A4	A4	A4	A3	A3	A3	A3	A3	A4	A3
YSA	A4	A5	A4	A3	A3	A3	A3	A5	A1	A4	A4	A9	A1	A3	A4	A5	A4	A2
Bağıl hata		3.1	0.3	0.1	1.8	0.1	0.2	2.1	6.0		0.1	17.1	3.0		0.4	4.8		0.5
Gerçek Aç s.	A3	A3	A4	A3	A4	A3	A3	A4	A4	A4	A4	A3	A3	A4	A3	A3	A4	A4
YSA	A2	A1	A4	A3	A5	A4	A3	A4	A4	A3	A4	A3	A4	A3	A3	A3	A4	A3
Bağıl hata	1.7	6.0			0.8	0.5			0.1	0.9		0.2	1.8	0.3				1.4
Gerçek Aç s.	A3	A4	A3															
YSA	A3	A4	A4															
Bağıl hata			0.5															

Bu durumda 35 aç eşleşmektedir. Doğruluk oranı %61.40 olmaktadır.

Aynı protein için hata payı 0.5 olarak seçilirse tablo aşağıdaki gibi olmaktadır :

Tablo 6.14. 1a9a proteini için 0,5 derece hata payı sınıf tablosu

Gerçek Açı s.	A4	A3	A4	A4	A3	A3	A3	A3	A4	A3	A3	A4	A4	A3	A4	A3	A3	A3
YSA	A4	A3	A5	A4	A3	A3	A3	A3	A4	A3	A3	A4	A4	A3	A4	A3	A3	A4
Bağıl hata			0.8			0.4				0.2								1.6
Gerçek Açı s.	A4	A3	A3	A3	A4	A3	A3	A4	A4	A4	A4	A3	A3	A3	A3	A3	A4	A3
YSA	A4	A5	A3	A3	A3	A3	A3	A5	A1	A4	A4	A9	A1	A3	A3	A5	A4	A3
Bağıl hata		3.1	0.3	0.1	1.8	0.1	0.2	2.1	6.0		0.1	17.1	3.0		0.4	4.8		0.5
Gerçek Açı s.	A3	A3	A4	A3	A4	A3	A3	A4	A4	A4	A4	A3	A3	A4	A3	A3	A4	A4
YSA	A2	A1	A4	A3	A5	A3	A3	A4	A4	A3	A4	A3	A4	A4	A3	A3	A4	A3
Bağıl hata	1.7	6.0			0.8	0.5			0.1	0.9		0.2	1.8	0.3				1.4
Gerçek Açı s.	A3	A4	A3															
YSA	A3	A4	A3															
Bağıl hata			0.5															

Bu durumda 42 açı eşleşmektedir ve doğruluk oranı %73,68 olmaktadır.

Aynı protein için hata payı 1 derece kabul edilecek olursa tablo aşağıdaki gibi olacaktır :

Tablo 6.15. 1a9a proteini için 1 derece hata payı sınıf tablosu

Gerçek Açı s.	A4	A3	A4	A4	A3	A3	A3	A3	A4	A3	A3	A4	A4	A3	A4	A3	A3	A3
YSA	A4	A3	A4	A4	A3	A3	A3	A3	A4	A3	A3	A4	A4	A3	A4	A3	A3	A4
Bağıl hata			0.8			0.4				0.2								1.6
Gerçek Açı s.	A4	A3	A3	A3	A4	A3	A3	A4	A4	A4	A4	A3	A3	A3	A3	A3	A4	A3
YSA	A4	A5	A3	A3	A3	A3	A3	A5	A1	A4	A4	A9	A1	A3	A3	A5	A4	A3
Bağıl hata		3.1	0.3	0.1	1.8	0.1	0.2	2.1	6.0		0.1	17.1	3.0		0.4	4.8		0.5
Gerçek Açı s.	A3	A3	A4	A3	A4	A3	A3	A4	A4	A4	A4	A3	A3	A4	A3	A3	A4	A4
YSA	A2	A1	A4	A3	A4	A3	A3	A4	A4	A4	A4	A3	A4	A4	A3	A3	A4	A3
Bağıl hata	1.7	6.0			0.8	0.5			0.1	0.9		0.2	1.8	0.3				1.4
Gerçek Açı s.	A3	A4	A3															
YSA	A3	A4	A3															
Bağıl hata			0.5															

Bu durumda da 45 aç ı eşleşmektedir ve doğruluk oran ı %78,95 olmaktadır.

Gör ıldüğü üzere aç ı sınıflarının sınır değ erlerinde biraz hata toleransı bırakmak tahminin daha kuvvetli olmasını sağlamaktadır.

7. SONUÇ

Canlılığın en önemli yapı taşlarından olan proteinler karmaşık fakat, karmaşık olduklarından daha fazla düzenli ve şaşırtıcı derecede yerli yerinde olmaları ile bugün olduğu gibi gelecekte de insanoğlunun ilgisini çekmeye devam edecektir. Özellikle hastalıkların tedavisinde hastalığa sebep olan mikrop yada virüs benzeri zararlı etkenlerin enzim yapılarının incelenmesi ve buna karşı ilaç geliştirilmesi gibi konularda proteinlerin yapılarının çözülmesine olan ihtiyaç bu konuyu sürekli gündemde tutmaya yetecektir.

Gelişen bilgisayar teknolojisi bu ve benzeri konulara çözüm getirme noktasında kolaylıklar sağladıkça gelişmeler akıl almaz hızlarla devam edecektir. Bundan 40-50 yıl öncesine kadar sadece delikli kartlarla bir takım verileri aktaran ve basit dört işlemleri gerçekleştirebilen bilgisayar sistemleri bugün her konuda tabiri caizse bilimin eli ayağı olmaya devam etmektedir.

Bu tezde incelenen proteinlerin sekonder yapıları, proteinin fonksiyonel yapısı hakkında ip uçları vermesi bakımından proteinlerin üç boyutlu yapıları ile beraber önemli bir konu olarak protein dünyasında yer tutmaktadır. Bizim bu tezde yapmış olduğumuz çalışma sekonder yapı tahminine bir yaklaşımdır. Bazı örnekler için son örnek proteinimizde olduğu gibi tahmin değerimiz düşük çıkabilmekle beraber sebepleri şu şekilde izah edilebilir:

Proteinler çok kompleks yapıları itibariyle, yapılarının tahmininde her zaman çok iyi bir başarıya el vermeyebilir. Çünkü aynı amino asit dizilimine sahip protein zincirlerinin bağ açıları şaşırtıcı bir şekilde farklı iki proteinde farklı değerler alabilmektedir. Bu etraflarında yer alan diğer amino asitlerin katlanma yapısından dolayı etkileşim oluşturacak şekilde yer almasındandır. Bu konuda tam bir yaklaşım gerçekleştirebilmek için, atomlar arası çekim kuvvetleri konusunda çok iyi bir inceleme yapmak gerekmektedir.

İkinci olarak tahminin başarısının düşük olma nedeni seçilen açılı sınırlarından kaynaklanmaktadır. Açılı sınırlarını değiştirmek bazen çözüm getirebilse de başka örnekler için sapmaya neden olabilir. Fakat ortalama açılı hatası ve bağılı hata dikkatlice incelendiğinde görülecektir ki hata olan sınıflarda açının başka sınıfa kaymasına neden olan değerler iki boyutlu yapıda çok da katlanma şeklini değiştirmeyecek hata değerleridir. Proteinler gibi çok hassas bir konuda çok küçük açılıların bile fonksiyonu ve aktiviteyi etkilediğini hatta işe yaramaz moleküller oluşturabildiğini düşünürsek daha iyiye yaklaşmaya çaba harcanması gerektiğini söyleyebiliriz. Bu, sadece bu çalışma için değil yapılan tüm çalışmalar için geçerlidir.

6. bölümün sonunda hata toleransının belirli değerlere çekilmesi göstermiştir ki belirli tolerans değerleri sonucunda hata iyice aşağı çekilebilmekte ve doğruluk oranı iyice

yükseltilebilmektedir. Açı sınıflarının bir tablo halinde bölünmesi klasik küme teorisi yöntemiyle sağlandığından çok küçük açı sapmaları başka sınıflara kaymalara neden olmaktadır. Bunun önüne geçmek için, sınıf ayırıcı denklemi kullanıcı tarafından tanımlamak yerine, araya bir uzman sistem koymak daha farklı ve güzel çözümler getirebilir.

Bu uzman sistem bir Neuro-Fuzzy sınıflandırıcı olabilir. Bu durumda kümeler kesin ayrık değerlerden ziyade, üyelik fonksiyonları yardımıyla bulanık değerler olursa sınıflara ait olma daha esnek olacaktır. Dolayısıyla eşleşmeler daha sık ve doğruluk oranları daha yüksek olacaktır.

Tabi bunun yanında proteinlerin yapılarındaki sapmaların molekül yapısını çok değiştirebildiği hatta işlevi tamamen etkisiz hale getirebildiğini de akıldan çıkarmamak gerekmektedir.

Geliştirmiş olduğumuz yaklaşımda eğitim açısının sayısı arttırıldıkça ve farklı örnek açı değerleri seçildikçe yaklaşım daha da iyiye doğru gidecektir.

KAYNAKLAR

- [1] Doç Dr. Ayşe CELAL, Türk Farmakoloji Derneği Klinik Farmakoloji Çalışma Grubu Elektronik Bülteni Yeni ilaç geliştirilmesi ,1 Ocak-Nisan 2005.
- [2] Prof. Dr. Edip KEHA, Prof. Dr. Ö. İrfan KÜFREVİOĞLU, Biyokimya, Şafak Yayınevi, Erzurum 1997.
- [3] Stephen A. Krawetz, David D. Womble, Introduction to Bioinformatics A theoretical and Practical Apporach, Humana Pres., 2003.
- [4] Prof. Dr. Taner ONAT, Prof. Dr. Kaya EMERK, Doç. Dr. Eser Y. SÖZMEN, İnsan Biyokimyası, Palme Yayıncılık, 2002.
- [5] Prof. Dr. Nedret KILIÇ, Lehninger Biyokimyanın İlkeleri, Palme Yayıncılık, 2005
- [6] Christopher K. Mathews, K.E. van Holde, Kevin G. Ahern, Biochemistry Third Edition, Inc. Benjamin/ Cummings , 2000.
- [7] Prof. Dr. Ercan ÖZTEMEL, Yapay Sinir Ağları, Papatya Yayıncılık, 2003
- [8] Simon Haykin, Second Edition Neural Networks A Comprehensive Foundation, Simon&Schuster/A Viacom Company, 1999.
- [9] Pierre Baldi¹, Søren Brunak², Paolo Frasconi³, Gianluca Pollastri¹, and Giovanni Soda³ , Bidirectional Dynamics for Protein Secondary Structure Prediction , Sequence Learning, LNAI 1828, pp. 80–104, 2000.
- [10] V.Di Francesco, 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.
- [11] 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.
- [12] Kuen-Pin Wu, Hsin-Nan Lin, Jia-Ming Chang, Ting-Yi Sung and Wen-Lian Hsu, Hyprosp : A Hybrid Protein Secondary Structure Prediction Algoritihm –A Knowledge –Based Approach, Institute of Information Science, Academia Sinica, Taipei Taiwan , 2004.
- [13] Yan Liv, Jaime Carbonell, Judith Klein-Seetharaman, Vanathi Gopalaknishnan, Context Sensitive Vocabulary and its Appliation in Protein Secondary Strucute Prediction, Sheffield, South Yorkshire, U.K. , SIGIR'04 July 25-59 , 2004.
- [14] Varun Aggorwal , Prediction of Protein Secondary Structure Using Genetic Proqraming, Summer Internship Project Report Stockholm University , Sweeden 2003.

- [15] Victor Robles, Maria S. Perez, Vanessa Herves, Jose M. Pena and Pedro Larranaga, Parallel Stochastic Search for Protein Secondary Structure Prediction , Springer-Verlag Berlin Heidelberg , 2004.
- [16] Volkan KURT, Protein Structure Prediction Using Decision Lists , Master Thesis Koç University,2005.
- [17] Howard Holley and Martin Karplus, Protein Secondary Structure Prediction with a Neural Network,Department of Chemistry, Harvard University Cambridge,1989.
- [18] İrem ERSÖZ, Secondary Structure Prediction of Hemoglobin by Using Combined Neural Networks, Master Thesis Çukurova University,2003
- [19] P. Baldi and S. Brunak. Bioinformatics: The Machine Learning Approach. The MIT Press, Cambridge, Massachusetts and London, England, 2001

ÖZGEÇMİŞ

Murat DEMİR

Fırat Üniversitesi
Mühendislik Fakültesi
Bilgisayar Mühendisliği Bölümü
23119, ELAZIĞ

E_Posta : mdemir@firat.edu.tr

1979	Eragrı'de doğdu.
1993-1996	Elazığ Lisesi'nden mezun oldu.
1997-2002	Fırat Üniversitesi Bilgisayar Mühendisliği bölümünden mezun oldu.
2002-	Fırat Üniversitesi Muş Meslek Yüksek Okulu'nda öğretim görevlisi olarak görev yapmaktadır.