

YALOVA ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

**PROTEİNLERDEKİ DÜZENSİZ BÖLGELERİN TESPİTİ İÇİN
KAOTİK VE FİZİKOKİMYASAL ÖZELLİKLER TABANLI
YENİ ÖZNİTELİK KODLAMA YÖNTEMLERİ GELİŞTİRİLMESİ**

YÜKSEK LİSANS TEZİ

SEVDANUR GENÇ

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

NİSAN 2015

YALOVA ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

**PROTEİNLERDEKİ DÜZENSİZ BÖLGELERİN TESPİTİ İÇİN
KAOTİK VE FİZİKOKİMYASAL ÖZELLİKLER TABANLI
YENİ ÖZNİTELİK KODLAMA YÖNTEMLERİ GELİŞTİRİLMESİ**

YÜKSEK LİSANS TEZİ

**SEVDANUR GENÇ
135105008**

Bilgisayar Mühendisliği Anabilim Dalı

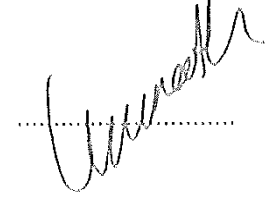
Bilgisayar Mühendisliği Programı

Tez Danışmanı: Doç. Dr. MURAT GÖK

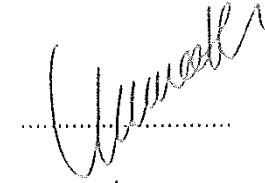
NİSAN 2015

YALOVA Üniversitesi Fen Bilimleri Enstitüsü'nün 135105008 numaralı Yüksek Lisans Öğrencisi Sevdanur GENÇ, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı "PROTEİNLERDEKİ DÜZENSİZ BÖLGELERİN TESPİTİ İÇİN KAOTİK VE FİZİKOKİMYASAL ÖZELLİKLER TABANLI YENİ ÖZNİTELİK KODLAMA YÖNTEMLERİ GELİŞTİRİLMESİ" başlıklı tezini aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

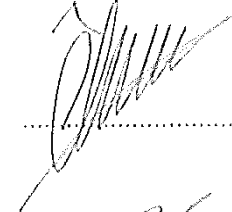
Tez Danışmanı : Doç. Dr. Murat GÖK
Yalova Üniversitesi



Jüri Üyeleri : Doç. Dr. Murat GÖK
Yalova Üniversitesi



Doç. Dr. Ahmet Turan ÖZCERİT
Sakarya Üniversitesi



Yrd. Doç. Dr. Osman Hilmi KOÇAL
Yalova Üniversitesi



Teslim Tarihi : 6 Şubat 2015
Savunma Tarihi : 20 Nisan 2015





Kıymetlilerim; Annem ve Abime,



ÖNSÖZ VE TEŞEKKÜR

Bu tez çalışmasında, proteinlerdeki düzensiz bölgelerin tespiti için öncelikle protein dizilimlerinin oluşmasında etkili olan fizikokimyasal özelliklerin kaotik yapı içerip içermedikleri araştırılmıştır. Sonra belirlenen fizikokimyasal özelliklere dayanan öznitelik kodlama yöntemleri geliştirilmiştir.

Yüksek lisans eğitimimin başlangıcından, "**Proteinlerdeki Düzensiz Bölgelerin Tespiti İçin Kaotik ve Fizikokimyasal Özellikler Tabanlı Yeni Öznitelik Kodlama Yöntemleri Geliştirilmesi**" isimli tez çalışmamın sonuçlandırılmasına kadar olan her aşamada beni yönlendiren değerli bilimsel tecrübelerini, farklı bakış açılarını, teşviklerini, cesaretlendirmelerini, yorumlarını, eleştirilerini, anlayışlarını, tavsiyelerini ve yardımlarını esirgemeyip katkılarıyla beni aydınlatan danışmanım Sayın Doç. Dr. Murat GÖK'e sonsuz teşekkürlerimi sunarım.

Aynı zamanda, bu tezin hazırlanma sürecinde her zaman büyük özveri göstererek, yardımlarını esirgemeyen, bilimsel tecrübeleriyle sonuca ulaşmamı sağlayan Sayın Yrd. Doç. Dr. Osman Hilmi KOÇAL'a saygılarımı ve teşekkürlerimi sunarım.

Yalova Üniversitesi Bilgisayar Mühendisliği Bölümünün tüm öğretim üyeleri ve elemanlarına teşekkür ederim.

Eğitimimde yüksek lisansa devam etmem konusunda beni teşvik eden, akademik deneyimleriyle yönlendiren ve her zaman desteğini esirgemeyen değerli hocam Sayın Doç. Dr. Ecir Uğur KÜÇÜKSİLLE'ye minnet ve şükranlarımı sunarım.

Ayrıca, çalışmalarım esnasında maddi ve manevi desteklerini hiçbir zaman esirgemeyip hayatım boyunca yanımda olan ve bugünlere gelmemde emeği ve katkısı bulunan, öncelikle dualarını hiç bir zaman eksik etmeyen kıymetli annem Liman AKÇAY GENÇ'e ve sonrasında bana hem abilik hem babalık yapan canım abim Selçuk GENÇ'e en içten teşekkürlerimi ve sevgilerimi sunarım. Son olarak, bu çalışmaya beni sürükleyen en büyük sebep olan, çağımızın hastalığı kanser nedeniyle 2005 yılında kaybettiğim değerli babam Naim GENÇ'e ithaf ediyorum.

Nisan 2015

Sevdanur GENÇ
Bilgisayar Mühendisi



İÇİNDEKİLER

Sayfa

ÖNSÖZ VE TEŞEKKÜR	vii
İÇİNDEKİLER	ix
KISALTMALAR	xi
ÇİZELGE LİSTESİ	xiii
ŞEKİL LİSTESİ	xv
ÖZET.....	xvii
SUMMARY.....	xix
1. GİRİŞ	1
1.1. Tezin Amacı	1
1.2. Literatür Araştırması	1
1.3. Hipotez	2
1.4. Tez Organizasyonu	2
2. PROTEİNLERDEKİ DÜZENSİZ BÖLGELER	5
2.1. Amino Asitler	5
2.1.1. Amino asitlerin fizikokimyasal özellikleri	7
2.2. Proteinler.....	8
2.2.1. Proteinlerin düzensiz bölgeleri	9
2.3. DisProt ve PDB Veri Setleri	10
3. LYAPUNOV ÜSTELLERİ	11
3.1. Kaotik Çekerler	11
3.2. Kaos Ölçme Yöntemleri	13
3.3. Lyapunov Üstellerinin Hesaplanması	14
3.3.1. Lyapunov Spektrumu ve analizi	15
4. ÖZİNİTELİK KODLAMA	19
4.1. Sınıflandırma	20
4.2. Öznelik Çıkarımı	20
5. GELİŞTİRİLEN ÖZİNİTELİK KODLAMA YÖNTEMLERİ	23
5.1. Protein Dizilimlerdeki Kaotik Yapının Analizi	23
5.2. En İyi Fizikokimyasal Özelliklere Göre Öznelik Kodlama Yöntemi.....	28
5.3. Lyapunov Özellikleri ve En İyi Fizikokimyasal Özelliklere Göre Öznelik Kodlama Yöntemi	29
5.4. Performans Değerlendirmesi	30
6. SONUÇLAR.....	33
KAYNAKLAR	35
ÖZGEÇMİŞ	39

KISALTMALAR

ALA	: Alanin
ARG	: Arginin
ASN	: Asparajin
ASP	: Aspartik Asit
CYS	: Sistein
GLN	: Glutamin
GLU	: Glutamik Asit
GLY	: Glisin
HIS	: Histidin
ILE	: İzölösın
LEU	: Lösın
LYS	: Lizin
MET	: Metiyonin
PHE	: Fenilalanin
PRO	: Prolin
SER	: Serin
THR	: Treonin
TRP	: Triptofan
TYR	: Trozin
VAL	: Valin
DNA	: Deoksiribonükleik Asit
RNA	: Ribonükleik Asit
AA	: Amino Asit
AAIndex	: Amino Asit Index
mRNA	: Messenger RNA
NMR	: Nükleer Manyetik Rezonans
CD	: Dairesel Dikroizm
MATLAB	: Matrix Laboratory
WEKA	: Waikato Environment for Knowledge Analysis
TISEAN	: Nonlinear Time Series Analysis
DisProt	: The Database Of Protein Disorder
PDB	: Protein Databank
LÜ	: Lyapunov Üstelleri
FKÖY	: En İyi Fizikokimyasal Özelliklere Göre Öznitelik Kodlama Yöntemi
LÜ - FKÖY	: Lyapunov Üstelleri ve En İyi Fizikokimyasal Özelliklere Göre Öznitelik Kodlama Yöntemi
FK	: Fizikokimyasal
MCC	: Mathew'in Korelasyon Katsayısı
SD	: Sınıf Doğruluğu



ÇİZELGE LİSTESİ

Sayfa

Çizelge 2.1 : Temel AA kalıntılarına ait isim, sembol ve FK özellikleri	5
Çizelge 2.2 : Amino asitlerin fizikokimyasal özelliklerine ait indeks tablosunu.....	7
Çizelge 2.3 : DisProt ve PDB verisetlerine ait bilgiler.....	10
Çizelge 3.1 : Lorenz denkleminin değişkenleri ve parametreleri	13
Çizelge 3.2 : LÜ metotları	15
Çizelge 3.3 : .lyap_spec parametreleri ve açıklamaları	17
Çizelge 5.1 : Bir fizikokimyasal özelliğe ait LÜ değerlerinin sonuçları	23
Çizelge 5.2 : Protein dizilimlerindeki düzensiz bölgeler üzerinde belirlenen LÜ değerleri	24
Çizelge 5.3 : FK için LÜ değerleri	28
Çizelge 5.4 : En iyi FK için LÜ değerleri	29
Çizelge 5.5 : En iyi FK için AAIndex değerleri	29
Çizelge 5.6 : En iyi FK'nın AAIndex değeri ile çarpımı	29
Çizelge 5.7 : Öznitelik çıkarım sonuçları	30
Çizelge 5.8 : Karmaşıklık Matrisi	30
Çizelge 5.9 : FKÖY 'e ait sınıflandırma sonuçları	31
Çizelge 5.10 : LÜ-FKÖY 'e ait sınıflandırma sonuçları	32
Çizelge 5.11 : Literatürdeki test yöntemleri ile karşılaştırılması	32



ŞEKİL LİSTESİ

Sayfa

Şekil 2.1 : Bir amino asitin yapısı	6
Şekil 2.2 : İki amino asitin bir peptit bağı oluşturması	6
Şekil 2.3 : Bir α helix ve bir β sheet yapısı	8
Şekil 2.4 : Düzensiz bir proteinin düzenli hale geçiş evresi	9
Şekil 3.1 : Lorenz çeker	12
Şekil 3.2 : Seçilen birbirine yakın iki ilk koşul ve evrimleri	14
Şekil 3.3 : lyap_spec komutunun LÜ hesaplanması için kullanımı	16
Şekil 3.4 : lyap_spec kullanımı sonucu LÜ sonuçları	17
Şekil 4.1 : Örüntü tanıma sisteminin genel diyagramı	19
Şekil 5.1 : Amino asitlere ait işaretlerin bir seri üzerindeki değişimleri	25
Şekil 5.2 : AA'lere ait işaretlerin bir seri üzerindeki değişimlerinin yakından gösterimi	26
Şekil 5.3 : Bir fizikokimyasal özelliğin faz uzayındaki kaotik yapısı	27
Şekil 5.4 : Bir fizikokimyasal özelliğe ait tüm LÜ işaretlerinin $D_E = 2$ boyutlarındaki faz uzayında gösterimi	27
Şekil 5.5 : Bir fizikokimyasal özelliğe ait tüm LÜ işaretlerinin $D_E = 3$ boyutlarındaki faz uzayında gösterimi	28



PROTEİNLERDEKİ DÜZENSİZ BÖLGELERİN TESPİTİ İÇİN KAOTİK VE FİZİKOKİMYASAL ÖZELLİKLER TABANLI YENİ ÖZNETELİK KODLAMA YÖNTEMLERİ GELİŞTİRİLMESİ

ÖZET

Canlının hayatını devam ettirmesinde kritik önem taşıyan proteinler, tüm biyokimyasal reaksiyonlarda yer almaktadır. Proteinleri oluşturan aminoasitlerin fizikokimyasal özellikleri, amino asitlerin bağlanma sıralarının ve proteinin üç boyutlu yapısının oluşmasında en önemli belirleyicidir. Amino asitler peptit bağlarıyla birbirlerine bağlanır, bu durum hücre içerisinde en düşük enerji seviyesine sebep olur. Böylelikle, protein zincirleri doğal durumunda katlanır ve proteini meydana getirirler. Proteinlerin her bir düzgün katlanması konformasyon olarak adlandırılırken aynı zamanda katlanamayan diğer bir ifadeyle düzensiz proteinlerde oluşabilmektedir. Düzensiz proteinler organizma içerisinde kanser, alzheimer, kardiyovasküler, kistik fibrozis ve diyabet gibi metabolik bir çok hastalıklığa sebep olmaktadır. Bu nedenle proteinlerde düzensiz bölgelerin tespiti hastalıkların tedavisi ve ilaç geliştirilmesi açısından önemli bir eşiktir.

Literatürde proteinlerde düzensiz bölgelerin tespiti için in vitro ve in silico çalışmalar yapılmaktadır. Bilgisayar ortamında gerçekleştirilen in silico çalışmalar in vitro çalışmalara göre maliyet ve zaman açısından üstünlüklere sahiptir. Buna göre bilgisayar ortamında düzensiz bölgeler çeşitli makine öğrenmesi algoritmaları ve biyoinformatik yöntemleri ile tespit edilmeye çalışılmaktadır. Bu tez çalışmasında, öncelikle protein dizilimleri fizikokimyasal özelliklere göre kodlanarak faz uzayında incelendi ve kaotik yapı içeren fizikokimyasal özellikler belirlendi. Sonrasında proteinlerdeki düzensiz bölgeleri tahmin etmek için belirlenen amino asitlerin fizikokimyasal özellikleri kullanılarak makine öğrenmesi algoritmaları için En İyi Fizikokimyasal Özelliklere Göre Öznitelik Kodlama Yöntemi (FKÖY) ve Lyapunov Üstelleri ve En İyi Fizikokimyasal Özelliklere Göre Öznitelik Kodlama Yöntemi (LÜ-FKÖY) olarak isimlendirilen iki öznitelik kodlama yöntemi geliştirilmiştir.

Proteinlerde düzensiz bölgelerin tespiti için DisProt ve PDB verisetleri üzerinde geliştirilen yöntemler sınıflandırıcı algoritmalarla uygulanmıştır. Elde edilen deneysel sonuçlarda LÜ-FKÖY, literatürde yer alan yöntemlere göre en yüksek performansı sergilemiştir.



DEVELOPING NEW ATTRIBUTE CODING METHODS TO PREDICTION OF DISORDERED PROTEIN REGIONS ON BASED CHAOTIC AND PHYSICOCHEMICAL PROPERTIES

SUMMARY

Proteins located in all biochemical reactions have vital importance for continue of organism life. Physicochemical properties of amino acids that constitute proteins are the most important to formation for three-dimensional structure of protein and binding orders of amino acids linked by peptide bonds. Moreover, this condition leads to the lowest energy level in the cell. In this way, protein chains are folded naturally and created proteins. Each of ordered proteins is called the conformation. And also, disordered proteins can be formed. Disordered proteins cause many metabolic disease in organisms such as cancer, alzheimer, cardiovascular, cystic fibrosis and diabetes. Therefore, prediction of disordered regions in proteins is a significant threshold in terms of development of treatment for the diseases.

In the literature, many studies have been applied to the prediction of disordered regions in protein in vitro and in silico environments. In silico study performances on computer has some advantages like cost and time-consume relatively in-vitro studies. Prediction of disordered regions have been determining by various machine learning algorithms and bioinformatics methods in silico. In this thesis, primarily, protein sequences coded by physicochemical properties and examined in phase space. Moreover, physicochemical properties on based chaotic structure was determined. And than, it has been developed using machine learning algorithms determined physicochemical properties of the amino acids such as Feature Encoding Method by Selected Best Physicochemical Properties and Features Encoding Method by Lyapunov Exponents and Selected Best Physicochemical Properties.

Proposed methods have been applied with classifier algorithms on DisProt and PDB datasets to the prediction of disordered region in proteins. According to obtained experimental results that our methods have demonstrated the highest performance than most methods in the literature.



1. GİRİŞ

1.1. Tezin Amacı

Düzensiz bölgeler içeren proteinler kanser, alzheimer, kardiyovasküler ve diyabet gibi metabolik hastalıklara sebep olmaktadır. Ayrıca düzensiz bölgelerin incelenmesi, transkripsiyon ve translasyonun regülasyonu, protein - DNA ve protein - protein birleşimlerinde önem taşımaktadır [1] [2]. Bu tezin amacı, proteinlerin işlevleri, yaşam sürelerini ve hastalıklarla olan ilişkilerini belirleyen fizikokimyasal özelliklerin karmaşık düzenlerinin kaotik bir yapıya dayandığını ispatlamak ve öncül fizikokimyasal özellikleri kestirmektedir. Sonrasında belirlenen fizikokimyasal özellikler temelinde proteinlerin düzensiz yerlerinin kestirimi için yeni öznelik kodlama teknikleri geliştirmektedir.

1.2. Literatür Araştırması

Proteinlerdeki düzensiz bölgelerin kestirimi için farklı teorik ve deneysel yaklaşımlar gerçekleştirilmiştir. X-Işını kristalografisi, NMR spektroskopisi, ultraviyoleye yakın dairesel dikroizm (CD) ve uzak ultraviyole CD gibi ıslak laboratuvar yöntemleri problemin çözümü için kullanılmıştır. Ancak, ıslak laboratuvar (in vitro) ortamında gerçekleştirilen bu deneyler hem masraflı hem de çevresel faktörler (laboratuvar şartları, insan faktörü) kaynaklı yanlış ölçümlere açıktır [3] [4]. Bu yüzden, hesaplamalı öngörü yöntemleri önemli bir alternatiftir. Bilgisayar ortamında (in silico) makine öğrenmesi temelli yapılan çalışmalarda, sınıflandırma yöntemleri sayesinde düzensiz bölgelerin tahmini yüksek doğruluk ile gerçekleştirilmesi hedeflenmektedir. Bu tez çalışmasının testleri in silico ortamda gerçekleştirilmiştir.

Literatürde RONN [5], DisEMBL [6], DISOPRED2 [7], VSL1 [8], VSL2B ve VSL2P [9] gibi düzensiz kalıntı bölgelerini kullanarak bir dizi hesaplama aracı geliştirilmiştir.

Ronn (The Regional Order Neural Network) tekniği ile proteinlerde düzensiz bölgelerin tespiti için örüntü tanıma çalışmalarında PDB'den elde edilen protein

dizilimleri üzerinde sinir ağırları kullanarak iyi performanslara ulaşılmıştır [6]. DisEMBL tekniğinde yapay sinir ağırları protein dizileri üzerinde eğitilerek probleme odaklanılmıştır [5]. DISOPRED2, düzenli ve düzensiz bölgeler arasındaki ayrımı bir destek vektör makinası (SVM) kullanarak çözüme ulaşılmıştır. VSL1 geliştiricileri düzensiz bölge tahmini için üç bileşenli bir mimarı geliştirilmiştir. Bu mimaride, kısa kalıntılar için VSL1-S, orta uzunluktaki kalıntılar için VSL1-M ve uzun uzunluktaki kalıntılar için VSL1-L yöntemini tespit için kullanılmıştır [10]. VSL2 mimarisi, VSL1 mimarisi üzerinden geliştirilerek kısa kalıntıya sahip düzensiz bölgelere VSL2-S, uzun kalıntıya sahip düzensiz bölgelere VSL2-L ismi verilen hesaplama yöntemleri kullanılmıştır [8].

1.3. Hipotez

Proteinlerde düzensiz bölgelerin kestirimi için kaos teorisi tabanlı seçilen fizikokimyasal özelliklerin protein dizilimlerinin kodlanmasında kullanılabileceği hipotezini ispatladık. Buna göre, her bir fizikokimyasal özelliğe göre oluşturulan öznitelik vektörleri faz uzayında incelenerek kaotik bir yapı içerip içermediği gözlemlendi. Böylece kaotik yapı içeren fizikokimyasal özellikler belirlendi. Bu fizikokimyasal özellikler temelinde düzensiz bölgelerin kestirimi için iki öznitelik kodlama yöntemi geliştirildi. FKÖY adı verilen ilk teknikte seçilen fizikokimyasal özelliklere göre ikinci teknikte ise Lyapunov Üstel (LÜ) değerleri ve seçilen fizikokimyasal özelliklere göre öznitelik matrisleri oluşturuldu. FKÖY, LÜ-FKÖY adı verilen teknikler son olarak literatürdeki benzer çalışmalarla kıyaslandı.

1.4. Tez Organizasyonu

Tez Organizasyonu aşağıda özetlenen altı bölümden oluşmaktadır.

Bölüm 1 : Giriş : Bu bölümde tez çalışmasına konu olan problemin tanımı, çalışmanın amacı, literatürde bu problemin çözümü üzerine yapılan çalışmaların özeti, tez organizasyonu hakkında bilgi sunulmaktadır.

Bölüm 2 : Proteinlerdeki Düzensiz Bölgeler : Bu bölümde proteinler ve düzensiz bölgeleri hakkında genel bilgi sunulmuştur. DisProt ve PDB veri setlerinden bahsedilmiştir.

Bölüm 3 : Lyapunov Üstelleri : Fizikokimyasal özellik içinde kaotik yapının nasıl araştırıldığı anlatılmıştır.

Bölüm 4 : Öznitelik Kodlama : Kaotik yapının ispatlanması ve yeni bir öznitelik yöntemi kullanılması için gerekli olan örüntü tanıma sistemlerinden bahsedilmiştir.

Bölüm 5 : Geliştirilen Öznitelik Kodlama Yöntemleri : Öncelikle protein dizilimindeki kaotik yapıdan bahsedilmiştir. Bununla birlikte geliştirilen hem en iyi fizikokimyasal özelliklere hem de Lyapunov üstellerine göre iki yeni öznitelik kodlama yönteminden bahsedilmiştir. Son olarak geliştirilen bu iki yeni öznitelik kodlama yöntemine göre güncel veri setleri üzerindeki performans değerleri sunulmuştur.

Bölüm 6 : Sonuç : Tez çalışmasına dair sonuçlar açıklanmıştır.

2. PROTEİNLERDEKİ DÜZENSİZ BÖLGELER

2.1. Amino Asitler

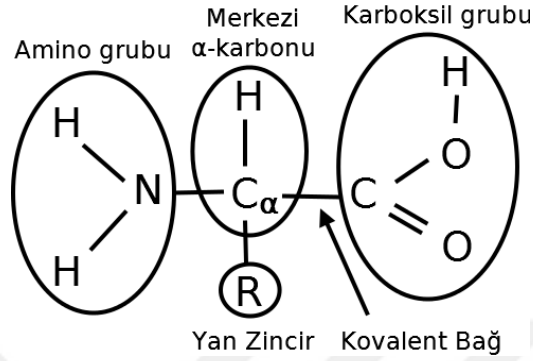
Amino asitler, proteinleri oluşturan temel yapı taşlarıdır. Proteinlerin içerisinde farklı sayıda ve dizide bulunan amino asitler, proteinlerin işlevinde çeşitliliklere sebep olmaktadır. Doğada 300 kadar farklı amino asit bulunmaktadır. Çizelge 2.1'de görülen 20 standart amino asit, DNA tarafından kodlanarak proteinler ve diğer biyomoleküllerin sentezinde kullanılırlar [11].

Çizelge 2.1 : Temel AA kalıntılarına ait isim, sembol ve FK özellikleri.

Sıra	Amino Asit	3 Harf	1 Harf
1	Alanin	Ala	A
2	Arginin	Arg	R
3	Asparajin	Asn	N
4	Aspartik Asit	Asp	D
5	Sistein	Cys	C
6	Glütamin	Gln	Q
7	Glütamik Asit	Glu	E
8	Glisin	Gly	G
9	Histidin	His	H
10	İzolösin	Ile	I
11	Lösin	Leu	L
12	Lizin	Lys	K
13	Metiyonin	Met	M
14	Fenilalanin	Phe	F
15	Prolin	Pro	P
16	Serin	Ser	S
17	Treonin	Thr	T
18	Triptofan	Trp	W
19	Trozin	Tyr	Y
20	Valin	Val	V

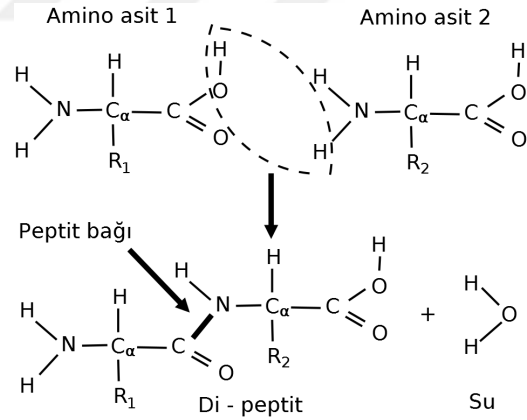
Bir amino asit hem amino grubu (-NH₂) hem de karboksil grubu (-COOH) içeren bir moleküldür. Amino asitlerin peptit bağlarıyla uç uca eklenmesiyle oluşturdukları kısa polimer zincirlere "Peptid", uzun polimer zincilere ise "Polipeptid" ya da "Protein" olarak adlandırılır. Protein zincirine dahil olmuş amino asitlere Kalıntı (Residue) denir. Fazla olan amino asitler harcanmaz ve depolanmaz, bunlar hücre içerisinde yakıt metabolizmasına dahil olmak üzere yakılmaktadırlar.

Şekil 2.1'de görüldüğü gibi bir amino asit dört kısımdan oluşur: α -karbonu (C_α), amino ($-NH_2$), karboksil ($-COOH$) ve yan zincir (R) grupları. Amino, karboksil ve yan zincir grupları kovalent bağlar ile merkezi α -karbonuna bağlanmışlardır.



Şekil 2.1 : Bir amino asitin yapısı.

İki amino asitin, peptid bağ oluşturabilmesi için amino ve karboksil grupları tepkimeye girer. Bu biyokimyasal tepkimede, bir amino asitin karboksil grubu başka bir amino grubuna bağlanmasıyla peptid bağı oluşur. Bu tepkime sonrasında Şekil 2.2'de görüldüğü gibi peptid bağı ile birlikte su ortaya çıkmaktadır.



Şekil 2.2 : İki amino asitin bir peptid bağı oluşturması.

Amino asitlerin fizikokimyasal özelliklerini her amino asitte bulunan özgün farklılıklar gösteren yan zincir grubu belirler [12].

2.1.1. Amino asitlerin fizikokimyasal özellikleri

Fizikokimya alanı, moleküllerin doğasını açıklamak için bu moleküllerde amino asitler ile olan etkileşimleri ve bu etkileşimler sırasında meydana gelen enerji alışverişlerini inceler. Amino asitler net yük, kimyasal reaktiflik, hidrojen bağlama, çözünürlük ve moleküler ağırlık gibi pek çok birbirinden farklı fizikokimyasal özelliklere sahiptirler. Her bir amino asit her bir fizikokimyasal özelliğe göre sayısal indeks değerleri bulunmaktadır. Fizikokimyasal özelliklerin her bir amino asite göre özgülleşmesi, proteinlerin işlevinin belirlenmesinde önemli rol oynadığı gibi sınıflandırılmalarına da yardımcı olmaktadır [11].

Amino asitlere ait fizikokimyasal özellik indeks değerlerine, AAIndex adı verilen çevrim içi bir veri tabanı aracılığı ile ulaşılabilir [13]. Çizelge 2.2'de AAIndex'ten alınan her bir amino asite göre fizikokimyasal özellik indeks değerleri görülmektedir. Tablodaki 544. fizikokimyasal özellik olan hidrofobiklik, amino asitler üzerinde hidrofobiklik derecelerini belirlemek amacıyla her bir amino asit için -6,04 ve 3.88 değerleri arasında değişkenlik göstermektedir.

Çizelge 2.2 : Amino asitlerin fizikokimyasal özelliklerine ait indeks tablosu.

Amino Asit	1. Fizikokimyasal Özellik	...	63. Fizikokimyasal Özellik	...	544. Fizikokimyasal Özellik
	Alpha - CH Kimyasal kaydırma	...	Büyüklik	...	Hidrofobiklik indeksi
A	4.35	...	2.5	...	-0.21
R	4.38	...	7.5	...	2.11
N	4.75	...	5	...	0.96
D	4.76	...	2.5	...	1.36
C	4.65	...	3	...	-6.04
Q	4.37	...	6	...	1.52
E	4.29	...	5	...	2.3
G	3.97	...	0.5	...	0
H	4.63	...	6	...	-1.23
I	3.95	...	5.5	...	-4.81
L	4.17	...	5.5	...	-4.68
K	4.36	...	7	...	3.88
M	4.52	...	6	...	-3.66
F	4.66	...	6.5	...	-4.65
P	4.44	...	5.5	...	0.75
S	4.5	...	3	...	1.74
T	4.35	...	5	...	0.78
W	4.7	...	7	...	-3.32
Y	4.6	...	7	...	-1.01
V	3.95	...	5	...	-3.5

AAIndex toplamda 3 bölümden oluşmaktadır. AAIndex1 20 sayısal değere sahip amino asit indeksleri, AAIndex2 amino asitlerin yer değiştirme matrisi ve AAIndex3 ise istatistiksel protein iletişim potansiyellerini barındırmaktadır. Bu tez çalışmasında 544 fizikokimyasal özellik içeren AAIndex1 (version 9.1) kullanılmıştır.

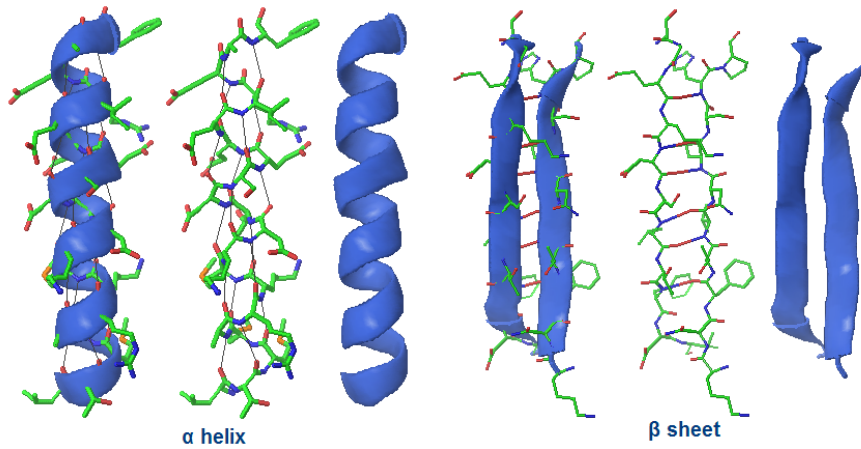
2.2. Proteinler

Proteinler bir veya birden fazla polipeptid zincirinden oluşmuş polimerlerdir. Amino asitlerin zincir halinde birbirlerine bağlanması sonucu oluşan büyük organik bileşiklerdir. Protein dizilimleri, serbest yapılı bir amino asit grubu ucu olan N terminalinden C terminali doğrultusunda ifade edilir.

Protein zincirleri birincil (primary), ikincil (secondary), üçüncül (tertiary) ve dördüncül (quarternary) olmak üzere dört çeşit yapı ile ifade edilirler.

Birincil yapılar, proteinlerin yapı taşı olan amino asitlerin birbirlerine bağlanış sıralarını gösteren basit bir yapıdır. Birincil yapı dışında kalan diğer üç yapı türüne ait proteinler üç boyutlu katlanma özelliklerine sahiptirler.

İkincil yapılar, hidrojen bağları ile düzenli bir geometriye ait terkarlanabilen yapılarıdır. Bu yapıya örnek olabilecek en yaygın örnek, Şekil 2.3'de görüldüğü gibi α sarmalı (α helix) ve β yaprağı (β sheet)'dir.



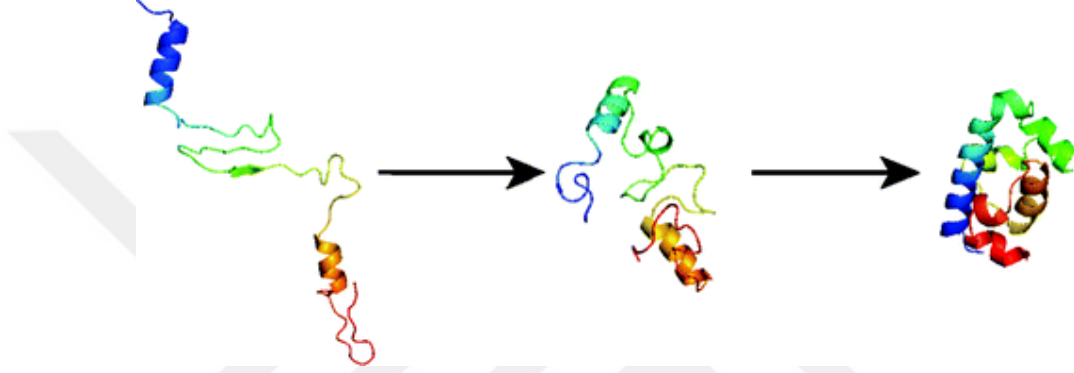
Şekil 2.3 : Bir α helix ve bir β sheet yapısı.

Üçüncül yapılar, proteinin doğal halini aldığı, α sarmalı (α helix) ve β yaprağı (β sheet) ile ifade edildikleri yapılarıdır.

Dördüncül yapılar, çoklu proteinleri (oligopeptit) ifade eden yapılarıdır.

2.2.1. Proteinlerin düzensiz bölgeleri

Amino asitler tarafından oluşturulan bir polipeptid zincirinin kendi yapısındaki karakteristik özelliklere göre şekillenmesine protein katlanması denir. Proteinler, işlevlerini yerine getirebilmeleri için katlanmaları gerekmektedir. Katlanma Şekil 2.4'de görüldüğü gibi, denaturasyon ve çökeltme olarak adlandırılan iki aşamadan meydana gelmektedir [14].



Şekil 2.4 : Düzensiz bir proteinin düzenli hale geçiş evresi [15].

Proteinlerdeki periyodik tekrarlanan yapılar düzenli bir yapıyı oluştururken, tesadüfi kıvrımlarsa düzensiz yapılar oluşturmaktadır. Proteinlerin düzensiz bölgelerinde α sarmalı ya da β yaprağı yapıları gerçekleşmez ve gelişi güzel kıvrımlar şeklinde oluşurlar. Oluşan bu kıvrımlar arasında belirli bir ilişki ve hidrojen bağları bulunmamaktadır [16] [17].

Düzensiz bölgeler içeren proteinler kanser, alzheimer, kardiyovasküler ve diyabet gibi metabolik hastalıklara sebep olmaktadır. Ayrıca düzensiz bölgeler, transkripsiyon ve translasyonun regülasyonu, protein - DNA ve protein - protein birleşimlerinde önem taşımaktadır [1] [2]. Bu nedenlerle proteinlerde düzensiz bölgelerin kestirimi ilaç geliştirilmesi açısından son derece önemlidir [18]. Düzensiz bölgelerin doğru tahminini sağlamak için yeni tekniklerin geliştirilmesi gerekmektedir ve böylece üst seviye yeni ilaçların geliştirilmesinin yolu açılacaktır.

Literatürde proteinlerdeki düzensiz bölgelerin kestirimi için üç farklı yaklaşım vardır. Bu yöntemlerden ilki canlı organizmaları üzerinde yapılan çalışmalardır (in vivo). İkincisi laboratuvar ortamlarında yapılan çalışmalardır (in vitro). Üçüncüsü ise bilgisayar ortamında yapılan çalışmalardır (in silico). In vivo ve in vitro ortamlarda yapılan çalışmalar hem laboratuvar ortamının deneyleri etkileyen insan hatalarına

açık olması hem de maliyetin yüksek olması bu ortamların büyük kısıtlarıdır. Ancak in silico ortamda yapılan deneysel çalışmalar hem problemi daha iyi simüle etmekte hem de çok yönlü yaklaşımlar bilgisayar yöntemleriyle uygulanabilmektedir.

2.3. DisProt ve PDB Veri Setleri

Bu çalışmada, proteinlerin düzensiz bölgelerinin kestirimi için DisProt (versiyon 6.02) [15] ve Protein Data Bank (PDB) [20] veri setleri kullanılmıştır. DisProt veri setinden düzensiz bölgelere sahip dizilimler, PDB'den tamamı düzenli bölgelere sahip protein dizilimleri alınmıştır. Çizelge 2.3'de de görüldüğü gibi DisProt'tan 96, PDB'den 273 olmak üzere 369 protein dizilimi kullanılmıştır.

Çizelge 2.3 : DisProt ve PDB Verisetlerine ait bilgiler.

Veri Seti	DisProt	PDB
Kullanılan Protein Sayısı	96	273
En Küçük Kalıntı Sayısı	49	43
En Büyük Kalıntı Sayısı	1861	926

3. LYAPUNOV ÜSTELLERİ

Yaşam içerisinde bir çok olayın içeriğinde düzensizlik, karmaşıklık ve rastgelelik görülmektedir. Günlük yaşamda bu tip düzensizliklerin meydana getirdiği olaylara Kaos ismi verilir. Kaotik sistemler, başlangıç koşullarına hassas bağıllık gösteren ve ölçülemeyen sistemlerdir [21]. Kaos bilimine göre, kaotik bir sistemin başlangıç koşullarındaki küçük bir değişiklik, sistemin gelecekteki durumunda büyük bir değişikliğe yol açabilmektedir.

Burada dikkat edilmesi gereken iki nokta vardır [22]:

1. Kaotik sistemin gelecekteki durumunda başlangıç koşulları hassas bir şekilde bağıllık göstermesi,
2. Sistemi kaotikleştiren duruma sahip başlangıç değişkenlerinin sayısında ve karmaşıklığında büyük bir artış gerçekleşmesi.

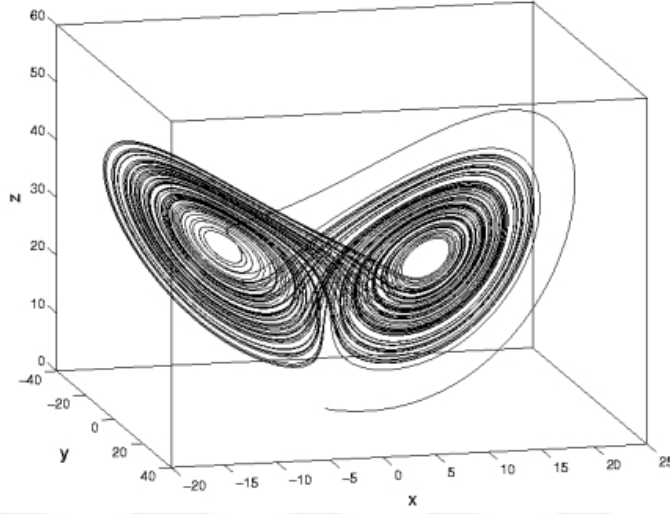
Kaotik sistemler tüm girdileri değerlendirip tüm sonucu ortaya koyarlar. Ortamı kaotik yapan en önemli etken ise değişkenlerin çok sayıda olmasıdır. Bu sebepten dolayı, kaotik yapılar rastgele her durumu almayıp, belli bir olasılıklar kümesi içerisinde hareket etmek zorunda kalırlar [23].

Özetle, kaos oldukça karmaşık bir düzendir. Bu durum deterministik kaos olarak bilinmektedir. Dolayısıyla, kaotik sistemler hem bir düzene uyarlar ve rastgele değildirler hem de başlangıç koşullarına göre çok hassas bir şekilde bağımlıdırlar [22].

3.1. Kaotik Çekerler

Bir sistemin kaotik davranışının sebebi çok fazla sayıda değişkenden etkilenmesidir. Kaos teorisinin matematiksel alt yapısını ilk olarak Edward Lorenz gerçekleştirmiştir. Lorenz, meteoroloji uzmanıdır ve bir çalışmasında deterministik kurallar üzerine eğilmiştir. Farkında olmadan çalıştığı sistem üzerindeki girdi değerlerinde ufak değişiklikler yapmıştır ve önceki çalışmalarına oranla farklı sonuçlar edinmiştir. Lorenz, bu farklılığı kaotik çekerler olarak adlandırmıştır.

Kaotik çekerler grafiksel gösterimler ile izah edilebilmekte olup, Şekil 3.1'de kelebek biçimli Lorenz çekerini gösterilmektedir.



Şekil 3.1 : Lorenz çeker.

Sistemin alabileceği muhtemel tüm durumları gösteren grafik alanı faz uzayıdır. Bir faz uzayının yapılandırılabilmesi için

$$X(t) = \{ x(t), x(t+\tau), \dots, x[t+(m-1)\tau] \} \quad (3.1)$$

formülü kullanılır [21]. Zaman gecikmesi olan (τ) ve yerleştirme boyutu olan (m) hesaplamalarda belirtilmesi gerekmektedir. Faz uzayındaki zaman serisinin analizinde zaman gecikmesi olan τ 'nin belirlenmesi ile yerleştirme boyutu m 'nin de değerleri ile birlikte kaotik zaman serisi incelenebilir. Bu iki kriterin sağlanması için öncelikle τ 'nin gerçekçi ve doğru seçilmesi gerekiyor. Bunun sebebi τ 'nin gerçekten uzak, büyük ya da küçük seçilmesi faz uzayını yapılandırmaz [21].

Şekil 3.1'deki Lorenz çekerinde faz uzayı içerisindeki zamana bağlı evrimler görülmektedir. Burada, Lorenz çekerlerinin alt yapısında kararsız bir denge bulunmaktadır. Merkezde bir nokta bulunmaktadır ve eğimler hep bu noktadan tekrar bir şekilde geçerler, aynı zamanda iki yöne yayıldıktan sonra bir eğim çizerler ve ardından tekrar merkez noktasına geri dönerler.

Kaotik çekerleri oluşturan parametlerin birbirlerine göre zamansal evrimine bakıldığında, çizgilerin asla iki kez aynı rotayı izlemediği, rotalar arasında hep bir farklılığın bulunduğu görülmektedir. Görülen bu yapı fraktal olarak adlandırılmaktadır. Kaotik çekerler tam sayı olmayan bir boyuta sahip olduklarından

dolayı fraktal boyut halini almışlardır. Bu boyuta sahip küme ise fraktal olarak isimlendirilir. Fraktal boyut, Mandelbrot tarafından geliştirilmiştir. Kaotik sistem çekerleri kesirli boyutlara sahip olmalarından dolayı fraktal boyut hesaplamaları sistemin kaotik olabileceğine ilişkin bir kriter olarak alınabilmektedir [24].

Bu fraktal yapıya ait Lorenz'in diferansiyel denklemleri,

$$\frac{dx}{dt} = \sigma (y - x) \quad (3.2)$$

$$\frac{dy}{dt} = x (\rho - z) - y \quad (3.3)$$

$$\frac{dz}{dt} = xy - \beta z \quad (3.4)$$

formülleriyle verilmiştir [25]. Lorenz denkleminde değişkenler ve parametreler Çizelge 3.1'de tanımlanmaktadır;

Çizelge 3.1 : Lorenz denkleminin değişkenleri ve parametreleri.

Değişkenler	
x	Konvektif hareketinin şiddetiyle orantılıdır.
y	Yükselen ve çöken akımlar arasındaki sıcaklık farkı ile orantılıdır.
z	Düşey sıcaklık profilinin lineer konumdan sapması ile orantılıdır.
Parametreler	
σ	Prandtl sayısı
ρ	Rayleigh sayısı

Prandtl sayısı akışkanın ısı geçiş özelliğini gösteren sayıdır ve Rayleigh sayısı da akışkanlar dinamiği ve ısı transferinde kullanılan boyutsuz bir sayı olan Grashof sayısı ile Prandtl sayısının çarpımı ile hesaplanabilmektedir. Lorenz, $\sigma = 10$, $\rho = 28$ ve $\beta = 8 / 3$ parametre değerleri için sistemin kaotik davranış ortaya koyduğu çözümleri elde etmiştir.

3.2. Kaos Ölçme Yöntemleri

Kaotik bir sistemin zaman içerisinde değişen parametrelerini gösteren ve bu sistemin zaman içerisinde nasıl bir davranış gösterdiğine ait var olan verilere zaman serileri adı verilmektedir. Bir zaman serisinin kaotik olup olmadığı, genellikle o sistemin davranışının faz uzayındaki görünümünün elde edilmesiyle incelenebilir. Matlab gibi yazılımların yardımıyla bir dizi karmaşık hesap gerektiren bu süreçler kolay bir şekilde algoritmalarla çözümlenebilmektedir.

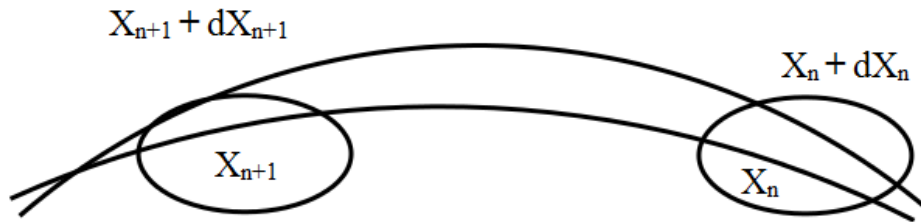
Çekerlerin oluşumunda gömme boyutu parametresinin bilinmesi gerekir. Gömme boyutu, kaotik bir sistemin davranışlarını etkileyen bağımsız dinamik kaynakların sayısını tahmin eden bir hesaplama yöntemidir. Üç boyuttan daha fazla gömme boyutları pek tercih edilmese de bazı karmaşık kaotik sistemlerde çok daha büyük boyutlu faz uzaylarına ihtiyaç duyulabilmektedir.

Gömme boyutu hesaplanması sonucunda zaman serisinin hangi zaman aralıklarında geciktirileceğine dair parametreyi de zaman gecikmesi belirlemektedir.

3.3. Lyapunov Üstellerinin Hesaplanması

Lyapunov üstelleri (LÜ), başlangıç şartlarına hassas bağlılık özelliğinin sayısal bir göstergesidir. LÜ, faz uzayındaki yörüngelerin birbirlerine olan mesafelerindeki yakınsama veya ıraksama hızlarını belirler. x boyutlu bir faz uzayı yine x adet farklı LÜ'ne eşit olmaktadır. Bu LÜ değerlerinden sadece bir tanesinin bile sıfırdan büyük olması o sistemin kaotik olmasının en önemli kanıtıdır. Bu kanıtın diğer bir önemli ayrıntısıysa, yörüngelerin üstel olarak ıraksamasıdır. Sonuçların zaman içerisindeki başlangıç koşullarına göre farklılıklar göstermeside aşırı duyarlılığı ile kaos betimlemektedir [26].

Faz uzayındaki yörüngeler, kararlı noktalardan oluştuğu gibi aynı zamanda tüm noktalar tek yörünge üzerine yayılmaktadır. Eğer LÜ'nden birisi λ pozitif değerliyse sistem kaotiktir denilmektedir. Bununla birlikte faz uzayındaki iki komşu λ değerlerine bağlı olacak şekilde birbirlerine ıraksayacaklardır [21] [22] [27].



Şekil 3.2 : Seçilen birbirine yakın iki ilk koşul ve evrimleri.

Başlangıçta birbirine ΔX_0 kadar yakın olan iki komşu nokta olsun. Bu noktaları sırasıyla X_0 ve $X_0 + \Delta X_0$ biçiminde temsil edilsin, bir tekrarlardan sonra bu iki komşu nokta Şekil 3.2'de ki gibi ayrışır [28].

$$\Delta X_I = f(X_0 + \Delta X_0) - f(X_0) \cong \Delta X_0 f'(X_0) \quad (3.5)$$

burada $f' = df/dX$ türevidir. X_0 noktasında uzaklaşmayı ifade eden LÜ λ tanımlanır. $e\lambda = |\Delta X_I / \Delta X_0|$ veya

$$\lambda = \ln |\Delta X_I / \Delta X_0| \cong \ln |f'(X_0)| \quad (3.6)$$

biçiminde ifadelendirilir. Yerel LÜ olarak isimlendirilen $|\Delta X_I / \Delta X_0|$ büyüklüğü $X - X_0$ 'daki gerilmeyi gösterir. Bu ifadenin mutlak değer içinde olması LÜ'nin pozitif olmasını temin eder [28].

$$\lambda = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \ln |f'(X_n)| \quad (3.7)$$

Bir bütün olarak sistemin LÜ'ü 3.7'deki denklemin birçok tekrarlama sonrasında ortalaması alınarak hesaplanır [28].

Time Series Analysis olarak bilinen Tisean yazılımı, açık kaynak kodlu ve ücretsizdir. Lineer olmayan belirleyici dinamik sistem temelli metotlar ile zaman serilerinin analizi yapılabilmekle birlikte, LÜ metodlarına ait algoritmalar Tisean paketi içerisinde bulunmaktadır [29]. Bu tez çalışması esnasında, dinamik sistemlerden LÜ ile ilgili algoritmalar Tisean paket yazılımı aracılığıyla Matlab ortamında hesaplanmıştır.

Çizelge 3.2'da Tisean yazılımında kullanılan LÜ ile ilgili kodlar ve açıklamaları bulunmaktadır.

Çizelge 3.2 : LÜ Metotları.

Açıklama	Program
Maximal Üstel	lyap_k , lyap_r
Lyapunov Spektrumu	lyap_spec
Boyutu Sonlu Üsteller	Fsle

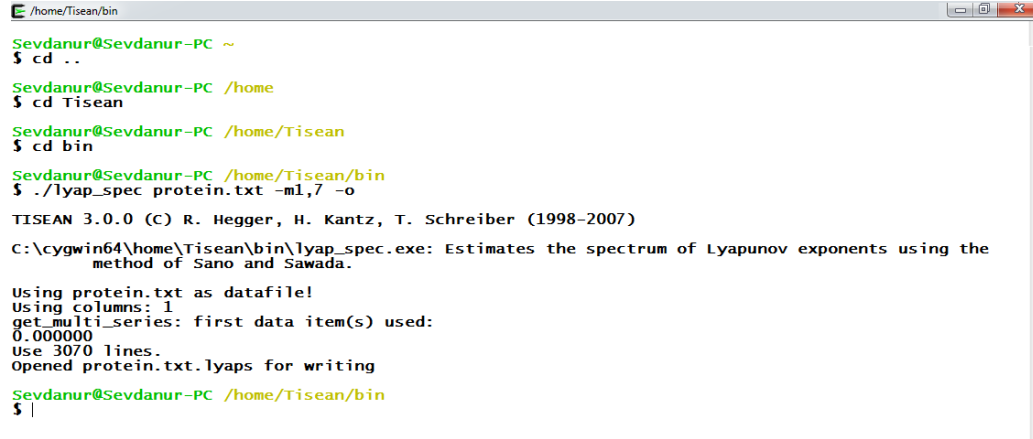
3.3.1. Lyapunov Spektrumu ve analizi

Tisean yazılımında Lyapunov Spektrum (lyap_spec) algoritması analiz yapmak amacıyla kullanılmaktadır.

Lyapunov Spectrum hesaplanması sadece maksimal üstten bile çok daha fazla çaba gerektirmektedir. Tisean yazılımı sayesinde LÜ değerlerinin spektrumlarıyla ilgili

tüm tahminleri hesaplayabilmektedir. Eğer d komponentleri ve gömme boyutu m değeri verilmişse $m*d$ üstel değerleri hesaplanabilmektedir [30].

Tisean yazılımı tarafından sistemin ekran çıktılarında ise $d*m+1$ sütunlar halinde oluşmaktadır. Birinci sütunda güncel yinelemeleri gösterirken diğer sütunlarda ise LÜ değerleri tahmin edebilmektedir. Listenin alt satırlarında ise yerel doğrusal modelin ortalama tahmin hataları, modelin montajı için kullanılan ortalama komşuluk boyutu ve son olarakta Kaplan - Yorke boyut tahminleri gösterilmektedir.



```
Sevdanur@Sevdanur-PC ~
$ cd ..
Sevdanur@Sevdanur-PC /home
$ cd Tisean
Sevdanur@Sevdanur-PC /home/Tisean
$ cd bin
Sevdanur@Sevdanur-PC /home/Tisean/bin
$ ./lyap_spec protein.txt -m1,7 -o
TISEAN 3.0.0 (C) R. Hegger, H. Kantz, T. Schreiber (1998-2007)
C:\cygwin64\home\Tisean\bin\lyap_spec.exe: Estimates the spectrum of Lyapunov exponents using the
method of Sano and Sawada.
Using protein.txt as datafile!
Using columns: 1
get_multi_series: first data item(s) used:
0.000000
Use 3070 lines.
Opened protein.txt.lyaps for writing
Sevdanur@Sevdanur-PC /home/Tisean/bin
$ |
```

Şekil 3.3 : lyap_spec komutunun LÜ hesaplaması için kullanımı.

Lyap_spec komutu ile LÜ'nin tüm spektrum tahminleri bulunmaktadır. Örneğin;

> lyap_spec 1-a_n.txt -m 1, 7 - o

komutu kullanılarak LÜ değerleri bulunmaktadır. Çizelge 3.3'de Tisean yazılımında kullanılan lyap_spec komutuna ait parametre ve açıklamaları bulunmaktadır.

Çizelge 3.3 : .lyap_spec parametreleri ve açıklamaları.

Seçenek	Açıklama	Varsayılan
-l#	Kullanılan veri sayısı	Dosyanın Tamamı
-x#	Göz ardı edilecek olan satırlar (ilk # tanesi)	0
-c#	Okunacak Sütun	1
-m#	Gömme boyutu aralığı	1, 2
-d#	Gecikme vektörleri için gecikme	1
-r#	Komşuluk başlangıç boyutu çalışma sırasında dinamik olarak değişir.	(Veri Aralığı) / 1000
-f#	Yeterli olmayan komşular bulunursa, komşuluk boyutunu artırma faktörü	1,2
-k#	Kullanılacak komşu sayısı	30
-n#	Yineleme sayısı	Noktaların Sayısı
-l#	Zaman serisinin sırasını ters çevirme, sahte üstelleri bulmada yardımcı olur.	Ters Çevirme Yok
-o[#]	Çıktı dosyasının adı Ayrıntı düzeyi;	Giriş_dosyasının_adi.lyaps
-v#	0: sadece acil mesajlar 1: giriş/çıkış mesajları eklenebilir	1
-h	Programın seçeneklerini gösterir	-

Tisean yazılımında kullanılan lyap_spec komutu sayesinde verilen serinin LÜ'ne ait sonuçlar Şekil 3.7'de ki gibi gösterilmektedir.

```

/home/Tisean/bin
Sevdanur@Sevdanur-PC /home/Tisean/bin
$ ./lyap_spec protein.txt.lyaps
TISEAN 3.0.0 (C) R. Hegger, H. Kantz, T. Schreiber (1998-2007)
C:\cygwin64\home\Tisean\bin\lyap_spec.exe: Estimates the spectrum of Lyapunov exponents using the
method of Sano and Sawada.
Using protein.txt.lyaps as datafile!
Using columns: 1
get_multi_series: first data item(s) used:
3063.000000
Line 2 ignored: #Average relative forecast errors:= 1.261866e-001
Line 3 ignored: #Average absolute forecast errors:= 4.274843e-002
Line 4 ignored: #Average Neighborhood Size= 1.391533e-001
Line 5 ignored: #Average num. of neighbors= 3.000000e+001
Line 6 ignored: #estimated KY-Dimension= 2.551171
Use 1 lines.
rescale_data: data ranges from 3.063000e+003 to 3.063000e+003. It makes
no sense to continue. Exiting!
Sevdanur@Sevdanur-PC /home/Tisean/bin
$

```

Şekil 3.4 : lyap_spec kullanımı sonucu LÜ sonuçları.



Örüntü tanıma sistemlerinde örüntüye ait ölçülebilir veya gözlenebilir bilgi olarak öznitelikler tanımlanabilir. Şekil 4.1'de de görüldüğü gibi örüntü verilerinin eğitim sürecinde iken özneliğin çıkartılması veya seçilmesiyle birlikte en uygun öznitelikler ayırt edilmektedir, ardından sınıflandırıcı öznitelik uzayını çeşitli bölümlere ayırabilmektedir. Geri besleme ile önışlem ve öznitelik çıkartılması veya seçilmesi optimize edilebilir. Sınıflandırma aşamasında ise, giriş örüntülerinin öznitelik ölçümlerine göre hangi sınıflara ait olduğuna karar verir [11].

4.1. Sınıflandırma

Örüntü tanımanın eğitim süresinde, öznitelik çıkarımından sonra karar verme aşaması olan sınıflandırma aşaması gerçekleşecektir. Bu aşamada, nesnelere sınıflandırılması için bir karar fonksiyonu elde edilir ve bu fonksiyona göre nesnelere sınıflandırılır.

N adet kümeden oluşan bir deneysel gözlem verileri için $D = \{x_i, y_i\}_{i=1,2,\dots,N}$ ve $x_i \in R^d$. Kümede bulunan tüm deneysel veriler birbirinden bağımsız ve özde dağılımlara sahiptirler. Sınıflandırma işlemi gerçekleştirildikten sonra her bir x_i 'nin sınıfına karar verilmiş olur. Burada y_i , x_i sınıfına karşılık gelen sınıf bilgisidir (etiket). Bu etiketlerin varlığı sınıflandırma işlemi için şarttır.

4.2. Öznitelik Çıkarımı

Özneliğin çıkarımı sınıflandırma açısından önemli bir rol oynamaktadır. Bunun sebebi, örüntüye ait karakteristik özellikler ne kadar net temsil edilirse sınıflandırıcı örüntüyü o kadar iyi tanır. Şekil 4.1'de de görüldüğü gibi bir örüntü tanıma sisteminde özellik çıkarımı önışlem adımında gerçekleştirilir. Eğer bir örüntü birden fazla öznitelik ile temsil ediliyorsa, tek bir öznitelik yerine bir öznitelik kümesi söz konusudur. n adet özneliğe sahip öznitelik kümesi ise n - boyutlu öznitelik vektörü ile temsil edilebilir. Özniteliklerin içinde bulunduğu n - boyutlu R uzayı ise öznitelik uzayı olarak isimlendirilmektedir [31][32]. Aynı zamanda örüntü çıkarımı, algoritma karmaşıklığı ve boyut problemi açısından etkin bir iyileştirme sağlamaktadır.

Proteinleri oluřturan amino asitlerin sayısal olarak temsili sınıflandırıcı algoritmanın örüntüyü tanımada önemli bir etkidir. Proteinlerin temsilinde amino asitlerin protein içindeki yeri, sayısı veya fizikokimyasal özellikleri öznitelik vektörler kümesi olarak ifade edilebilir [11].

Biyoinformatik alanında kullanılan belli başlı öznitelik yöntemleri řunlardır: birimdik, ağırlık tabanlı, yer deęiřtirme matrisleri tabanlı, n grams, kalıntı çiftleri, OETMAP ve OEDICHO [33].





5. GELİŞTİRİLEN ÖZİNİTELİK KODLAMA YÖNTEMLERİ

Proteinlerde düzensiz bölgelerin tahmini için DisProt ve PDB veri setlerindeki protein dizilimlerinin her bir fizikokimyasal özelliğe göre kaotik yapı içerip içermediği araştırıldı.

5.1. Protein Dizilimindeki Kaotik Yapının Analizi

Bir proteinin, bir fizikokimyasal özelliğe göre kaotik yapı içerip içermediğini anlamak için fizikokimyasal özelliğe ait değerler protein dizilimi içerisinde karşılık gelen kalıntının yerine yazılarak protein dizilimleri sayısal olarak ifade edildi. Böylece N adet amino asite sahip bir protein 544 x N boyutunda bir matris ile temsil edildi. İkinci olarak, Bölüm 2'de anlatıldığı gibi bu öznelik matrisine ait LÜ değerleri elde edildi.

Çizelge 5.1 : Bir fizikokimyasal özelliğe ait LÜ değerlerinin sonuçları.

Lyapunov Üstel Sonuçları

Length :	3848		
Signals belong to each amino acids of a protein :			
9.044008e-001	4.018379e-001	2.226827e-001	1.310360e-001
4.904960e-002	-1.972889e-002	-1.013593e-001	-1.959663e-001
-3.338790e-001	-7.586726e-001		
Average relative forecast errors :	1.157800e+000		
Average absolute forecast errors :	2.145047e+001		
Average Neighborhood Size :	2.398085e+001		
Average num. of neighbors :	3.000000e+001		
Estimated KY-Dimension :	10.000000		

Protein dizilimlerinin kaotik bir yapı sergiledikleri, 10 LÜ değerlerinden en az birinin artı sonuç olması ile anlaşılmaktadır. Ancak daha kararlı kaotik yapı elde edebilmek için sonuçların yarı yarıya işaret değişimlerinin olması gerektiği kriterini göz önüne aldık. Örneğin, Çizelge 5.1'de görüldüğü gibi beş artı beş eksi sonucun gelmesi, ilgili fizikokimyasal özelliğin sadece bu protein yapısında kararlı bir kaotik yapı dağılımı gösterdiğinin ispatıdır. Nitekim ilgili fizikokimyasal özelliğe göre

kodlanan protein dizilimine faz uzayında bakıldığında Lorenz dağılımı gibi kaotik, pseudo periyodik bir yapının olduğu da görülmüştür.

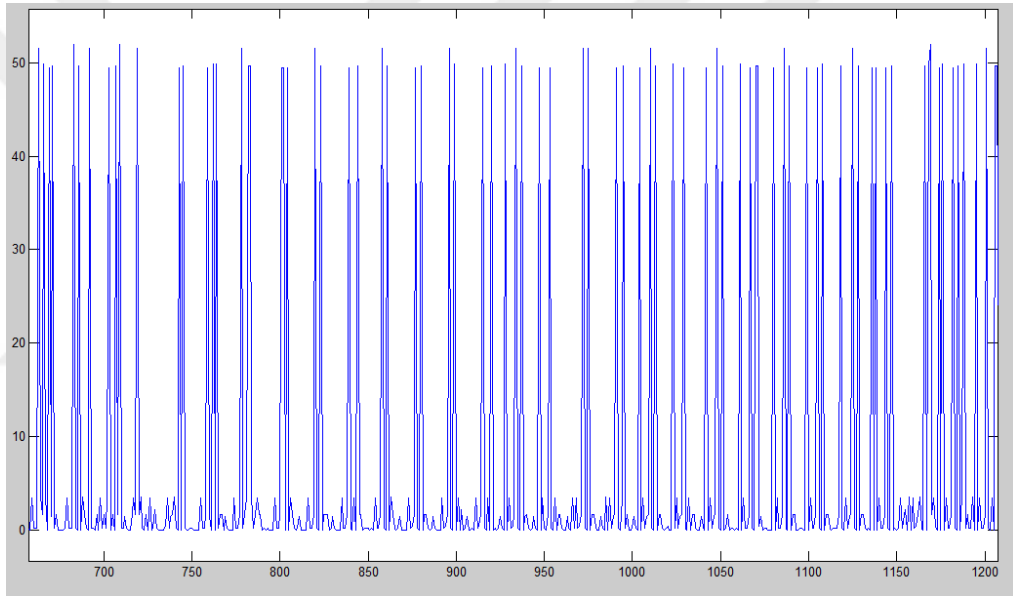
Başka veri setlerinde ve bu veri setlerine ait olan proteinler de farklı sonuçlara rastlanılabilir. Diğer bir ifade ile fizikokimyasal özelliğin kaotik yapı gösterip göstermeyeceği proteine özgüdür.

Çizelge 5.2: Protein dizilimlerinde düzensiz bölgeler üzerinde belirlenen LÜ değerleri.

FK Özellikler	Veri Setindeki Tekrar Sayısı	FK Özellikler	Veri Setindeki Tekrar Sayısı	FK Özellikler	Veri Setindeki Tekrar Sayısı
3	1	156	2	393	1
4	1	168	1	395	287
7	1	173	3	396	278
13	1	174	42	397	1
14	1	175	19	400	34
16	312	197	1	401	10
18	85	199	1	419	13
20	1	201	1	423	11
25	1	205	1	424	115
46	1	206	1	426	1
50	1	211	1	432	2
51	1	212	1	433	7
64	1	219	1	437	103
70	10	233	1	453	314
90	2	241	2	454	1
93	238	246	1	456	1
94	1	248	1	462	1
96	33	257	1	468	1
108	1	282	1	470	1
111	1	312	1	472	1
116	1	313	1	476	13
126	13	324	1	481	1
133	3	326	1	482	1
134	1	336	36	495	2
136	1	343	1	504	1
137	1	346	1	509	36
141	1	362	1	512	1
142	1	365	2	530	1
144	1	369	3	531	1
148	1	372	1	533	1
151	1	374	1	538	2
155	267	383	2	539	1

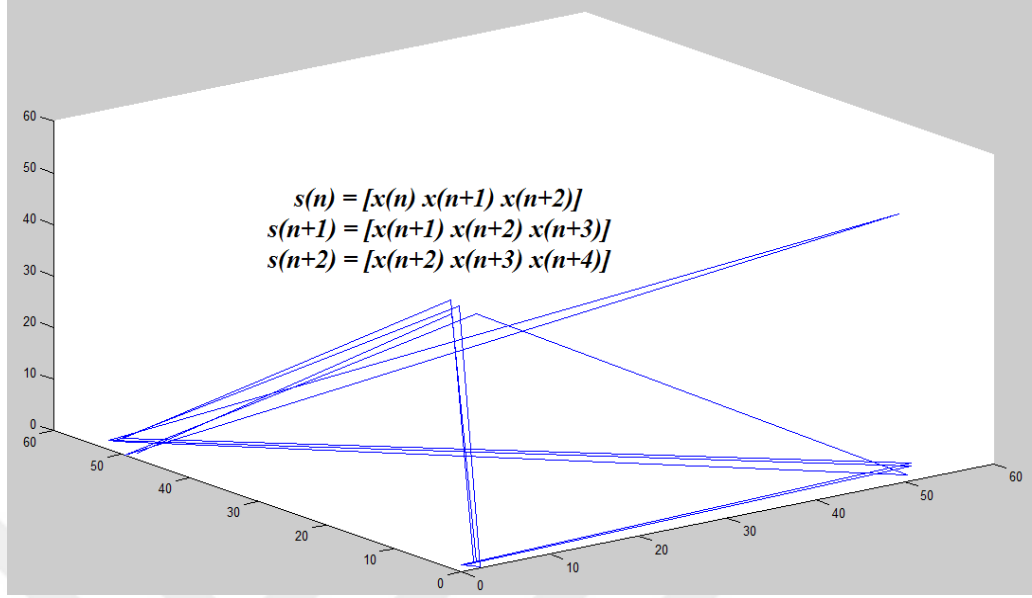
Matlab ve Tisean yazılım ortamlarında yapılan deneyler sonucunda her bir fizikokimyasal özelliğin veri setinde kaç protein dizilimi üzerinde kaotik yapı gösterdiği Çizelge 5.2'de görüldüğü gibi belirlenmiştir.

.Lyap dosya uzunluğundaki verilerin koordinat sistemindeki dağılımlarını incelemek için öncelik olarak bu özelliklerin iki boyutlu bir eksenle fraktal bir yapı sergileyip sergilemedikleri hakkında ayrı ayrı testler yapılmaktadır. Örneğin 3590 uzunluğundaki protein diziliminin 400. fizikokimyasal özelliğe ait amino asitlerin işaretlerinin bir seri üzerindeki değişimleri Şekil 5.1'de görüldüğü iki boyutlu eksen üzerinde elde edilmiştir.

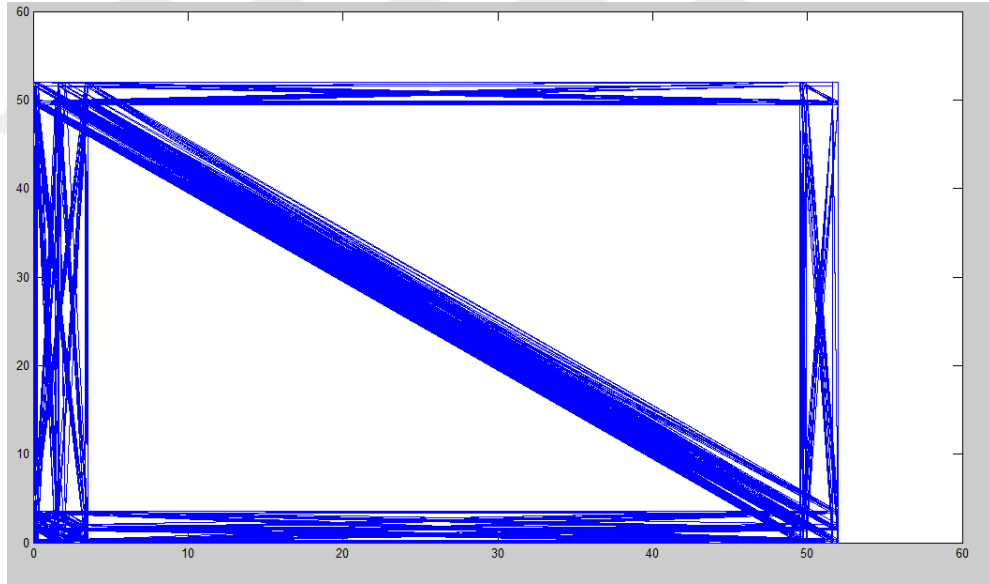


Şekil 5.1 : Amino asitlere ait işaretlerin bir seri üzerindeki değişimleri.

Şekil 5.2'de ise iki boyutlu eksen üzerinde ifade edilen protein diziliminin detayı görülmektedir.

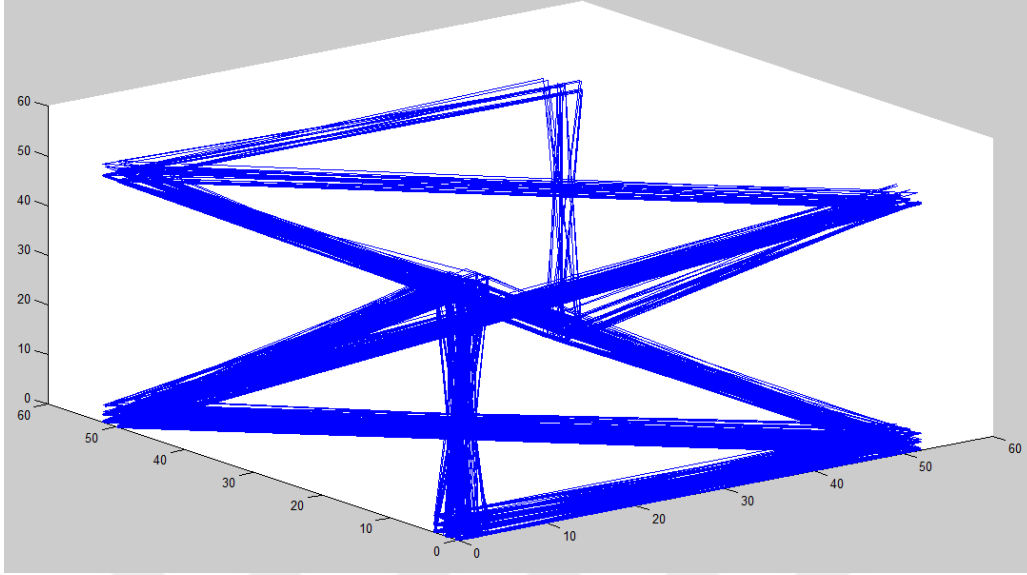


Şekil 5.3 : Bir fizikokimyasal özelliğinin faz uzayındaki kaotik yapısı.



Şekil 5.4 : Bir fizikokimyasal özelliğe ait tüm LÜ işaretlerinin $D_E = 2$ boyutlarındaki faz uzayında gösterimi.

Şekil 5.3 incelendiğinde ilk 20 LÜ işaretinin kaotik bir yapı sergilediği görülüyor. Seçilen fizikokimyasal özelliğinin tüm işaretlerinin $D_E = 2$ faz uzayındaki gösterimi Şekil 5.4 'de gösterilmektedir. Şekil 5.5' de ise $D_E = 3$ faz uzayındaki görünümü bulunmaktadır.



Şekil 5.5 : Bir fizikokimyasal özelliğe ait tüm LÜ işaretlerinin $D_E = 3$ boyutlarındaki faz uzayında gösterimi.

5.2. En İyi Fizikokimyasal Özelliklere Göre Öznitelik Kodlama Yöntemi

En iyi fizikokimyasal özelliklere göre öznitelik kodlama yöntemi (FKÖY) Çizelge 5.3'te görülen en çok kaotik yapı özelliği gösteren 6 fizikokimyasal özelliğe göre öznitelik kodlama yapılmıştır. Buna göre her bir fizikokimyasal özelliğin her bir protein dizilimine göre 1×10 boyutunda bir vektörü elde edilmiştir. Böylece 6 fizikokimyasal özellik için 6×10 boyutunda bir matris elde edilmiştir. Sınıf bilgisi de eklenince matris 6×11 boyutunda olmaktadır.

Tez çalışmasında, hem AAIndex hem de LÜ değerleri, z-skor normalizasyon yöntemi [34] ile normalize edilmiştir.

Çizelge 5.3 : FK için LÜ değerleri.

	1.LÜ	2.LÜ	3.LÜ	4.LÜ	5.LÜ	6.LÜ	7.LÜ	8.LÜ	9.LÜ	10.LÜ
FK - 1	0.297	0.138	0.061	0.048	0.011	-0.018	-0.031	-0.071	-0.179	0
FK - 2	0.643	0.454	0.135	0.053	-0.023	-0.048	-0.143	-0.214	-0.285	-0.783
FK - 3	-0.004	-0.041	-0.029	-0.06	-0.046	-0.128	-0.095	-0.169	-0.296	-1.144
...
FK - 6	-0.038	-0.054	-0.049	-0.077	-0.107	-0.158	-0.163	-0.258	-0.342	-0.835

5.3. Lyapunov Üstelleri ve En İyi Fizikokimyasal Özelliklere Göre Öznitelik Kodlama Yöntemi

Lyapunov Üstelleri ve En İyi Fizikokimyasal Özelliklere Göre Öznitelik Kodlama Yöntemi (LÜ-FKÖY) göre Çizelge 5.2'de görülen en çok kaotik yapı özelliği gösteren 6 fizikokimyasal özelliğe göre öznitelik kodlama yapılmıştır. Buna göre her bir fizikokimyasal özelliğin her bir protein dizilimine göre 1×10 boyutunda bir vektörü elde edilmiştir. Böylece altı fizikokimyasal özellik için 6×10 boyutunda bir matris elde edilmiştir. Sınıf bilgisi de eklenince matris 6×11 boyutunda olmaktadır.

ACDEE peptitini yöntemimiz ile kodlayalım. Önce proteine ait 1×10 boyutunda her bir FK için Çizelge 5.4'de görüldüğü gibi LÜ değerleri elde edilir.

Çizelge 5.4 : En iyi FK için LÜ değerleri.

	1.LÜ	2.LÜ	3.LÜ	4.LÜ	5.LÜ	6.LÜ	7.LÜ	8.LÜ	9.LÜ	10.LÜ
FK - 1	0.297	0.138	0.061	0.048	0.011	-0.018	-0.031	-0.071	-0.179	0

İkinci adımda AAIndex'den en iyi FK karşısına gelen her bir AA değeri Çizelge 5.5'te görülen LÜ değerleri ile çarpılır. Örneğin

Çizelge 5.5 : En iyi FK için AAIndex değerleri.

	A	C	D	Y
FK - 1	121.9	121.4	117.5	121.7

Bu durumda örneğin A amino asiti için,

Çizelge 5.6 : En iyi FK'nın AAIndex değeri ile çarpımı.

	1.LÜ	2.LÜ	3.LÜ	4.LÜ	5.LÜ	6.LÜ	7.LÜ	8.LÜ	9.LÜ	10.LÜ
FK - 1	36.204	16.847	7.533	5.948	1.426	-2.255	-3.778	-8.764	-21.905	0

biçiminde 1×10 'luk kalıntıya özgü vektör elde edilir. Bu kodlama işlemi proteindeki diğer kalıntılar için tekrarlanır böylece beş amino asite sahip bir protein için 30×10 boyutunda bir öznitelik matrisi elde edilir.

5.4. Performans Değerlendirmesi

Proteinlerin düzensiz bölgelerinin kestirimi için geliştirilen FKÖY öznelik çıkarım yöntemi DisProt ve PDB veri setleri üzerinde uygulanmıştır. Toplamda 369 kalıntı üzerinde altı en iyi FK'nın kaç adet kalıntıda rastlanıldığına dair tekrar sayıları Çizelge 5.7'de sunulmuştur.

Çizelge 5.7 : Öznelik çıkarım sonuçları.

FK Özellik	Veri Setindeki Tekrar Sayısı
16	312
93	238
155	267
395	287
396	278
453	314

Performans metrikleri WEKA yazılım platformunda 10 kat çapraz doğrulama test tekniği ile elde edilmiştir. 10 kat çapraz doğrulama test tekniğinde veri seti rastgele 10 kümeye bölünür. Her bir kümedeki düzensiz protein sayısı ve düzenli protein sayısı eşittir. Her bir küme için Çizelge 5.8'de görülen karmaşıklık matrisi elde edilir.

Çizelge 5.8 : Karmaşıklık Matrisi.

		Öngörülen Sınıf	
		Sınıf 1 (Artı)	Sınıf 2 (Eksi)
Gerçek Sınıf	Sınıf 1 (Artı)	Doğru Artı (DA)	Yanlış Eksi (YE)
	Sınıf 2 (Eksi)	Yanlış Artı (YA)	Doğru Eksi (DE)

Karmaşıklık matrisinden sınıf doğruluğu (SD), duyarlılık, özgünlük, Mathew'in korelasyon katsayısı (MCC), f-skor performans metrikleri elde edilir:

$$SD = \frac{(DA+DE)}{(DA+YA+DE+YE)} \quad (5.2)$$

$$Duyarlılık = \frac{DA}{DA+YE} \quad (5.3)$$

$$Özgünlük = \frac{DE}{DE+YA} \quad (5.4)$$

$$MCC = \frac{(DAxDE-YAxYE)}{\sqrt{(DA+YA)x(YA+DE)x(DE+YE)x(YE+DA)}} \quad (5.5)$$

$$F - Skor = \frac{2 (DA)}{2 (DA) + (YE) + (YA)} \quad (5.6)$$

Test bittiğinde performans metriklerinin ortalaması alınır.

Geliştirilen FKÖY ve LÜ-FKÖY öznitelik kodlama yöntemleri, Navie Bayes [35], BayesNet [36], Random Forest [37], kNN [38], C4.5 [39], SMO [40], AdaBoost [41] ve MLP [42] sınıflandırma algoritmaları ile DisProt ve PDB veri setleri üzerinde test edilmiştir. Çizelge 5.9'da FKÖY yönteminin performans değerleri görülmektedir.

Çizelge 5.9 : FKÖY 'e ait sınıflandırma sonuçları.

Sınıflandırıcı	SD	Duyarlılık	Özgünlük	MCC	F-Skor
kNN	72,72	0,727	0,661	0,107	0,197
NavieBayes	73,92	0,739	0,740	-0,006	0,0013
BayesNet	69,67	0,697	0,617	0,097	0,268
Random Forest	70,87	0,709	0,607	0,126	0,284
C4.5	74,62	0,746	0,709	0,128	0,083
SMO	73,99	0,740	0,740	-	0
AdaBoost	73,99	0,740	0,740	-	0
MLP	74,50	0,745	0,725	0,120	0,039

Elde edilen deneysel sonuçlara göre FKÖY yöntemi C4.5 sınıflandırma algoritması değerleri ile en iyi performansı sergilemiştir. Buna karşın, BayesNet yöntemiyle en düşük değerleri vermiştir.

LÜ-FKÖY yönteminin performans değerleri Çizelge 5.10'da görülmektedir.

Çizelge 5.10 : LÜ-FKÖY 'e ait sınıflandırma sonuçları.

Sınıflandırıcı	SD	Duyarlılık	Özgünlük	MCC	F-Skor
kNN	99,87	0,999	0,003	0,996	0,997
NavieBayes	77,23	0,772	0,772	0,007	0,013
BayesNet	99,50	0,995	0,016	0,985	0,988
Random Forest	99,99	1	0	0,999	0,999
C4.5	99,95	1	0,001	0,998	0,999
SMO	77,23	0,772	0,772	-	0
AdaBoost	86,50	0,865	0,424	0,580	0,605
MLP	89,36	0,894	0,342	0,680	0,705

Elde edilen deneysel sonuçlara göre LÜ-FKÖY yöntemi Random Forest sınıflandırma algoritması değerleri ile en iyi performansı sergilemiştir. Buna karşın, hem Navie Bayes hem de SMO yöntemleriyle en düşük değerleri vermiştir.

Çizelge 5.11'de görülen proteinlerin düzensiz bölgelerin tespiti ile ilgili literatürde yer alan çalışmalar, geliştirdiğimiz LÜ-FKÖY ile karşılaştırılmıştır.

Çizelge 5.11 : Literatürdeki test yöntemleri ile karşılaştırılması.

Çalışma Grubu	SD	Duyarlılık
DisEMBL [6]	61.9	25.9
RONN [5]	61.8	34.7
DISOPRED2 [7]	74.2	53.8
VSL1 [8]	80.7	75.6
VSL2B [9]	77.3	73.7
VSL2P [9]	79.4	78.4
LÜ-FKÖY	99.9	1

Elde edilen sonuçlara göre, geliştirdiğimiz yeni öznelik kodlama yöntemi hem sınıf doğruluğu hem de duyarlılık açısından en iyi sonuçları vermiştir.

6. SONUÇLAR

Bu tez çalışmasında, proteinlerin düzensiz bölgelerinin tespiti için kaos teorisi tabanlı fizikokimyasal özellikler seçilmiş ve bu fizikokimyasal özelliklere dayalı yeni iki tane FKÖY ve LÜ-FKÖY adı verilen öznelik kodlama yöntemleri geliştirilmiştir.

Öncelikle DisProt ve PDB verisetleri üzerinde kaotik özellik gösteren altı fizikokimyasal özellik belirlenmiştir. Bu belirlenen fizikokimyasal özellikler, daha sonra geliştirilen yöntemlerde de kullanılmıştır.

İkinci aşamada bu özellikler temelinden proteindeki her bir kalıntı FKÖY yönteminde kodlanmıştır.

Üçüncü aşamada ise kaotik özellik gösteren bu altı fizikokimyasal özelliğin elde edilen 10 LÜ değeri ile amino asit indeks değerleri birleştirilerek LÜ-FKÖY gerçekleştirilmiştir.

LÜ-FKÖY ile DisProt ve PDB verisetleri üzerinde % 99 sınıf doğruluğu ve 0,99 f - skor performans değerleri elde edilmiştir. Literatürdeki benzer çalışmalarla kıyaslandığında LÜ-FKÖY 'nin proteinlerin düzensiz bölgeleri tespiti probleminde en iyi sonucu vermiştir.

Gelecekte yapılacak olan çalışmalarda,

a. Geliştirilen FKÖY ve LÜ-FKÖY yöntemlerine PAM ya da BLOSUM yer değiştirme matrisleri bilgilerinin eklenmesi ile daha yüksek başarımlar elde edilebilir.

b. Birleştirilmiş (Ensemble) sınıflandırıcılar kullanılarak performans daha yukarı çekilebilir.

c. Geliştirilen öznelik kodlama yöntemlerinden elde edilen yüksek değerli deneysel sonuçlar, bu yöntemlerin başka protein örüntü tanıma çözümlerine de uygulanabileceğini göstermektedir.

d. FKÖY ve LÜ-FKÖY yöntemlerine dayanan araştırmacıların kullanabileceği bir web sunucuları hazırlanabilir.



KAYNAKLAR

- [1] **DYSON, H. J., WRIGHT, P. E.**, Nature Rev Mol Cell Biol, 6, 197-208, 2005.
- [2] **FENG, Z-P, ZHANG, X., HANB, P., ARORA, N., ANDERS, R. F., NORTON, R. S.**, Molecular & Biochemical Parasitology 150, 256–267, 2006.
- [3] **RINGE, D., PETSKO, G. A.**, 1986. Study of protein dynamics by xray diffraction. Methods Enzymol.131:389–433.
- [4] **UVERSKY, V. N., GILLESPIE, J. R., FINK, A. L.**, Why are natively unfolded proteins unstructured under physiologic conditions? Proteins.41:415–427, 2000.
- [5] **LINDING, R., JENSEN, LJ., DIELLA, F., BORK, P., GIBSON, TJ., RUSSELL, RB.**, Protein disorder prediction: implications for structural proteomics. Structure (Camb) 2003, 11:1453-1459.
- [6] **YANG, ZR., THOMSON, R., MCNEIL, P., ESNOUF, RM., RONN:** the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics 21:3369-3376, 2005.
- [7] **WARD, JJ., SODHI, JS., MCGUFFIN, LJ., BUXTON, BF., JONES, DT.**, Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol, 337:635-645, 2004.
- [8] **PENG, K.**, Length-dependent prediction of protein intrinsic disorder. BMC bioinformatics 7.1 (2006): 208.
- [9] **PENG, K., RADIVOJAC, P., VUCETIC, S., DUNKER, A.K., OBRADOVIC, Z.**, Length-Dependent Prediction of Protein Intrinsic Disorder, BMC Bioinformatics 7:208, 2006.

- [10] **OBRADOVIC, Z., PENG, K., VUCETIC, S., RADIVOJAC, P., DUNKER, A.K.**, Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins* 2005, 61(Suppl 7):176-182.
- [11] **GÖK, M.**, "HIV-1 Proteaz Enziminin Kesme Konumlarının Tespitinde Yeni Öznitelik Vektörleri", Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, 2011.
- [12] **BARNES, M., GRAY, I.**, *Bioinformatics for Genetics*. John Wiley & Sons Inc, 2003.
- [13] **KAWASHIMA, S., KANEHISA, M.**, "AAindex: amino acid index database, *Nucleic Acids Res.* 20 (1): 374, 2000.
- [14] **LEVINTHAL, C.** "Are there pathways for protein folding." *J. Chim. phys* 65.1 (1968): 44-45.
- [15] **BOWMAN, G.R., VOELZ, V.A., PANDE, V.S.**, "Atomistic Folding Simulations of the Five-Helix Bundle Protein λ_{6-85} ", *Journal of the American Chemical Society*, 133 (4), pp 664–667, 2011.
- [16] **WRIGHT, P. E., DYSON, H. J.**, Intrinsically unstructured proteins: reassessing the protein structure-function paradigm, *Journal of Molecular Biology*, vol. 293, no. 2, pp. 321–331, 1999.
- [17] **DUNKER, A.K., BROWN, C.J., LAWSON, J.D., IAKOUCHEVA, L.M., OBRADOVIC, Z.**, Intrinsic disorder and protein function. *Biochemistry* 2002, 41:6573–6582.
- [18] **CHENG, Y., LEGALL T., OLDFIELD, C.J., MUELLER, J.P., VAN, Y.Y., ROMERO, P., CORTESE, M.S., UVERSKY, V.N., DUNKER, A.K.**, Rational drug design via intrinsically disordered protein. *Trends Biotechnol*, 2006, 24:435–442.
- [19] **VUCETIC, S., OBRADOVIC, Z., VACIC, V., RADIVOJAC, P., PENG, K., IAKOUCHEVA, L.M., CORTESE, M.S., LAWSON, J.D., BROWN, C.J., SIKES, J.G.**, DisProt: a database of protein disorder. *Bioinformatics*, 21:137-140, 2005.

- [20] **GIBSON, T.J., RUSSELL, R.B., JENSEN, L.J., BORK, P., LINDING, R., DIELLA, F.**, Protein disorder prediction: implications for structural proteomics. *Structure*, 11, 1453–1459, 2003.
- [21] **YILMAZ, D., GÜLER, N.F.**, "Kaotik Zaman Serisinin Analizi Üzerine Bir Araştırma", *Gazi Üniversitesi Mühendislik Fakültesi Dergisi*, 2006.
- [22] **TIRNAKLI, U.**, "Kaos Geçiş Eşiğindeki Dinamik Sistemler", *ITAP-IARS Ders Notları*, 2007
- [23] **YAMAMOTO, Y.**, "Detection of Chaos and Fractals from Experimental Time Series", *Modern Techniques in Neuroscience Research*, Editör: Windhorst, U, ve H. Johansson, Springer-Verlag, Berlin Heidelberg New York Tokyo, 669-687, 1999.
- [24] **MANDELBROT, B.B.**, "The Fractal Geometry of Nature" W.H. Freeman, New York, 1982.
- [25] **YILMAZ, N.**, "Elektroensefalogram (EEG) Dalgalarında Lyapunov Üstelleri", *İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi*, 2010.
- [26] **GENÇ, F.**, "Bölgesel Depremlerin İncelenmesinde Zaman Serisi Yönetimi", *İstanbul Üniversitesi, Fen Bilimleri Fakültesi, Yüksek Lisans Tezi*, 2012.
- [27] **ECKMANN, J.P., RUELLE, D.**, "Ergodic Theory of Chaos and Strange Attractors", *Reviews of Modern Physics*, Cilt 57, No 3, 617-656, 1985.
- [28] **KOÇAK, K.**, "Kaotik Davranış Kriteri Olarak Fraktal Boyut Değişimi ve Dinamik Sistemlere Uygulanması", *İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi*, 1996.
- [29] **KİRAZ, S.** "Tisean Programı ile Lyapunov Üstelleri Tahmini", *Yalova Üniversitesi, Bilgisayar Mühendisliği Bölümü, Lisans Tezi*, 2014.
- [30] **SANO, M., SAWADA, Y.**, (1085), "Measurement Of The Lyapunov Spectrum From A Chaotic Time Series", *Phys. Rev. Lett.* 55, 1082.

- [31] **ÇEVİKALP, H.**, Feature Extraction Techniques in High-dimensional Spaces: Linear and non-linear Approaches, Ph.D., Nashville, Tennessee, 2005.
- [32] **KUNCHEVA, L.I.**, Combining pattern classifiers, John Wiley & Sons Inc., New Jersey, 2004.
- [33] **GÖK M., ÖZCERİT, A.T., OĞUL H.**, Predicting HIV-1 Protease Cleavage Site Using Support Vector Machines with Physicochemical Properties. INISTA 2010, Kayseri, Turkey, pp. 260-263, 2010.
- [34] **ALTMAN, E.**, "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," Journal of Finance, September 1968.
- [35] **RISH, I.** 2001. An empirical study of the naive Bayes classifier. IJCAI 2001 workshop on empirical methods in artificial intelligence. 3(22) IBM, New York, USA.
- [36] **MURPHY, K.** "The bayes net toolbox for matlab." Computing science and statistics 33.2 (2001): 1024-1034.
- [37] **RAMON, D., ANDRES, S.A.**, "Gene selection and classification of microarray data using random forest." BMC bioinformatics 7.1 (2006): 3.
- [38] **GOWDA, K. C., KRISHNA, G.** 1979. The condensed nearest neighbor rule using the concept of mutual nearest neighborhood. IEEE Transactions on Information Theory 25(4): 488-490.
- [39] **QUINLAN, J. R.** C4. 5: programs for machine learning. Elsevier, 2014.
- [40] **KEERTHI, S. S.**, "Improvements to Platt's SMO algorithm for SVM classifier design." Neural Computation 13.3 (2001): 637-649.
- [41] **RATSCH, G., ONODA, T., MULLER, K-R.**, "Soft margins for AdaBoost." Machine learning 42.3 (2001): 287-320.
- [42] **SANKAR, P.K., MITRA, S.**, "Multilayer perceptron, fuzzy sets, and classification." IEEE Transactions on Neural Networks 3.5 (1992): 683-697.

ÖZGEÇMİŞ

Ad Soyad : Sevdanur GENÇ

Doğum Yeri ve Tarihi : İstanbul - 03.04.1983

E-Posta : sevdanurgenc@yahoo.com

Web : www.sevdanurgenc.com



Yüksek Lisans : Yalova Üniversitesi - Bilgisayar Mühendisliği

Lisans : Süleyman Demirel Üniversitesi - Bilgisayar Mühendisliği

Ön Lisans : Süleyman Demirel Üniv. Uluborlu MYO - Bilgisayar Programcılığı

Yayın Listesi:

- Küçüksille, E.U., Genç, S.N., Özger, F., 2012 "Mobil Bulut Bilişim ve Geleceği", "Mobile Cloud Computing And The Future", Akademik Bilişim 2013, Aralık 2012
- Küçüksille, E.U., Genç, S.N., Karabulut, Y.E, 2013 "Dünya da Siber Güvenlik Stratejileri ve Bir Siber Güvenlik Stratejisinin Oluşumu.", ISDFS 2013, Mayıs 2013

TEZDEN TÜRETİLEN YAYINLAR/SUNUMLAR

- Genç, S.N., Gök, M., Koçal, O. H., 2015 "The Chaotic Structure of Bacterial Virulence Protein Sequences ", Computer Science & Information Technology, pp. 147–155
- Genç, S.N., Gök, M., Koçal, O. H., 2015 "Prediction Of Disordered Protein Regions Based On Chaos Theory And Physicochemical Properties Of Amino Acids"
- Genç, S.N., Gök, M., Durgut, A., 2015 "Proteinlerde Düzensiz Bölgelerin Makine Öğrenmesi Yöntemleri İle Tespiti"