

1. PROTEİN DİZİLİMLERİNİN SAYISAL İFADELERE DÖNÜŞTÜRÜLMESİ

Protein dizilimlerinin çeşitli biyoenformatik uygulamaları ve yapay zekâ uygulamalarında kullanılabilmesi için, bu dizilimlerin sayısal ifadelere dönüştürülmesi gerekmektedir. Bu doğrultuda çeşitli kategorilere ait protein haritalama teknikleri bulunmaktadır. Karakter-tabanlı, fizikokimyasal-tabanlı, evrim-tabanlı, yapı-tabanlı, makine öğrenmesi tabanlı ve algoritma tabanlı olmak üzere altı farklı kategori bulunmaktadır.

1.1. Fizikokimyasal-Tabanlı Protein Haritalama Teknikleri

Bu yöntemde protein dizilimleri, proteinlerin kimyasal bilgilerine dayanarak sayısal ifadelere dönüştürülmektedir. Bu doğrultuda kullanılan yöntemler hidrofobiklik, Meiler parametreleri, Atchley faktörleri ve EIIP yöntemleridir.

1.1.1. Hidrofobiklik Protein Haritalama Tekniği

Bir proteinin biyolojik fonksiyonu, protein diziliminin kendisi ve protein diziliminin kimyasal özellikleri tarafından belirlenmektedir. Yirmi adet amino asidin yüzlerce fizikokimyasal özellikleri bulunmaktadır. Bunların en önemlisi polarize edilebilirlik, Van der Waals hacmi, iyonizasyon sabiti, erişilebilir solvent yüzey alanı ve hidrofobiklik. Bu yöntemde yirmi adet amino asit yan zincirinin her birinin hidrofilik ve hidrofobik özellikleri ele alınmaktadır. Amino asitlerin hidrofobiklik özellikleri, amino asit dizilimlerini karşılaştırmak ve analiz etmek için etkili bir yol olarak sıklıkla kullanılmaktadır. Hidrofobiklik protein sayısallaştırma tekniği bir boyutlu sayısal verilerden oluşmaktadır. Tablo 1’de amino asitlerin hidrofobiklik değerleri verilmiştir.

Tablo 1. Amino asitlerin hidrofobiklik değerleri

Amino asit kodu	Hidrofobiklik değeri	Amino asit kodu	Hidrofobiklik değeri
A	1.8	M	1.9
C	2.5	N	-3.5
D	-3.5	P	-1.6
E	-3.5	Q	-3.5
F	2.8	R	-4.5
G	-0.4	S	-0.8
H	-3.2	T	-0.7
I	4.5	V	4.2
K	-3.9	W	-0.9
L	3.8	Y	-1.3

Tablo 1’de verilen değerler doğrultusunda $P(N) = [MWFYPC \dots]$ biçimindeki bir protein dizilimi, hidrofobiklik yöntemi ile $C(N) = [1.9 - 0.9 \ 2.8 - 1.3 - 1.6 \ 2.5 \dots]$ şeklinde sayısal hale getirilir.

1.1.2. Atchley Faktörleri

Atchley faktörleri ise, protein dizilimlerinde sıklıkla görülen metrik problemlerini gidermek için önerilmiş bir yöntemdir. Protein dizilimlerinde sıklıkla bulunan varyans ve kovaryansa değişkenleri, son yıllarda gelişen teknolojiler ile amino asit kodlarının alfabetik sıralanmasına ve bilgi teorisine dayanarak sayısallaştırılmaktadır. Bu yöntemde toplamda beş faktör bulunmaktadır. Bu faktörler ikincil yapı (iy), moleküler hacim (mh), polarite (p), protein elektrostatik yük (ey) ve kodon yoğunluğu (ky) faktörleridir. Beş adet faktör bulunduğu için beş boyutlu bir yöntemdir. Tablo 2’de amino asitlerin Atchley faktör değerleri verilmiştir.

Tablo 2. Amino asitlerin Atchley faktör değerleri

Amino asit kodu	p	iy	mh	p	ey
A	-0.591	-1.302	-0.733	1.570	-0.146
R	1.538	-0.055	1.502	0.440	2.897
N	0.945	0.828	1.299	-0.169	0.933
D	1.050	0.302	-3.656	-0.259	-3.242
C	-1.343	0.465	-0.862	-1.020	0.255
E	1.357	-1.453	1.477	0.113	-0.837
Q	0.931	-0.179	-3.005	-0.503	-1.853
G	-0.384	1.652	1.330	1.045	2.064
H	0.336	-0.417	-1.673	-1.474	-0.078
I	-1.239	-0.547	2.131	0.393	0.816
L	-1.019	-0.987	-1.505	1.266	-0.912
K	1.831	-0.561	0.533	-0.277	1.648
M	-0.663	-1.524	2.219	-1.005	1.212
F	-1.006	-0.590	1.891	-0.397	0.412
P	0.189	2.081	-1.628	0.421	-1.392
S	-0.228	1.399	-4.760	0.670	-2.647
T	-0.032	0.326	2.213	0.908	1.313
W	-0.595	0.009	0.672	-2.128	-0.184
Y	0.260	0.830	3.097	-0.838	1.512
V	-1.337	-0.279	-0.544	1.242	-1.262

Tablo 2’de verilen değerler doğrultusunda $P(N) = [ARN \dots]$ biçimindeki bir protein dizilimi, Atchley faktörleri ile $C(N) = [[-0.591 - 1.302 - 0.733 \ 1.570 - 0.146][1.538 - 0.055 \ 1.502 \ 0.440 \ 2.897][0.945 \ 0.828 \ 1.299 - 0.169 \ 0.933]]$ şeklinde sayısal hale getirilir.

1.1.3. EHP Protein Haritalama Tekniği

EIIP protein sayısallaştırma yöntemi ilk olarak DNA ve proteinler arasındaki etkileşimleri belirlemek için kullanılmıştır. EIIP yönteminin ilk aşamasında protein dizilimleri Fourier dönüşümü ile sinyallere ayrıştırılmıştır. Ayrıştırılma işleminin ardından sinyallerden güç spektrum yoğunluğu değerleri hesaplanmıştır. Son olarak her bir güç spektrum değeri, ilgili amino asitlere tanımlanmış ve sayısallaştırma işlemi gerçekleştirilmiştir. Tablo 3'te amino asitlerin EIIP değerleri verilmiştir.

Tablo 3. Amino asitlerin EIIP değerleri

Amino asit kodu	EIIP değeri	Amino asit kodu	EIIP değeri
M	0.0823	Q	0.0761
W	0.0548	S	0.0829
F	0.0946	A	0.0373
Y	0.0516	N	0.0036
P	0.0198	G	0.0050
C	0.0829	R	0.0959
T	0.0941	I	0
H	0.0242	D	0.1263
V	0.0057	E	0.0058
L	0	K	0.0371

Tablo 3'te verilen değerler doğrultusunda $P(N) = [MWFYPC \dots]$ biçimindeki bir protein dizilimi, EIIP yöntemi ile $C(N) = [0.0823 \ 0.0548 \ 0.0946 \ 0.0516 \ 0.0198 \ 0.0829 \dots]$ şeklinde sayısal hale getirilir.

1.2. Evrim-Tabanlı Protein Haritalama Teknikleri

Evrimsel protein sayısallaştırma yöntemlerinde ise, protein dizilimleri, proteinlerin evrimsel bilgilerine dayanarak sayısallaştırılmaktadır. Proteinlerin evrimsel bilgileri, filogenetik ağaçlardan ya da dizilim hizalama işlemlerinden elde edilmektedir. Evrim-tabanlı protein sayısallaştırma kategorisinde, PAM250 (Point Accepted Mutation – Nokta Kabul Edilen Mutasyon) ve BLOSUM62) olmak üzere iki farklı sayısallaştırma yöntemi değerlendirilmiştir.

1.2.1. BLOSUM Matrisi

BLOSUM matrisi, biyoenformatik alanında sıklıkla kullanılan ve protein dizilimlerini hizalamak için başvurulan yöntemlerin başında gelmektedir. BLOSUM matrisleri, evrimsel olarak birbirinden farklı protein dizilimleri arasındaki hizalamaları puanlamak için kullanılmaktadır. BLOSUM80, BLOSUM45 gibi farklı türde matrisleri bulunmaktadır. Matrislerin yanındaki sayılar benzerlik değerini ifade etmektedir. Evrimsel açıdan birbirine en yakın iki organizma arasındaki benzerlikler yüksek skor değerleri almaktadır. Bundan dolayı, BLOSUM matrislerinde skor ne kadar yüksek ise, benzerlik de o kadar fazla olmaktadır. Örneğin, birbirine tür olarak yakın olan iki organizmanın protein dizilimleri hizalanırken BLOSUM80 matrisi kullanılmaktadır. Buna zıt olarak ise, birbirine tür olarak uzak olan iki organizmanın proteinleri ise BLOSUM45 matrisi ile hizalanabilmektedir. Ancak, organizmaların uzaklık ya da yakınlık durumları tam olarak bilinmediğinde ise BLOSUM62 matrisine başvurulmaktadır. Tablo 4'te amino asitlerin BLOSUM62 değerleri verilmiştir.

Tablo 4. Amino asitlerin BLOSUM62 değerleri

Amino asit kodu	BLOSUM62 değeri	Amino asit kodu	BLOSUM62 değeri
A	4	L	4
R	5	K	5
N	6	M	5
D	6	F	6
C	9	P	7
Q	5	S	4
E	5	T	5
G	6	W	11
H	8	Y	7
I	4	V	4

Tablo 4'te verilen değerler doğrultusunda $P(N) = [ARNDCQ \dots]$ biçimindeki bir protein dizilimi, BLOSUM62 yöntemi ile $C(N) = [4\ 5\ 6\ 6\ 9\ 5 \dots]$ şeklinde sayısal hale getirilir.

1.2.2. PAM250 Matrisi

PAM250 matrisi, hizalanan peptit dizilimlerini puanlayarak protein dizilimlerindeki benzerlikleri belirlemek için kullanılan bir yöntemdir. PAM250 matrisindeki değerler, bilinen homolojiye sahip hizalanmış protein dizilimlerinin kıyaslanmasından ve gözlemlenen nokta mutasyonlarının belirlenmesinden elde edilmiştir. Tablo 5'te amino asitlerin PAM250 değerleri verilmiştir.

Tablo 5. Amino asitlerin PAM250 değerleri

Amino asit kodu	PAM250 değeri	Amino asit kodu	PAM250 değeri
A	2	L	6
R	6	K	5
N	2	M	6
D	4	F	9
C	4	P	6
Q	4	S	3
E	4	T	3
G	5	W	17
H	6	Y	10
I	5	V	4

Tablo 5'te verilen değerler doğrultusunda $P(N) = [ARNDCQ \dots]$ biçimindeki bir protein dizilimi, PAM250 yöntemi ile $C(N) = [2 \ 6 \ 2 \ 4 \ 4 \ 4 \dots]$ şeklinde sayısal hale getirilir.

1.3. Yapı-Tabanlı Protein Haritalama Teknikleri

Yapı-tabanlı protein sayısallaştırma yöntemlerinde, proteinler, proteinlerin yapı bilgisine dayanarak sayısallaştırılır. Bu bilgiler, proteinlerin birincil, ikincil, üçüncül ve dördüncül yapılarından gelmektedir. Bu kategoride çeşitli sayıda yöntem bulunsa da diğer kategorilere nazaran daha az kullanılmaktadır. Bunun nedenlerinden birisi protein yapı bilgilerinin elde edilmesindeki zorluktur. Bu kategoride Miyazawa enerjileri ve Micheletti potansiyelleri en çok kullanılan yöntemlerdendir.

1.3.1. Miyazawa Enerjileri Protein Haritalama Tekniği

Miyazawa enerjileri, protein dizilimlerindeki kalıntılar arasındaki enerjileri belirlemek için önerilen bir yöntemdir. Bu yöntemde protein dizilimlerinin temas enerjileri regresyon katsayıları ile elde edilir. Miyazawa enerjilerinde enerji değerleri, ikincil yapı enerjileri ve üçüncül yapı enerjileri olmak üzere iki terime ayrılmıştır. Üçüncül yapı enerjileri, proteinlerin kalıntı-kalıntı arasındaki temas enerjilerinin toplanmasıyla elde edilir. İkincil yapı enerjileri ise atom zincirleri arasındaki etkileşimlere ve atom zincir-yan zincir arasındaki etkileşimlerine göre hesaplanır. Tablo 6'da proteinlerin Miyazawa enerji değerleri verilmiştir.

Tablo 6. Amino asitlerin Miyazawa enerji değerleri

Amino asit kodu	Miyazawa enerji değeri	Amino asit kodu	Miyazawa enerji değeri
A	-0.02	L	-0.32
R	0.08	K	0.30
N	0.10	M	-0.25
D	0.19	F	-0.33
C	-0.32	P	0.11
Q	0.21	S	0.11
E	0.15	T	0.05
G	-0.02	W	-0.27
H	-0.02	Y	-0.23
I	-0.28	V	-0.23

Tablo 6’da verilen değerler doğrultusunda $P(N) = [ARNDCQ \dots]$ biçimindeki bir protein dizilimi, Miyazawa enerjileri ile $C(N) = [-0.02 \ 0.08 \ 0.10 \ 0.19 \ -0.32 \ 0.21 \dots]$ şeklinde sayısal hale getirilir.

1.3.2. Micheletti Potansiyelleri Protein Haritalama Tekniği

Diğer bir yapı-tabanlı protein sayısallaştırma yöntemi olan Micheletti potansiyelleri, proteinler arasında ortaya çıkan etkileşimlerdeki potansiyel enerjiye dayanmaktadır. Micheletti potansiyellerinin ana fikri, en uygun protein etkileşimlerini belirlemektir. Tablo 7’de proteinlerin Micheletti potansiyel değerleri verilmiştir.

Tablo 7. Amino asitlerin Micheletti potansiyel değerleri

Amino asit kodu	Micheletti potansiyel değeri	Amino asit kodu	Micheletti potansiyel değeri
A	-0.001461	L	-0.000782
R	0.009875	K	0.005109
N	-0.001962	M	0.031655
D	-0.000531	F	-0.013128
C	-0.002544	P	-0.003621
Q	0.006456	S	-0.000802
E	0.008438	T	0.003269
G	0.000990	W	0.131813
H	0.001314	Y	-0.007699
I	0.006801	V	0.001445

Tablo 7’de verilen değerler doğrultusunda $P(N) = [ARNDCQ \dots]$ biçimindeki bir protein dizilimi, Micheletti potansiyelleri ile $C(N) = [-0.001461 \ 0.009875 \ -0.001962 \ -0.000531 \ -0.002544 \ 0.006456 \dots]$ şeklinde sayısal hale getirilir.

1.4. Karakter-Tabanlı Protein Haritalama Teknikleri

Karakter-tabanlı protein sayısallaştırma yöntemlerinde, protein dizilimleri belirli bir bilgiye ihtiyaç duyulmadan sayısallaştırılmaktadır. Diğer kategoriye ait protein sayısallaştırma yöntemlerinde, proteinler yapı bilgilerine, evrimsel bilgilere ya da kimyasal bilgilere dayanarak sayısallaştırılırken karakter-tabanlı yöntemlerde böyle bir yaklaşım bulunmamaktadır. Bu kategoride CPNR ve ikili-kodlama olmak üzere iki adet protein sayısallaştırma yöntemi üzerinde durulmuştur.

1.4.1. CPNR Protein Haritalama Tekniği

CPNR protein sayısallaştırma yöntemi ilk kez protein fonksiyonlarını kıyaslamak için önerilmiş ve kullanılmıştır. Önerilen protein sayısallaştırma yönteminde, protein dizilimleri kodonlara göre sayısallaştırılmaktadır. Sayısallaştırma işleminin ilk aşamasında amino asitler kodon sayılarına ayrıştırılmıştır. Ardından asal sayılar kodon sayılarına göre amino asit dizilimlerine atanmıştır. Tablo 8’de amino asitlerin CPNR değerleri verilmiştir.

Tablo 8. Amino asitlerin CPNR değerleri

Amino asit kodu	CPNR değeri	Amino asit kodu	CPNR değeri
M	1	Q	29
W	2	S	31
F	3	A	37
Y	5	N	41
P	7	G	43
C	11	R	47
T	13	I	53
H	17	D	59
V	19	E	61
L	23	K	67

Tablo 8’de verilen değerler doğrultusunda $P(N) = [MWFYPC \dots]$ biçimindeki bir protein dizilimi, CPNR protein haritalama tekniği ile $C(N) = [1\ 2\ 3\ 5\ 7\ 11 \dots]$ şeklinde sayısal hale getirilir.

1.4.2. Bir-Sıcak Protein Haritalama Tekniği

Bir-sıcak protein haritalama tekniğinde, protein dizilimlerindeki amino asitler “0” ve “1” değerleri ile temsil edilmektedir. En yaygın olarak kullanılan ikili kodlama yöntemidir. Bu yöntemde, yirmi adet standart amino asidin her biri yirmi boyutlu ikili vektör ile sayısallaştırılmaktadır. Sayısallaştırma işleminin ilk aşamasında amino asitler alfabetik bir şekilde sıralanır. Ardından *i*. sıradaki amino asit değeri “1” değerini alırken, diğerleri “0” değerini almaktadır. Tablo 9’da amino asitlerin bir-sıcak değerleri verilmiştir.

Tablo 9. Amino asitlerin bir-sıcak değerleri

Amino asit kodu	Bir-sıcak değeri	Amino asit kodu	Bir-sıcak değeri
A	10000000000000000000	M	00000000001000000000
C	01000000000000000000	N	00000000000100000000
D	00100000000000000000	P	00000000000010000000
E	00010000000000000000	Q	00000000000001000000
F	00001000000000000000	R	00000000000000100000
G	00000100000000000000	S	00000000000000010000
H	00000010000000000000	T	00000000000000001000
I	00000001000000000000	V	00000000000000000100
K	00000000100000000000	W	00000000000000000010
L	00000000010000000000	Y	00000000000000000001

Tablo 9’da verilen değerler doğrultusunda $P(N) = [AC \dots]$ biçimindeki bir protein dizilimi, bir-sıcak protein haritalama tekniği ile $C(N) = [10000000000000000000 01000000000000000000 \dots]$ şeklinde sayısal hale getirilir.

1.4.3. Tam Sayı Protein Haritalama Tekniği

Bu yöntemde ise amino asit kodları alfabetik olarak sıralanır ve sıralanmasının ardından 1'den 20'ye kadar olan tam sayıya bu kodlara atanır. Tablo 10'da amino asitlerin tam sayı değerleri verilmiştir.

Tablo 10. Amino asitlerin tam sayı değerleri

Amino asit kodu	Tam sayı değeri	Amino asit kodu	Tam sayı değeri
A	1	L	11
R	2	K	12
N	3	M	13
D	4	F	14
C	5	P	15
Q	6	S	16
E	7	T	17
G	8	W	18
H	9	Y	19
I	10	V	20

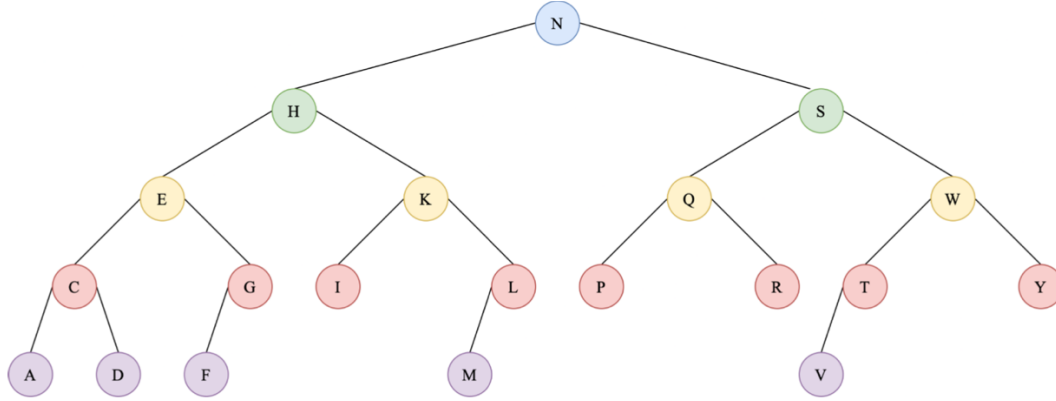
Tablo 10'da verilen değerler doğrultusunda $P(N) = [ARNDCQ \dots]$ biçimindeki bir protein dizilimi, tam sayı protein haritalama tekniği ile $C(N) = [1\ 2\ 3\ 4\ 5\ 6\ \dots]$ şeklinde sayısal hale getirilir.

1.5. Algoritma-Tabanlı Protein Haritalama Teknikleri

Algoritma-tabanlı protein haritalama tekniklerinde ise amino asit dizilimleri belirli bir algoritmik yapıya dayanarak sayısal ifadelere dönüştürülmektedir. Karakter-tabanlı yaklaşımlarda olduğu gibi belirli bir bilgiye gerek duyulmamaktadır. Ancak, bu tekniğin en büyük avantajı hızlı bir sayısallaştırma işlemi yapmasıdır. Bu kategoride değerlendirilen protein haritalama teknikleri ise AVL-ağacı ve FIBHASH yöntemleridir.

1.5.1. AVL-Ağacı Protein Haritalama Tekniği

AVL-ağacı protein haritalama tekniğinde ilk olarak amino asitler alfabetik olarak sıralanmış ve AVL-ağacına eklenmiştir. Tüm amino asitler AVL-ağacına eklendikten sonra gerekli işlemler yapılmış ve ağacın dengeli olması sağlanmıştır. Şekil 1’de amino asitlerin AVL-ağacına eklenmesi gösterilmiştir.



Şekil 1. Amino asitlerin AVL-ağacına eklenmesi

Ağaca eklendikten sonra her bir amino asidin derinlik değerleri hesaplanmış ve amino asit dizilimleri sayısal ifadelerle dönüştürülmüştür. Tablo 11’de amino asitlerin AVL-ağacı değerleri verilmiştir.

Tablo 11. Amino asitlerin AVL-ağacı değerleri

Amino asit kodu	Tam sayı değeri	Amino asit kodu	Tam sayı değeri
A	4	L	3
R	3	K	2
N	0	M	4
D	4	F	4
C	3	P	3
Q	2	S	1
E	2	T	3
G	3	W	2
H	1	Y	3
I	3	V	4

Tablo 11’de verilen değerler doğrultusunda $P(N) = [ARNDCQ \dots]$ biçimindeki bir protein dizilimi, AVL-ağacı protein haritalama tekniği ile $C(N) = [4\ 3\ 0\ 4\ 3\ 2 \dots]$ şeklinde sayısal hale getirilir.

1.5.2. FIBHASH Protein Haritalama Tekniği

FIBHASH protein haritalama tekniğinde amino asit dizilimlerine ilk aşamada Fibonacci sayıları atanmıştır. Doğada yirmi adet amino asit bulunduğu için Fibonacci serisinin yirminci sayısına kadar değerler göz önünde bulundurulmuştur. Ancak belirli bir seriden sonra, sayıların boyutu arttığı için bu değerler algoritma analizinde sıklıkla kullanılan bir tablo olan hash tablosuna atanmıştır. Bu sayede verilerin 1 ve 20 arasında olması sağlanmıştır. Aynı indekse gelen amino asit dizilimlerinin değerleri ise doğrusal sondalama yapılarak yeni indekslere aktarılmıştır. Tablo 12’de amino asitlerin FIBHASH değerleri verilmiştir.

Tablo 12. Amino asitlerin FIBHASH değerleri

Amino asit kodu	Tam sayı değeri	Amino asit kodu	Tam sayı değeri
A	1	L	15
R	10	K	14
N	7	M	9
D	3	F	5
C	2	P	16
Q	17	S	11
E	4	T	18
G	8	W	19
H	13	Y	20
I	6	V	12

Tablo 12’de verilen değerler doğrultusunda $P(N) = [ARND CQ \dots]$ biçimindeki bir protein dizilimi, FIBHASH protein haritalama tekniği ile $C(N) = [1\ 10\ 7\ 3\ 2\ 17 \dots]$ şeklinde sayısal hale getirilir.

Uygulama Soruları

1. WVTTLHLKGPHAQI şeklinde verilen protein dizilimini hidrofobiklik protein haritalama tekniği kullanarak sayısal ifadeye dönüştürünüz.
2. LAQM şeklinde verilen protein dizilimini Atchley faktörleri protein haritalama tekniği kullanarak sayısal ifadeye dönüştürünüz.
3. DFKFWNYQMGRQNRI şeklinde verilen protein dizilimini EIIP protein haritalama tekniği kullanarak sayısal ifadeye dönüştürünüz.
4. GHDDKVGAMD MELPA şeklinde verilen protein dizilimini BLOSUM62 protein haritalama tekniği kullanarak sayısal ifadeye dönüştürünüz.
5. DLQKFPWLYYWLRGC şeklinde verilen protein dizilimini PAM250 protein haritalama tekniği kullanarak sayısal ifadeye dönüştürünüz.
6. FKGD E MTHKEKPFSY şeklinde verilen protein dizilimini Miyazawa enerjileri protein haritalama tekniği kullanarak sayısal ifadeye dönüştürünüz.

7. HERGKVCPFMQEIMA şeklinde verilen protein dizilimini Micheletti potansiyelleri protein haritalama tekniği kullanarak sayısal ifadeye dönüştürünüz.

8. CTTVMIPQHGNAAHF şeklinde verilen protein dizilimini CPNR protein haritalama tekniği kullanarak sayısal ifadeye dönüştürünüz.

9. WNCQH şeklinde verilen protein dizilimini bir-sıcak protein haritalama tekniği kullanarak sayısal ifadeye dönüştürünüz.

10. SFWYFLGQSCSWLKE şeklinde verilen protein dizilimini tam sayı protein haritalama tekniği kullanarak sayısal ifadeye dönüştürünüz.

11. YAGFQPQNMLLGPIQ şeklinde verilen protein dizilimini AVL-ağacı protein haritalama tekniği kullanarak sayısal ifadeye dönüştürünüz.

12. GVSPADAESCIIMQY şeklinde verilen protein dizilimini FIBHASH protein haritalama tekniği kullanarak sayısal ifadeye dönüştürünüz.

13. PMWRITIFKQAKIP şeklinde verilen bir protein dizilimini sayısal hale getirecek kırmızı-siyah ağaçları tabanlı yeni bir yöntem geliştiriniz. Geliştirilecek olan yöntem sadece verilen protein dizilimini kapsayacaktır. İlk olarak amino asit dizilimleri alfabetik olarak sıralanacak ve ona göre ağaca yerleştirme işlemi yapılacaktır. En son ağaçta bulunan her bir amino asit kodunun derinlik değeri belirlenecek sayısallaştırma işlemi tamamlanacaktır.

14. YAHTGVPSNHNILVWNNGVSKKRNE şeklinde verilen bir protein dizilimini sayısal hale getirecek Lucas serisi tabanlı yeni bir yöntem geliştiriniz. Geliştirilecek olan yöntem sadece verilen protein dizilimini kapsayacaktır. İlk olarak amino asit dizilimleri alfabetik olarak sıralanacak ve amino asit kodlarına Lucas sayıları atanacaktır. Lucas sayılarının boyutu fazla olduğu için ardından hash tablosu kullanılacak ve sayısallaştırma işlemi tamamlanacaktır. Tabloda herhangi bir çakışma olması durumunda ise karesel sondalama kullanılacaktır.

Çözümler

1. WVTTLLEHLKGPHAQI şeklinde verilen protein dizilimini hidrofobiklik protein haritalama tekniği ile aşağıdaki gibi sayısal ifadeye dönüştürülür:

$$C(N) = [-0.9 \ 4.2 \ -0.7 \ -0.7 \ 3.8 \ -3.5 \ -3.2 \ 3.8 \ -3.9 \ -0.4 \ -1.6 \ -3.2 \ 1.8 \ -3.5 \ 4.5]$$

2. LAQM şeklinde verilen protein dizilimini Atchley faktörleri protein haritalama tekniği ile aşağıdaki gibi sayısal ifadeye dönüştürülür:

$$C(N) = [[-1.019 \ -0.987 \ -1.505 \ 1.266 \ -0.912] \ [-0.591 \ -1.302 \ -0.733 \ 1.570 \ -0.146] \ [0.931 \ -0.179 \ -3.005 \ -0.503 \ -1.853] \ [-0.663 \ -1.524 \ 2.219 \ -1.005 \ 1.212]]$$

3. DFKFWNYQMGRQNRI şeklinde verilen protein dizilimini EIIP protein haritalama tekniği ile aşağıdaki gibi sayısal ifade dönüştürülür:

$$C(N) = [0.1263 \ 0.0946 \ 0.0371 \ 0.0946 \ 0.0548 \ 0.00360.0516 \ 0.0761 \ 0.0823 \ 0.0050 \ 0.0959 \ 0.0761 \ 0.0036 \ 0.0959 \ 0]$$

4. GHDDKVGAMDMELPA şeklinde verilen protein dizilimini BLOSUM62 protein haritalama tekniği ile aşağıdaki gibi sayısal ifadeye dönüştürülür:

$$C(N) = [6 \ 8 \ 6 \ 6 \ 5 \ 4 \ 6 \ 4 \ 5 \ 6 \ 5 \ 5 \ 4 \ 7 \ 4]$$

5. DLQKFPWLYYWLRGC şeklinde verilen protein dizilimini PAM250 protein haritalama tekniği ile aşağıdaki gibi sayısal ifadeye dönüştürülür:

$$C(N) = [4 \ 6 \ 4 \ 5 \ 9 \ 6 \ 17 \ 6 \ 10 \ 10 \ 17 \ 6 \ 6 \ 5 \ 4]$$

6. FKGDETHKEKPFSY şeklinde verilen protein dizilimini Miyazawa enerjileri protein haritalama tekniği ile aşağıdaki gibi sayısal ifadeye dönüştürülür:

$$C(N) = [-0.33 \ 0.30 \ -0.02 \ 0.19 \ 0.15 \ -0.25 \ 0.05 \ -0.02 \ 0.30 \ 0.15 \ 0.30 \ 0.11 \ -0.33 \ 0.11 \ -0.23]$$

7. HERGKVCQPMQEIMA şeklinde verilen protein dizilimini Micheletti potansiyelleri protein haritalama tekniği ile aşağıdaki gibi sayısal ifadelere dönüştürülür:

$$C(N) = [0.001314 \ 0.008438 \ 0.009875 \ 0.000990 \ 0.005109 \ 0.001445 \ -0.002544 \ -0.003621 \ -0.013128 \ 0.031655 \ 0.006456 \ 0.008438 \ 0.006801 \ 0.031655 \ -0.001461]$$

8. CTTVMIPQHGNAAHF şeklinde verilen protein dizilimini CPNR protein haritalama tekniği ile aşağıdaki gibi sayısal ifadeye dönüştürülür:

$$C(N) = [11 \ 13 \ 13 \ 19 \ 1 \ 53 \ 7 \ 29 \ 17 \ 43 \ 41 \ 37 \ 37 \ 17 \ 3]$$

9. WNCQH şeklinde verilen protein dizilimini bir-sıcak protein haritalama tekniği ile aşağıdaki gibi sayısal ifadeye dönüştürülür:

$$C(N) = [00000000000000000010 \ 00000000000100000000 \ 01000000000000000000 \ 00000000000001000000 \ 0000001000000000000000]$$

10. SFWYFLGQSCSWLKE şeklinde verilen protein dizilimini tam sayı protein haritalama tekniği ile aşağıdaki gibi sayısal ifadeye dönüştürülür:

$$C(N) = [16 \ 18 \ 14 \ 19 \ 14 \ 11 \ 8 \ 6 \ 16 \ 5 \ 16 \ 18 \ 11 \ 12 \ 7]$$

11. YAGFQPQNMLLGPIQ şeklinde verilen protein dizilimini AVL-ağacı protein haritalama tekniği ile aşağıdaki gibi sayısal ifadeye dönüştürülür:

$$C(N) = [3 \ 4 \ 3 \ 4 \ 2 \ 3 \ 2 \ 0 \ 4 \ 3 \ 3 \ 3 \ 3 \ 2]$$

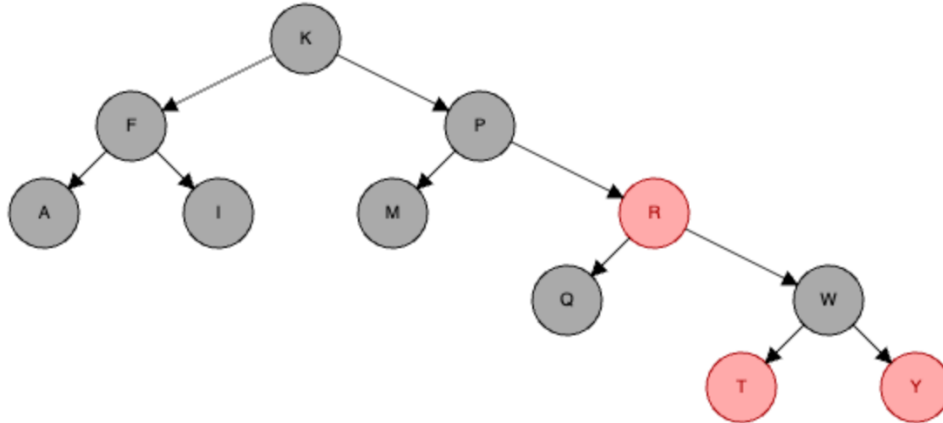
12. GVSPADAESCIIMQY şeklinde verilen protein dizilimini FIBHASH protein haritalama tekniği ile aşağıdaki gibi sayısal ifadeye dönüştürülür:

$$C(N) = [8 \ 12 \ 11 \ 16 \ 1 \ 3 \ 1 \ 4 \ 11 \ 2 \ 6 \ 6 \ 9 \ 17 \ 20]$$

13. PMWRITIYFKQAKIP şeklinde verilen bir protein dizilimini ilk olarak alfabetik sıralanır. Sıralama işlemi şu şekilde gerçekleşir:

A, F, I, K, M, P, Q, R, T, W ve Y

Ardından sırasıyla kırmızı-siyah ağaca eklenir. Ekleme işleminin ardından ağacın son durumu şu şekilde oluşur. Ardından her bir düğüm için derinlik değeri hesaplanır ve sayısallaştırma işlemi tamamlanır.



$$C(N) = [1 \ 2 \ 3 \ 2 \ 2 \ 4 \ 2 \ 4 \ 1 \ 0 \ 3 \ 2 \ 0 \ 2 \ 1]$$

14. YAHTGVPSNHNILVWNNGVSKKRNE şeklinde verilen bir protein dizilimini ilk olarak alfabetik biçimde sıralanır. Sıralama işleminin ardından dizilimler şu şekilde olur:

A, E, G, H, I, K, L, N, P, R, S, T, V, W, Y

Ardından her bir kodona Lucas serisindeki sayılar eklenir ve aşağıdaki değerler elde edilir:

$$A = 2, E = 1, G = 3, H = 4, I = 7, K = 11, L = 18, N = 29, P = 47, R = 76, S = 123, T = 199, V = 322, W = 521, Y = 843$$

Ardından hash tablosu oluşturulur ve değerler atanır. Toplamda 15 adet amino asit olduğu için hash tablosunun boyutu 15 olacaktır. Çakışmalar karesel sondalama ile giderilmiştir. Buna göre sayısallaştırma için gerekli olan değer şu şekilde olur:

W = 0, E = 1, A = 2, G = 3, H = 4, L = 5, P = 6, I = 7, T = 8, Y = 9, R = 10, K = 11, S = 12, V = 13, N = 14

En son sayısallaştırma işlemi ise şu şekilde gerçekleştirilir:

C(N) = [9 2 4 8 3 13 6 12 14 4 14 7 5 13 0 14 14 3 13 12 11 11 10 13 1]