

# Dna Dizilimlerinin Sayısallaştırılması İçin Entropi Tabanlı Yeni Bir Sayısal Haritalama Tekniği

Bihter Daş

İbrahim Türkoğlu

Yazılım Mühendisliği Fırat Üniversitesi

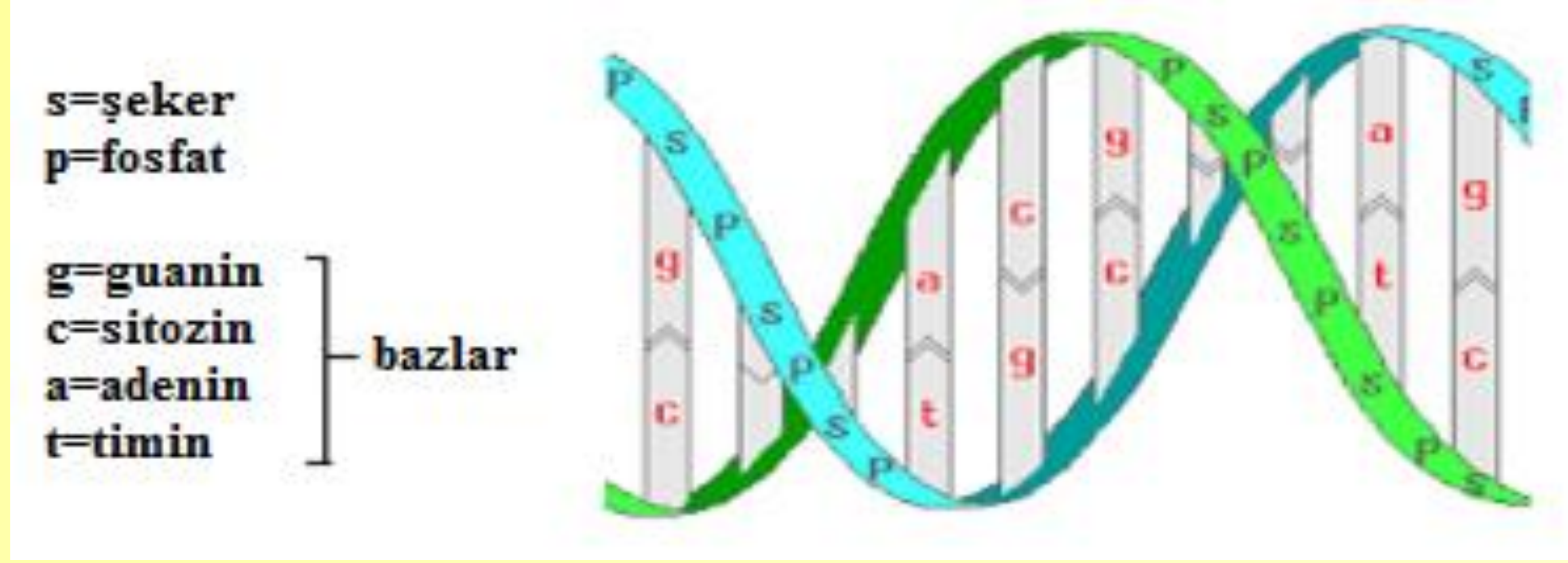
Elazığ, Türkiye

bihter@ualberta.ca & bdas@firat.edu.tr iturkoglu@firat.edu.tr

Bu çalışmanın amacı, DNA dizilimlerini sayısal sinyal işleme uygulamalarında kullanabilmek için yeni bir sayısal haritalama tekniği önermektir.

- Her bir kodon Shannon denkleminin geliştirilmiş kesirli bir türevi olan fraksiyonel entropi denklemi ile haritalanmaktadır.
- DNA dizilimdeki ekzon bölgelerini tespiti için Tekli Değer Ayrışımı, Ayrık Fourier Dönüşümü yöntemi kullanılmaktadır.
- Önerilen haritalama tekniğinin performansı varolan diğer sayısal haritalama teknikleriyle karşılaştırılmaktadır.
- Her ekzon pozisyonunun tespiti için yöntemin doğruluk oranı ve ortalama başarımları yaklaşık olarak %96 dir.

## Giriş



Şekil 1. Bir DNA'nın yapısı

- DNA, genetik talimatları taşıyan bir nükleik asittir.
- DNA dizilimleri, DNA bölgesindeki A,G,C,T nükleotid bazlarının sırasının belirlenmesi işlemidir.
- DNA dizilimlerinde protein kodlayan bölgelere ekzon, protein kodlamasına katılmayan bölümlere intron denir.

>gi|41223386|emb|AJ229040.1| Homo sapiens 959 kb contig between AML1 and CBR1 on chromosome 21q22GATCACTGAGCAT,TTTGCAGGGTTGTTCTCTGAGGGCTGTGAGGAGAATCTGTTCTGGGCCTCTGACTTGGCAGATGGTCATCTCTCTCTCTCTGTCCTCCCTCTCTCTCTCTGTACCGCTCAAGTCTGTGCAAGACACCACCTGTGGAGCTTCTCACTCAGCTCTGGGCTTCGGCTCCATGAGCTCTAGTTAATCTCCATGTGGCTCCAGCAATCTGTGCAACATAAAGCACAATATGAATTTTATAGTAAAAATAAAGATTGAAAGCTTCTCAGTGCCGGGCACTGTGATAAGGACTTTAAACCGGTACTATAATGATGTAACTGGGATTGTCTGGCCAAATGGGTGATGGCCACCCCACTACAGTCTACCAGGATGACAGAGCACTGGGTGCAGACCAGGAGAGTGGGAGGAGTGGCAGAGGCTTGCTCATGCTGTGTGTGTGTGTGCACACCAATGTGTATACAGCATGCATGGATCCAATAGGTCTGACCCCGGCCCATCTCAGTGGCTATCTGGTTGGTGGCTTTTAAATTTTTAGTGGTGGTGGGAGTAAAGGATTAACACACACACACACACACACACACACAGATTCCTTGGATGTTAGGTGAGAAAAGTCC'ACTGGCAATACGGGAAATTTTATAGTCAGCACATTTTAAATTAGCCAGACACGAACCTTAGCCGTTCCGCTGAATACATCAGCACAGAGTGCAGAACCTTGCAAGAGGGTCTCTGAAATTAGGATACTATCAGAATCTCAGGAGTATTTCTCATGTAGCATCTTTTAGATAAGTCCATGTGAGAGGTGAAAACTCAGGTTGTTCTGTTTTCGCAAGTTCTGTGGCATATGAAGCATTAATATTTCTCTCTTTCTTACTTACTCTCCAGCTACCTTTCTCTTTAAGGCCAGGCCGATGTTACACGTTGTCTGTGGCTCTTTCTCAAGAGGACTGGTTAGACATTTTGTAGAATGTTCCCATTTGCAATGTGTCTGTATGTTTCTGTGATTAGCAGGCTTTGGGTTTCTGGGAAGAATACCATGAAGGCCCTTATCTAACAGTCATATTGTGGTTAGGGCTCAACATATAAATTTTGCAGGGACACAATTCAATCCCTAACAGAGCTTTTTTTTTTTTTTTTCTTGTGAGAT

Şekil 2. NCBI veritabanından alınan AJ229040 genine ait bir DNA dizilim örneği

## Önerilen Entropi Tabanlı Sayısal Haritalama Tekniği

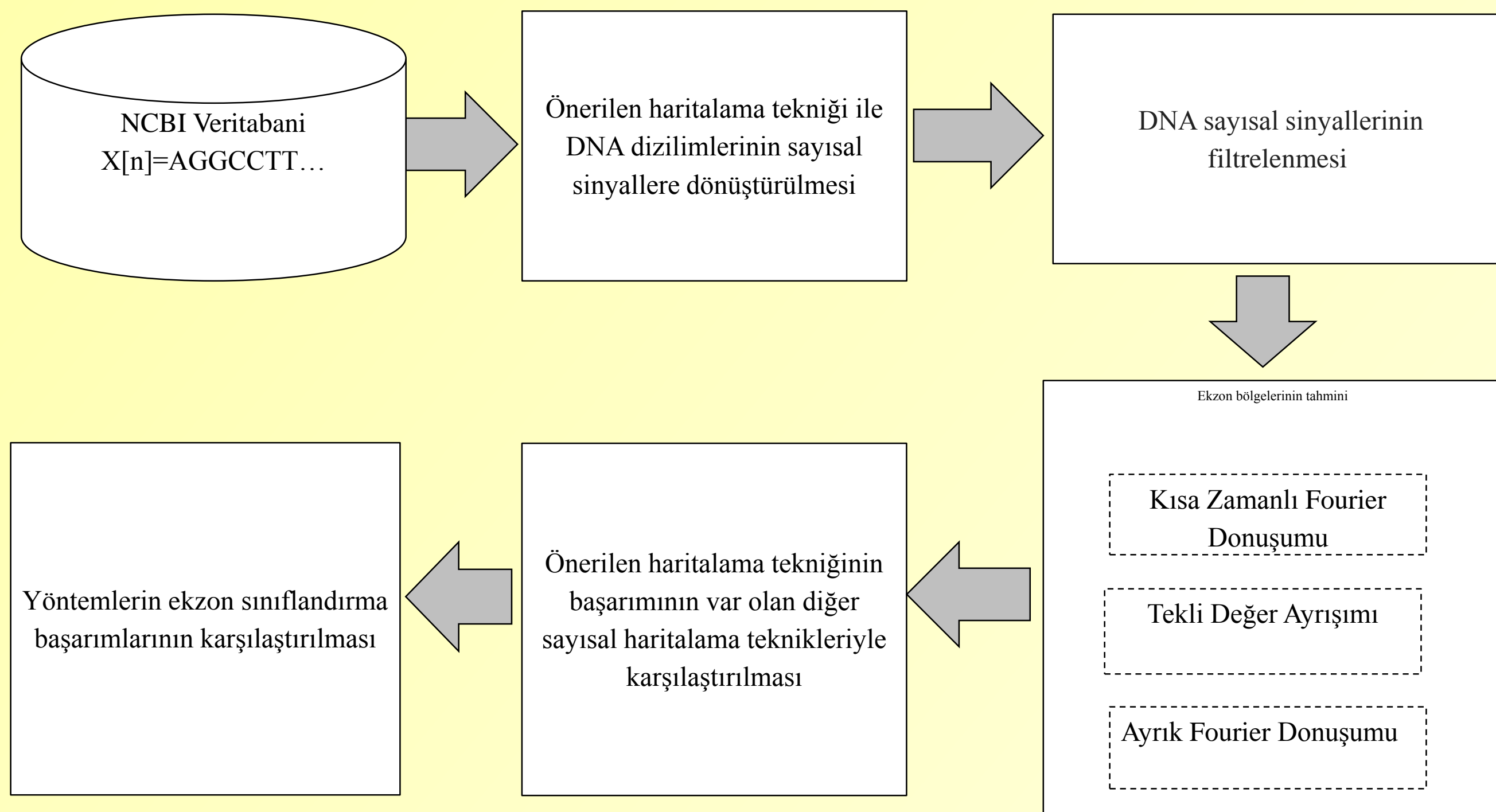
Dizilimdeki analog değerlerini sayısal değerlere dönüştürmek için fraksiyonel Shannon Entropi yöntemi kullanılmıştır.

$$Sf = -\sum_i [(-p(x_i))^{\alpha} p(x_i) \log(p(x_i))]$$

$p(x_i)$  değeri bir DNA dizilimindeki her bir kodonun tekrarlaması sıklığıdır.

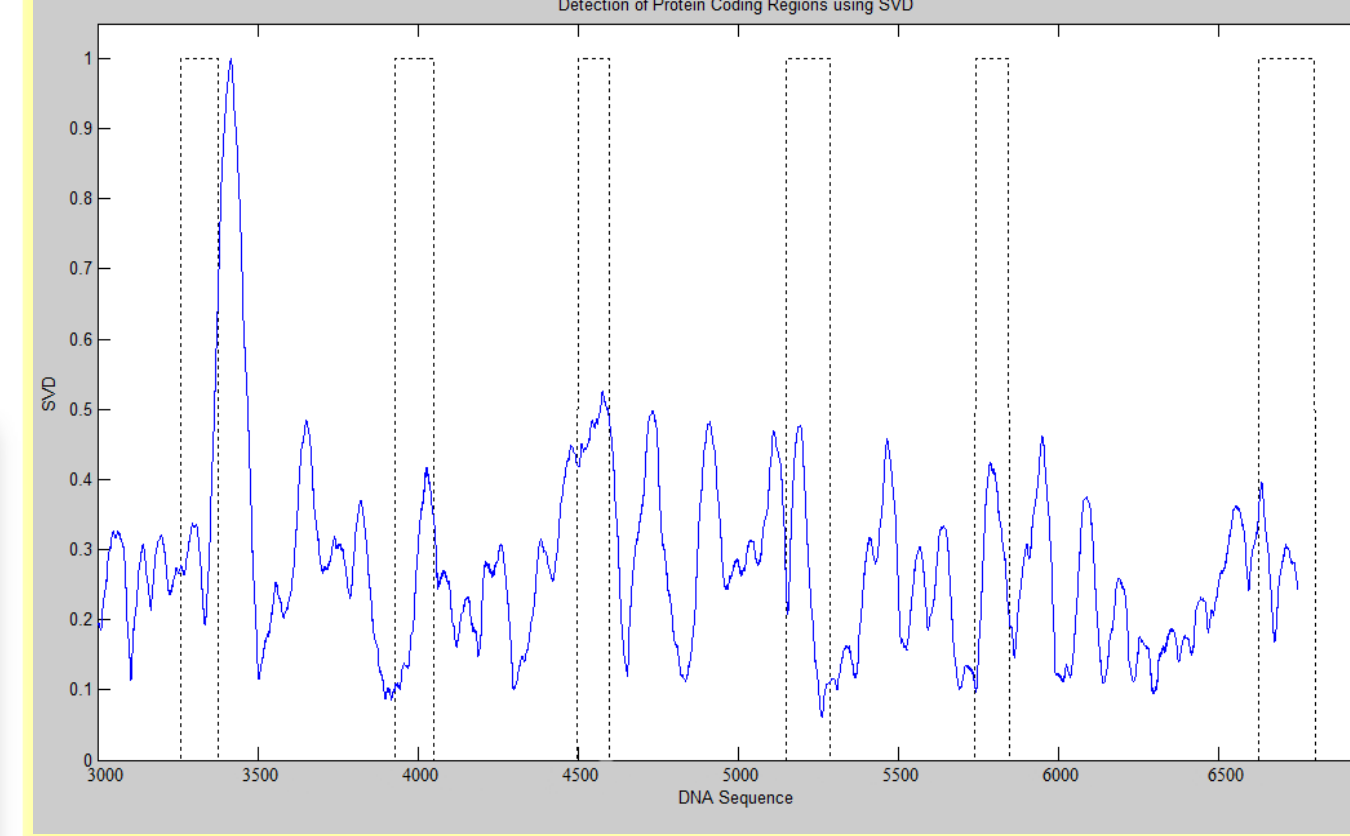
$$\alpha = \frac{1}{\log(p(x_i))}$$

## Ekzon Bölgelerinin Tespitinde Haritalama Tekniğinin Performansı

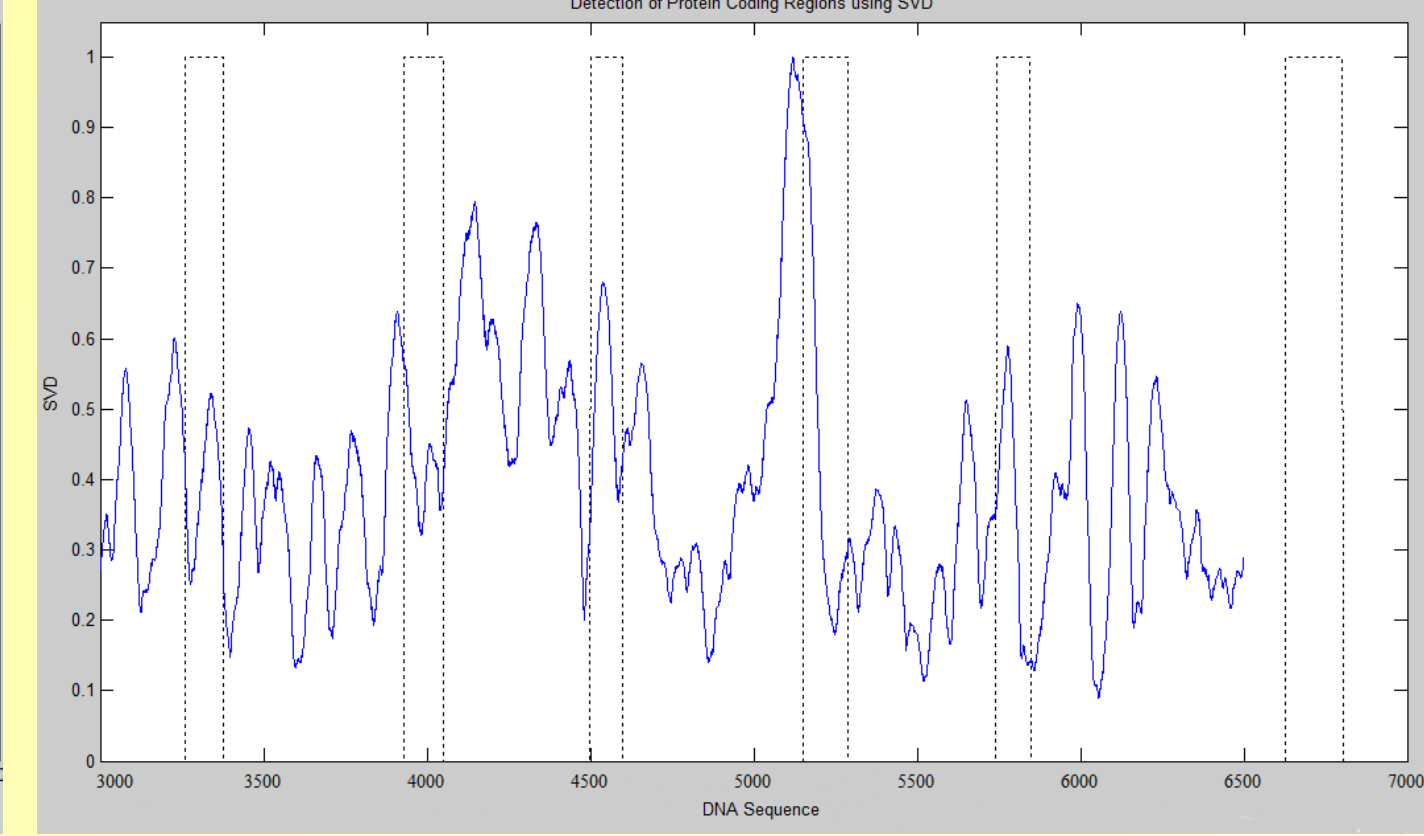


Şekil 3. AFD, KZFD ve TDA yöntemlerinde önerilen sayısal haritalama tekniği ile ekzon bölgelerinin belirlenmesine yönelik uygulamanın blok diyagramı

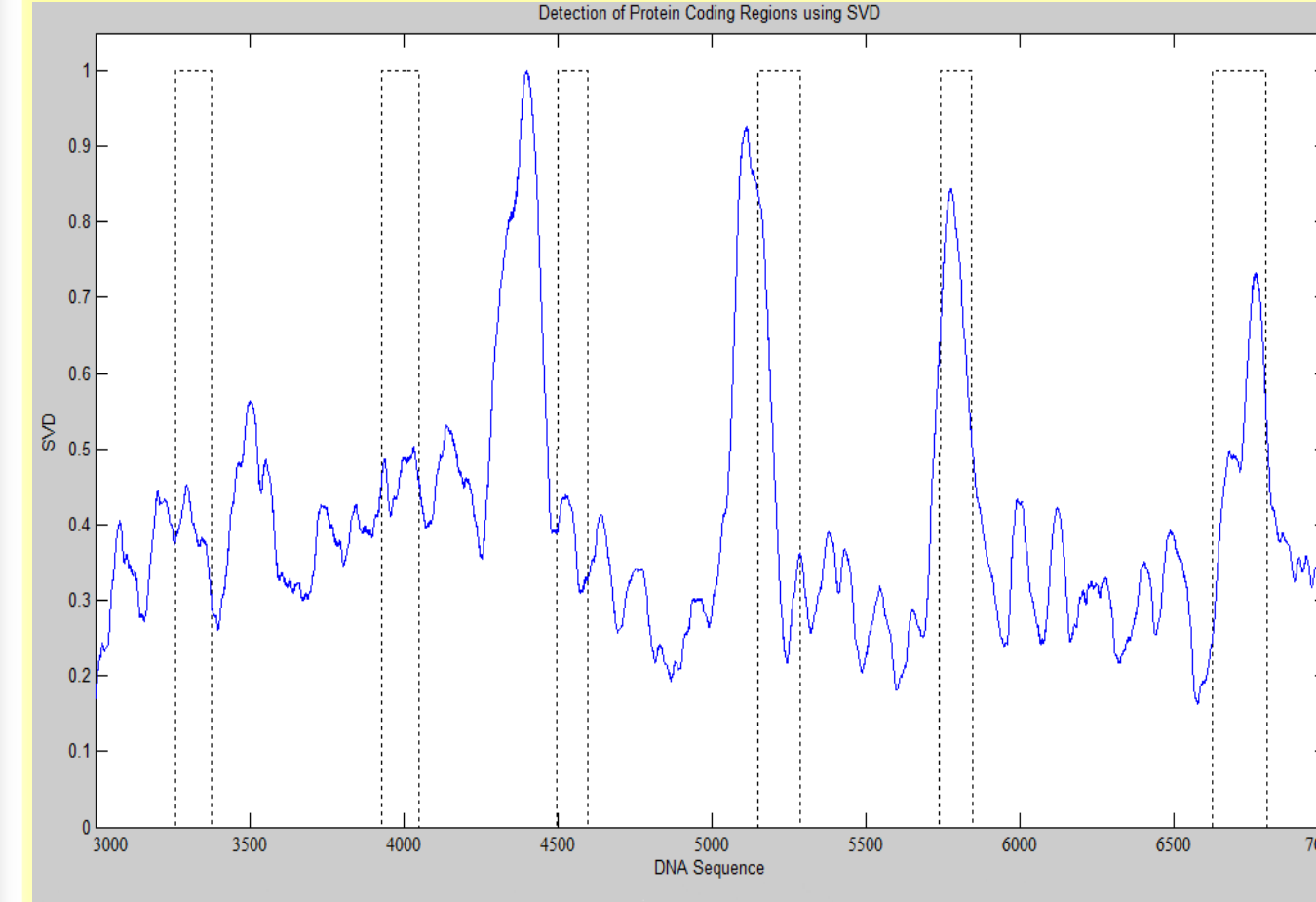
## Deneyisel Sonuçlar



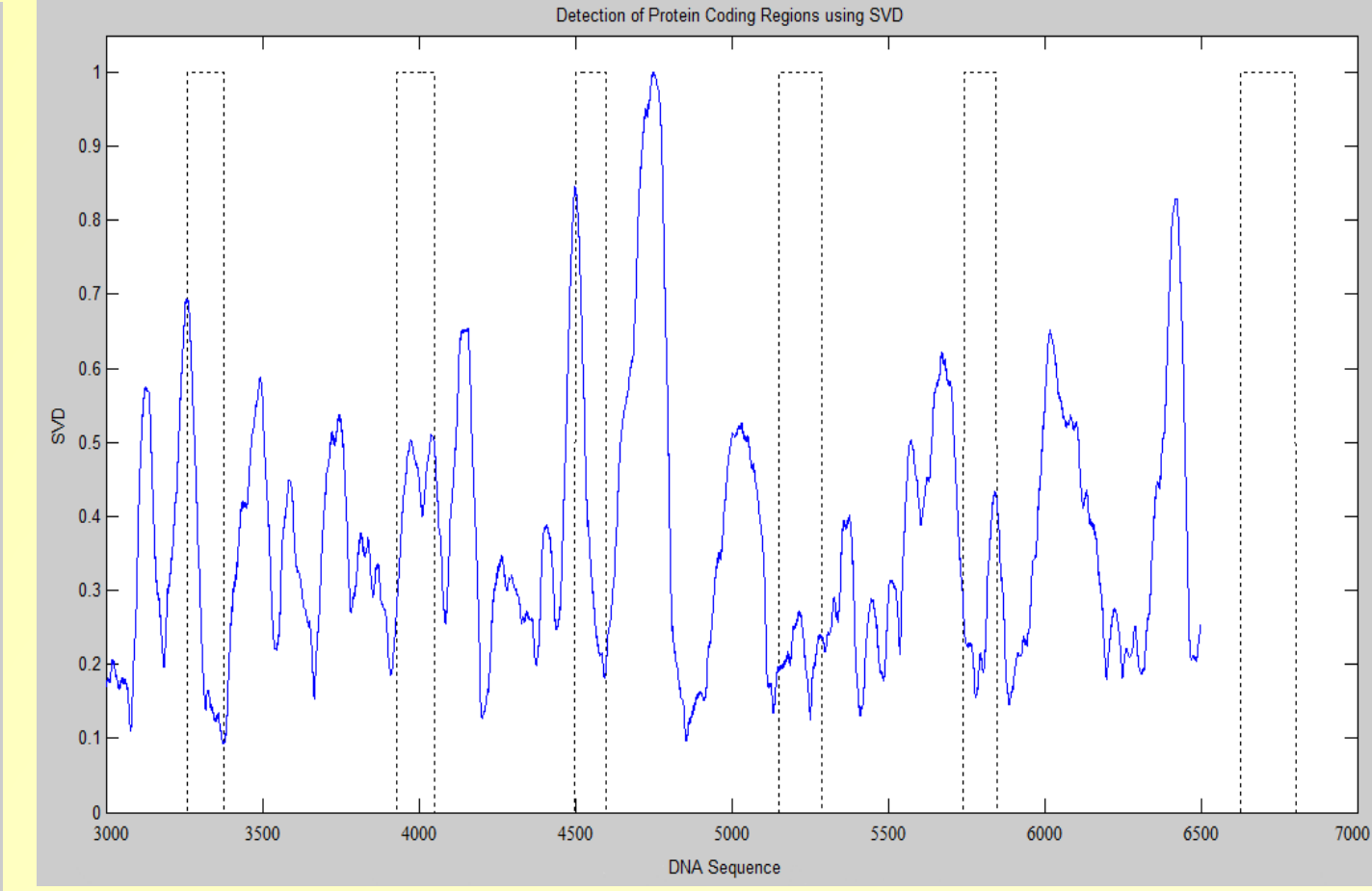
a. Complex haritalama tekniği



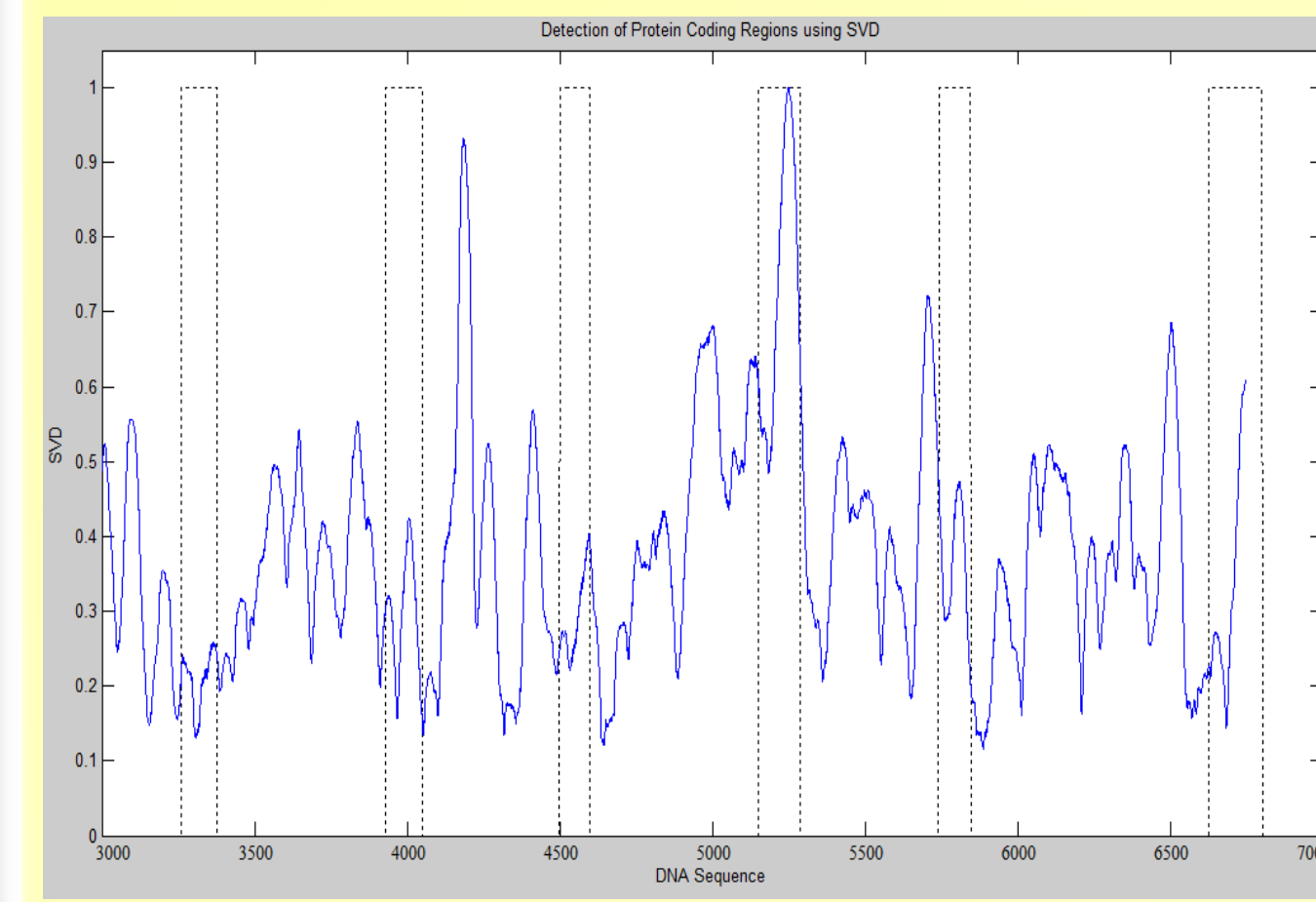
b. EIIP haritalama tekniği



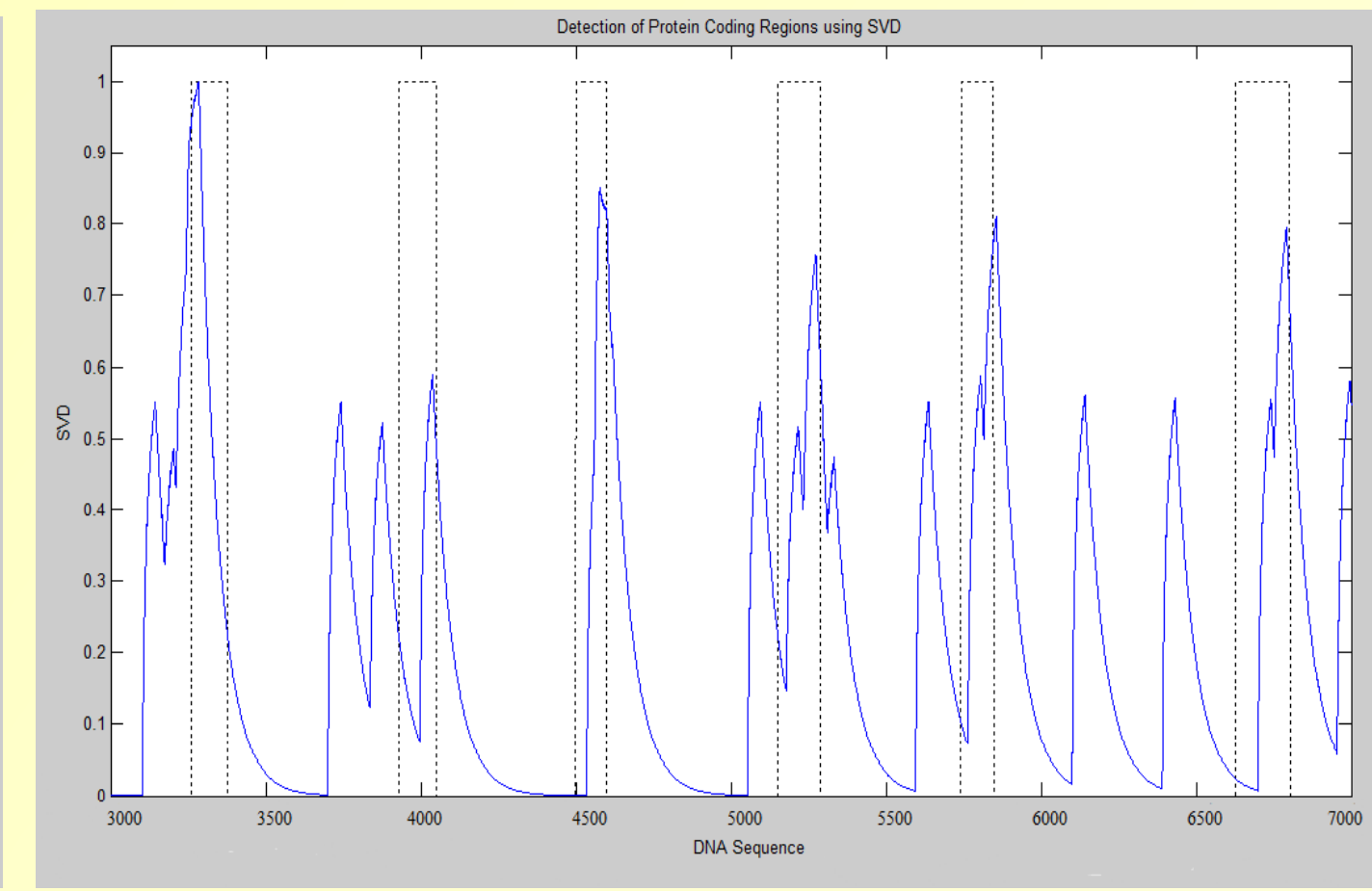
c. Integer haritalama tekniği



d. Real haritalama tekniği

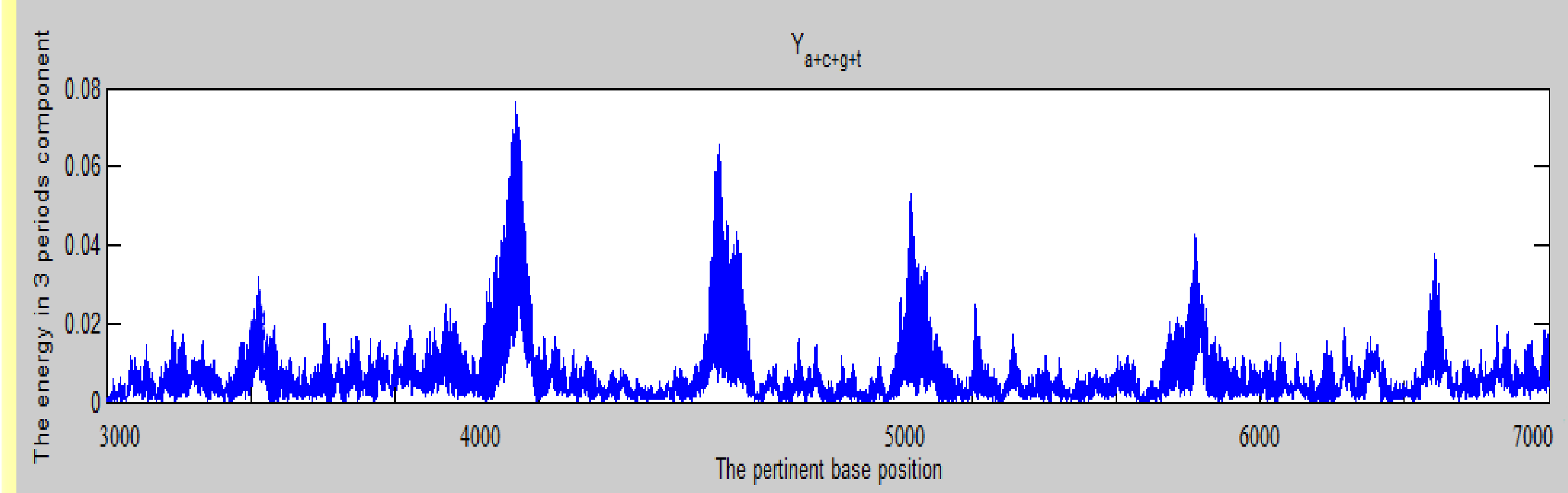


e. Atomic haritalama tekniği



e. Entropi tabanlı haritalama tekniği

Şekil 4. Ekzon bölgelerinin taranmasına yönelik Tekli Değer Ayrışımı yönteminde farklı sayısal haritalama tekniklerinin performanslarının karşılaştırılması



Şekil 5. Entropi tabanlı haritalama tekniğinin Ayrık Fourier Dönüşümü yöntemindeki ekzon bölgelerinin taranmasındaki başarımı

Tablo 1. Esik değere göre sayısal haritalama tekniklerinin doğruluk oranlarının karşılaştırılması

Haritalama Tekniği	Esik değeri	Protein Kod Bölgeleri Doğruluk Oranı (%)
Integer Tekniği	0.0463	86,125
EIIP Tekniği	0.000142	77,53
DNA Walk Tekniği	0.0305	90,82
Paired Numeric Tekniği	0.0290	87,79
Entropi Tabanlı Haritalama Tekniği	0.0275	96,114

## Sonuç

- DNA dizilimindeki protein kodlayan bölgelerin (ekzon) tespiti için yeni bir sayısal haritalama yaklaşımı önerilmiştir.
- Dizilimdeki her bir kodon Shannon denkleminin geliştirilmiş fraksiyonel türevi ile sayısallaştırılır.
- Önerilen sayısal haritalama teknik 6 ekzon bölgesini bulmada büyük bir başarı göstermiştir.
- Önerilen haritalama tekniğinin başarımı TDA, AFD yöntemlerinde test edilmiştir.
- Önerilen teknik %96.116 başarımla göstermiştir.