

Kuvarsa Bağlanabilen Peptitler için Evrimsel Strateji ile Skor Matris Optimizasyonu

Score Matrix Optimization for Quartz Binding Peptides Using Evolutionary Strategy

Bariş Şenliol, Zehra Çataltepe

Bilgisayar Mühendisliği Bölümü
İstanbul Teknik Üniversitesi, İstanbul, Türkiye
{senliol, cataltepe}@itu.edu.tr

Özetçe

İnorganik materyallere yüksek oranda özgün olarak bağlanabilen peptitlerin bulunabilmesi, tiptan elektroniğe kadar birçok alanda uygulaması olan, aktif bir araştırma konusudur. Yeni peptitlerin bilgisayar ortamında yaratılması kolay, fakat bağlanma özelliklerinin deneysel olarak tespiti zaman alıcı ve pahalı bir işlemdir. Yapısal olarak benzer peptitler genellikle benzer fonksiyonları yerine getirmektedir. Blossum, PAM ve Gonnet gibi genel skor matrisleri amino-asit dizilerinin benzerliklerinin bulunmasında Needleman-Wunsch veya Smith-Waterman gibi hizalama algoritmaları tarafından sıklıkla kullanılmaktadır. Genel matrisler kullanılan veri kümesine uyumlu olmayabilir. İstenen peptit ailesine veya fonksiyona göre optimize edilmiş skor matrisleri, yeni yaratılan veya keşfedilen peptitlerin yüksek doğrulukta sınıflandırılmasında önemli bir katkıda bulunabilir. Bu çalışmada, kuvarsa bağlanan peptitler için, evrimsel strateji (ES) kullanılarak peptit sınıfına özel skor matrisleri yaratılmış ve bu matrislerin performansları yine kuvarsa özel olarak Ören v.d. (2007) tarafından oluşturulmuş Quartz I skor matrisi ile karşılaştırılmıştır. Evrimsel strateji ile optimize edilmiş matrislerin TSS, toplam benzerlik skoru, ölçütü kullanıldığında, güçlü, orta ve zayıf bağlanan peptitlerin sınıflandırılması ve bağ kuvvetlerine sıralanması işlemlerinde Quartz I matrisine göre toplam doğruluk oranı açısından yüzde 30'a ve evirtim sayısı açısından ise yüzde 40'a varan daha iyi performanslar sergilediği gözlenmiştir.

Abstract

Finding new peptides that can bind to some inorganic materials with high affinity and specificity is an active research problem with many application areas ranging from medicine to electronics. While it is easier to design peptides in-silico, in-vitro testing of binding is a slow and expensive process. Peptides with similar structure usually have similar functions. Score matrices, such as Blossum, PAM and Gonnet are used to score amino acid replacements in alignment algorithms such as Needleman-Wunsch or Smith-Waterman which measure similarity between two sequences. These score matrices may not represent the frequency of amino-acids replacement for experimental data. Score matrices which are optimized for specific peptide class or function have an important effect on accurate classification of artificial or newly discovered peptides. In this paper, score matrices that are specific to quartz binding peptides are derived using an

evolutionary strategy (ES) and performance of these matrices are compared with the Quartz I matrix of Ören et.al. (2007). Results show that the matrices which obtained from the ES have better performance than Quartz I matrix when TSS (total similarity score) is used as a performance measurement criterion. Using ES scoring matrices and TSS for classifying peptides as strong, moderate or weak has performed 30% better in terms of accuracy and 40% better in terms of the number of inversions of affinity ordering of peptides than Quartz I matrix.

1. Giriş

Proteinler arasındaki ilişkilerin bulunması biyoinformatikteki önemli konulardan biridir. Bu ilişkilerin bulunabilmesi için öncelikle proteinlerin sahip oldukları fonksiyonların tanınması gerekir. Fonksiyonların tanınabilmesine yönelik yapılan çalışmalarda genellikle proteinin içinde bulunan peptitlerden, kısa amino-asit sekanslarından faydalanılır [1].

Önceki çalışmalar göstermiştir ki doğada bulunan benzer fonksiyonlara sahip proteinler evrimsel, biyokimyasal ve biyofiziksel kısıtlardan ötürü benzer amino-asit dizilerine sahiplerdir [2]. Bu nedenle fonksiyonu bilinmeyen bir sekans ile karşılaşıldığında ilk olarak fonksiyonu bilinen sekanslarla benzerliğine bakılır [3]. Proteinlerin birbirleri ile sahip oldukları benzerlikler evrimsel bağlantıların anlaşılmasının yanı sıra yapısal ve fonksiyonel korunumlar konusunda da kanıt olarak kullanılabilir [2].

Benzerlik ölçme işleminde protein sekans hizalama algoritmalarından faydalanılır. Protein sekans hizalama algoritmaları biyologların, bir proteinin evrimsel geçmişini veya evrimlerinden kalan kilit fonksiyonel kalıntıları belirleyebilmek gibi analizleri yapılabilme amacıyla sıklıkla kullandıkları araçlardan biridir [2].

Proteinlerin birbirleri ile benzerlikleri protein sekans hizalama gibi dinamik programlamaya dayalı yöntemlerle hesaplanabilmektedir [2,3]. Global hizalama yapan Needleman-Wunsch [4] algoritması ile yerel hizalama yapan Smith-Waterman [5] algoritması gibi yöntemler iki proteinin benzerlik skorunun hesaplanmasında kullanılmaktadır. Bu tip yöntemlerin anahtar parçası skor matrisi adı verilen matrislerdir. Skor matrisi, proteinlerin hizalanması sonucunda eşleşen her amino-asit için bir maliyet barındıran genellikle simetrik bir matristir. Algoritmalar benzerlik skorlarını hesaplarlarken skor matrislerinden ve boşluk penaltılarından yararlanmaktadır. Özellikle proteinlerin hizalanmasında

kullanılan BLOSUM [6], PAM [7] ve Gonnet [8] adında hazır skor matrisleri mevcuttur. Bu matrisler, tüm protein fonksiyonlarının, fonksiyonel ve fonksiyonel olmayan bölgelerin ve türlerin mevcut sekans verilerinden yararlanarak hesaplanan mutasyon oranlarının ortalaması alınarak oluşturulmuştur [1].

Bazı veri kümelerinde protein sekanslarının amino-asit frekansları genel karakteristiğe uymamaktadır [1, 9]. Üstelik bu ender sayılabilecek bir durum da değildir. Bu gibi durumlarda daha önceden oluşturulmuş skor matrisleri veri kümesine uyumlu olarak çalışmayabilir. Bu nedenle veri kümesinin daha doğru anlaşılabilmesi ve sonuçlar alınabilmesi amacıyla amino-asit frekanslarını daha iyi yansıtacak, veri kümesine daha uyumlu bir skor matrisinin bulunması gerekir. Yapılan çalışmalar var olan matrislerin, matrisin hedef frekans değerleri değiştirilerek veya veri kümesine benzer peptitlerle baştan oluşturularak veri kümesine uygun hale getirilebileceğini göstermiştir [9]. Bu konuda yapılan başka çalışmalarda ise standart veya rastgele değerlere sahip matrisler ile sınıflandırıcılardan geri beslemeler kullanılarak veri kümesine özel matrisler bulunabildiği kaydedilmiştir [1].

Skor matrislerinin var olan veya rastgele oluşturulmuş matrislerden geri beslemeler ile veri kümesine uygun hale getirilebilmesinde iki farklı yaklaşım kullanılmaktadır. Birinci yaklaşımda optimizasyon doğrusal bir şekilde tek matris üzerinden performansı arttıracak değişiklikler ile yapılmakta iken ikinci yaklaşımda evrimsel stratejilerden faydalanılmaktadır. Evrimsel stratejilerin (ES) biyoinformatik dahil bir çok optimizasyon probleminde diğer yöntemlerden daha iyi sonuç verdiği görülmüştür [1].

Bu çalışmada evrimsel stratejiden yararlanılarak kuvarsa bağlanabilen peptitler üstünde skor matris optimizasyonu yapılmıştır. Kuvars gibi inorganik materyallere bağlanabilen peptitler üzerine yapılan çalışmalar sayesinde son yıllarda metaller, minerallere ve oksitlere bağlanabilen birçok peptit bulunmuştur. Bu peptitlerin elektronik ve tıp gibi alanlarda inorganik materyallerin sentezi, kurulumu ve şekillendirilmesi için kullanılabilirliği oluşu bu konuya ilgiyi daha da arttırmaktadır [2]. Veri kümesinin amino-asit frekanslarının yanı sıra aynı fonksiyona sahip peptitleri doğru bir benzerlik ölçütü ile karşılaştırarak bilgisayar ile yeni peptitler üretmek araştırmacıların eğildiği konulardan biridir. Bu durum veri kümesine özel skor matrislerinin oluşturulmasını zorunlu kılmaktadır [2]. Bu çalışmada skor matrisi optimizasyonu üstüne kuvarsa bağlanabilen peptitlerden yararlanarak deneyler yapılmış ve daha önce kuvarsa özel olarak bulunmuş skor matrisinin sonuçları ile karşılaştırılarak geliştirilen yöntemin başarısı ortaya koyulmuştur.

2. Önceki Çalışmalar ve Veri Kümesi

Daha önce kuvarsa bağlanabilen peptitler üzerine yapılan bir çalışmada [2] PAM250 matrisinden iteratif değişikliklerle güçlü bağa sahip peptitlerin diğerlerinden ayrılması sağlanmaya çalışılmıştır. Kuvars veri kümesinde [2] 10 güçlü, 14 orta ve 15 zayıf olmak üzere 39 tane peptit bulunmaktadır. Bu çalışma boyunca peptitler arası benzerlikler TSS (toplam benzerlik skoru) adı verilen kritere göre ölçülmüştür. Her iterasyonda güçlü bağa sahip peptitlerin diğer güçlüler ile sahip olduğu ortalama benzerlik skoru, TSSs-s, ile zayıf bağa sahip peptitlerin güçlüler ile sahip olduğu ortalama benzerliğin, TSSs-w, arası açılmaya çalışılmıştır. Ayrıca yapılan deneyler ile zayıf bağlanan peptitlerin kendi aralarında düşük ortalama benzerlik skorlarına TSSw-w sahip olduklarını gösterilmiştir. Bu çalışmada sınıflandırıcı olarak her peptitin

güçlü bağlandığı bilinen peptitler ile ortalama benzerlik skoru, TSSp-s, kullanılmıştır. Bu skora göre sıralanan peptitlerden ilk onda (güçlü sayısı) kaç adet güçlü peptitin olduğu doğruluk derecesi olarak kabul edilmiştir.

Yukarıda belirtilen iteratif yöntem ile elde edilen Quartz I [2] adındaki matris Blosum62 ve PAM250 matrislerinden doğruluk oranı açısından çok daha iyi sonuç vermiştir. Ayrıca veri kümesinin amino-asit frekansından faydalanarak üretilen yapay peptitler, Quartz I'den faydalanılarak var olanlar ile karşılaştırılmıştır. Güçlü bağa sahip olabilecekleri veri kümesinin eşik değerlerinin yardımı ile belirlendikten sonra seçilen bir kısım yapay peptit, laboratuvar ortamında gerçekleştirilmiştir. Bu peptitlerle yapılan deneyde, veri setindeki peptitlerden daha iyi bağlantı derecesine sahip peptitler gözlenmiştir. Bu durum veri kümesine özel skor matrisi optimizasyonun bilgisayarla peptit üretiminde gerekliliğini ve faydasını göstermektedir.

3. Evrimsel Strateji ile Skor Matris Optimizasyonu

Bu çalışmada skor matrisinin optimizasyonu, tek matris üzerinden iteratif bir yöntem yerine evrimsel stratejiden faydalanılarak sağlanmıştır. Evrimsel strateji, genetik algoritmalar arasında kendi kendine adapte olabilen mutasyon adım büyüklükleri sayesinde veriye göre sonuca yakınsayan başarılı bir optimizasyon yöntemidir [10].

Uygulanan evrimsel algoritmanın genel özellikleri Tablo 1'de, sözde kodu Şekil 1'de gösterilmiştir.

Tablo 1: Genel Özellikler

Gösterim	Gerçel Sayılı Vektörler
Üreme	Yerel Ayrık
Mutasyon	Kendinden Adapteli Adım Büyüklükleri
Ebeveyn Seçimi	Uniform Rastgele
Yaşayan Seçimi	Elitist ($\mu+\lambda$)
Durma Şartı	10 İterasyon Değişim Olmaması

Yapılan bu çalışma kapsamında yerel minimumlara takılma şansı daha az olması nedeniyle [10] elitist ($\mu+\lambda$)-ES kullanılmıştır. Skor matrisi simetrik kabul edilip diyagonal dâhil üst yarısı 325 gerçel sayıdan oluşan bir vektör (kromozom) ile temsil edilmiştir. Ayrıca her skor matrisinin 325 gerçel sayıdan oluşan bağımsız mutasyon adım büyüklüğü σ vektörü (strateji parametreleri) de bulunmaktadır. Her matrisin strateji parametreleri -2 ile 2 arasındaki tam sayılardan rastgele şekilde ilklendirilmekte ve nesilden nesile (1) ile adapte olmaktadır. Adaptasyon işlemi için gereken genel ve bireysel öğrenme oranları [10]'da belirtildiği gibi sırası ile $\tau_0 = 1/\sqrt{2 \cdot 325}$ ve $\tau = 1/\sqrt{2 \cdot \sqrt{325}}$ alınmıştır.

$$\vec{\sigma} := \exp(\tau_0 \cdot N(0,1)) \cdot (\sigma_1 \cdot \exp(\tau \cdot N(0,1)), \dots, \sigma_N \cdot \exp(\tau \cdot N(0,1))) \quad (1)$$

Evrimsel strateji $\mu=10$ ve $\lambda=70$ parametreleri ile önceki çalışmada kullanılan PAM250 matrisinin eşleniği olan Blosum45 matrisinin farklı mutasyon değerlerine sahip on adet kopyası kullanılarak başlamaktadır. Algoritma boyunca 2 bireyin evliliğinden 1 çocuk oluşmakta ve çocuğun matris ve mutasyon vektörü değerleri yerel ayrık yani rastgele şekilde aneden veya babadan parçalar halinde gelmektedir. Bu şekilde 70 adet çocuk oluşturulmakta ve yeterlilik değerlerine göre ebeveyn ve çocuklardan en iyi 10 tanesi bir sonraki nesile aktarılmaktadır.

```

1 Başla
2   g := 0;
3   oluştur(Eμ, Eσμ);
4   Fμ := yetfonk1(Eμ);
5   TSS := ortTSSfark(Eμ);
6   Tekrar
7     Devam i := 1'den λ Yap
8       S := seç_ebeveyn(Eμ);
9       Ci := üreme(S);
10      Cσi := üreme(S);
11      Cσi := σ_mutasyon(Cσi);
12      Ci := mutasyon(Ci, Cσi);
13      Fi := yetfonk1(Ci);
14      TSSci := TSSfark(Ci);
15     Bit
16     C := ele_cocuk(C, TSSci, TSS);
17     Eμ := seç_yaşayan(E, C, Fμ, F, μ);
18     Fμ := yetfonk1(Eμ);
19     TSS := ortTSSfark(Eμ);
20     g := g+1;
21   Dön durma şartı sağlanmadıysa
22     Fμ := yetfonk2(Eμ);
23     D := maks_doğruluk(Eμ);
24     g := 0;
25   Tekrar
26     Devam i := 1'den λ Yap
27       S := seç_ebeveyn(Eμ);
28       Ci := üreme(S);
29       Cσi := üreme(S);
30       Cσi := σ_mutasyon(Cσi);
31       Ci := mutasyon(Ci, Cσi);
32       Fi := yetfonk2(Ci);
33       Dci := doğruluk(Ci);
34     Bit
35     C := ele_cocuk(C, Dci, D);
36     Eμ := seç_yaşayan(Eμ, C, Fμ, F, μ);
37     Fμ := yetfonk2(Eμ);
38     D := maks_doğruluk(Eμ);
39     g := g+1;
40   Dön durma şartı sağlanmadıysa
41     Sonuc=seç_min_yetfonk2(Eμ)
42 Bit

```

Şekil 1: (μ+λ)-ES sözde kodu

Yeterlilik ölçümü için 2 farklı fonksiyon kullanılmaktadır. İlk fonksiyona göre çalışan ilk bölüm durma kriterini sağlandıktan sonra o andaki ebeveyn bireyler diğer yeterlilik fonksiyonuna göre çalışacak ikinci bölümün ilk bireylerini oluşturmaktadır. Yöntemin sonucu ancak 2. yeterlilik fonksiyonu ile bir yakınsama sağlandığında ortaya çıkmaktadır. İlk fonksiyon (2) güçlü bağa sahip bireyleri diğerlerinden ayırarak önceki çalışmada kullanılan sınıflandırıcıya göre ilk on sıraya çekmeye çalışırken, 2. fonksiyon (3) güçlü olanları sıralamada yukarıda tutarken aynı anda diğerlerini gerçek bağ yapabileme gücüne göre sıralanması için kullanılmaktadır. Her 2 fonksiyon da bireylerde minimum noktaları olan 0'a yakınsamaya çalışmaktadır.

$$\text{yet1} = (\text{topla}(\text{sıra}(\text{güçlü})) - 55) / (\text{güçlü doğruluk oranı}) \quad (2)$$

$$\text{yet2} = \text{topla}(\text{evirtim}(\text{tüm})) / (\text{güçlü doğruluk oranı}) \quad (3)$$

(2) ve (3)'ten görülebileceği üzere ilk fonksiyon peptitler TSSp-s skorlarına göre sıralandıktan sonra güçlülerin kaçınıcı sırada olduklarını toplamakta ve en iyi koşulda olacakları yerlerin toplamını çıkararak sonuç üretmektedir. Kendini minimize etmeye çalışan bu fonksiyon güçlülerin gerçek sıralamalarına bakmadan hepsinin olabildiğince üst sıralarda olmasını sağlamaya yöneliktir. Ancak hesaplanan değer güçlülerin ilk 10 da bulunma oranına bölünerek yüksek doğruluk değerine sahip ama kötü skor üreten matrislere de şans verilmesi sağlanmaktadır. İkinci fonksiyon ise tüm peptitlerin gerçek olması gerek yerlerinden ne kadar uzakta olduklarını (evirtim, inversions) hesaplayarak toplamakta bu sayede her matris için bir gerçekten uzaklık skoru çıkarmaktadır. Güçlü peptitlerin doğruluk oranı ile ölçüklenen bu fonksiyonda da güçlü doğruluk oranı yüksek ama nispeten kötü evirtim skoruna sahip matrislere de şans verilmesi sağlanmaktadır.

Önceki çalışmada olduğu gibi TSSs-s ve TSSw-s arası farkı açmak için birinci yeterlilik fonksiyonu boyunca bazı bireylere penaltı uygulanmaktadır. O anki ebeveynlerin sahip olduğu ortalama fark değeri referans kabul edilip o değerden ufak değere sahip çocuk bireylerin seçilebilmesi engellenmekte ve bu sayede dengeli bir şekilde artan bir fark yaratılmaktadır. İkinci yeterlilik fonksiyonu boyunca ise ilk bölümün sağlamış olduğu güçlü peptitlerin yüksek doğruluk oranının koruması için aynı doğruluk oranını sağlayamayan çocukların seçilebilmesi engellenmektedir. Bu sayede son durumda yeterli TSS farkına sahip yüksek doğruluk oranında bireylerin olması sağlanmaktadır.

TSS hesapları için gereken tüm peptidlerin birbirleri ile benzerlik skorlarının bulunduğu benzerlik matrisinin oluşturulmasında C ile yazılmış Balign [11] programından faydalanılmıştır. Bu çalışmada skor matrisi ile birlikte kullanılan boşluk açma ve boşluk uzatma penaltıları sırasıyla birçok çalışmada uygun olduğu gösterilen 10 ve 1 değerleri ile sabitlenmiştir [1, 2].

Evrimsel algoritma ikinci yeterlilik fonksiyonu için bir noktaya yakınsadıktan sonra, yani ebeveynlerin ortalama ikinci yeterlilik fonksiyonu değerleri 10 iterasyon boyunca değişmediği anda sonlamaktadır. Bu noktada 10 ebeveyninden en düşük 2. yeterlilik fonksiyonuna sahip birey algoritmanın sonucu olan skor matrisi olarak seçilmektedir.

4. Deneysel Sonuçlar

4.1 Deneysel Kurulumu

Deneysel daha önce belirtilen 39 peptitten oluşan kuvars veri kümesi üzerinde uygulanmıştır. Beş farklı ölçüm; güçlü, orta, zayıf peptit doğruluk oranları, TSSs-s ve TSSs-w arası fark, TSSfark, ve toplam evirtim sayısı performans kriteri olarak belirlenmiş ve bu değerler 20 farklı çalışma için gözlenmiştir. Doğruluk değerleri, skor matristen yararlanarak çıkartılmış benzerlik matrisinden yararlanarak bulunan TSSp-s değerlerine göre hesaplanmaktadır. Bu değerlere göre yapılan sıralamada güçlülerin 1-10 arası, ortaların 11-24 arası ve zayıfların 25-39 arası bulunmaları kendileri için doğruluk oranı olarak hesaplanmıştır. Genel performansın değerlendirilmesi için evirtim ve TSS için medyan, doğruluk oranları için ise ortalama değerler kullanılmıştır.

4.2 Sonuçlar

Yirmi farklı çalışmada elde edilen sonuçların medyan ve ortalama değerleri karşılaştırma yapabilmek amacıyla Quart Z

'in performans değerleri ile birlikte Tablo 2’te gösterilmektedir. Tablo 3’de ise her çalışma ile ilgili ölçümler yer almaktadır.

Tablo 2: Evrimsel Strateji ile Quartz I karşılaştırması

	Ölçüm Kriterleri				
	Doğ-G	Doğ-O	Doğ-Z	Evirtim	TSSfark
Ort	1	0.757	0.773	171	28.3776
Qtz	0.8	0.571	0.733	238	19.4178

Tablo 3: 20 Farklı Çalışma Sonuçları

#	Ölçüm Kriterleri				
	Doğ-G	Doğ-O	Doğ-Z	Evirtim	TSSfark
1	1	0.786	0.8	138	25.6606
2	1	0.786	0.8	170	34.562
3	1	0.857	0.867	158	57.7712
4	1	0.643	0.667	176	35.1204
5	1	0.786	0.8	186	30.539
6	1	0.786	0.8	168	23.6075
7	1	0.857	0.867	162	27.9149
8	1	0.643	0.667	172	41.7946
9	1	0.786	0.8	154	26.0111
10	1	0.786	0.8	162	35.9801
11	1	0.643	0.667	218	26.3868
12	1	0.857	0.867	174	25.4598
13	1	0.786	0.8	154	52.0261
14	1	0.786	0.8	168	26.5847
15	1	0.714	0.733	182	20.3613
16	1	0.714	0.733	192	43.5815
17	1	0.786	0.8	136	35.9298
18	1	0.786	0.8	198	28.8404
19	1	0.643	0.667	198	22.7522
20	1	0.714	0.733	192	20.5034

İstatistiksel olarak evrimsel strateji sonuçlarının Quartz I’den daha iyi olduğunu gösterebilmek amacıyla doğruluk oranları için one-sample t-test, evirtim sayısı ve TSSfark değerleri için ise Wilcoxon signed-rank testleri %5 anlamlılık derecesinde yapılmıştır. T-test için p değeri ve güven aralığı, Wilcoxon için ise yalnız p değeri hesaplanmıştır. Sonuçlar Tablo 4’te gösterilmektedir. Sonuçlara göre tüm ölçütler için evrimsel algoritma Quartz I’den daha iyi sonuç vermiştir.

Tablo 4: İstatistiksel Karşılaştırma

	Ölçüm Kriterleri				
	Doğ-G	Doğ-O	Doğ-Z	Evirtim	TSSfark
P	0	3.9941 e-10	0.0142	8.7949 e-5	8.8575 e-5
CI	[1 1]	[0.724 0.790]	[0.742 0.804]	-	-

5. Sonuç

Bu çalışma ile skor matrislerin belli bir peptid sınıfına veya ailesine optimize edilebilmesi konusunda evrimsel strateji kullanan bir yaklaşım geliştirilmiştir. Kuvars üzerinde yapılan deneylerin gösterdiği üzere, deneysel olarak başarılı kabul edilmiş matrislerden bile daha iyi performans sergileyebilen bu yaklaşım sayesinde daha iyi sınıflandırma performansları elde edilebilmektedir. İstenilen sınıftan peptitlerin birbirlerine daha benzer, diğerlerinden daha ayrık olması ve bunu yaparken de bağlanabilme gücüne göre gerçek sıraya uygun

hale getirilmesi, yapay olarak üretilen peptitlerin performansları konusunda daha iyi bir öngörü sağlayacaktır. Bu nedenle skor matris optimizasyonunda evrimsel stratejinin kullanımı sınıflandırma performansındaki artışın yanı sıra, yapay peptit üretiminde daha efektif olunabilmesini de beraberinde getirecektir.

6. Teşekkür

İki yazar da Tübitak projesi 105E164 tarafından kısmen desteklenmiştir. Yazar Şenliol, Turckcell Teknoloji’ye ve Türkiye Bilişim Derneği’ne yüksek lisans bursu için teşekkür eder. Kuvars bağlanan peptitler problemi ile bizi tanıştıran ve bilgilendiren, İTÜ Moleküler Biyoloji ve Genetik Bölümü’nden Dr. Candan Tamerler’e ve Washington Üniversitesi, Malzeme Bilimi ve Mühendisliği Bölümü’nden Dr. Mehmet Sarıkaya’ya, Nature Inspired Computing yüksek lisans dersinde verdiği bilgiler için İTÜ Bilgisayar Mühendisliği Bölümü’nden Dr. Şima Etaner-Uyar’a ve ayrıca İTÜ yüksek lisans öğrencileri Eser Aygün’e balıgn programı için, Erdem Çiçek ve Sanjarberk Hudaiberdiev’e doğrusal arama ile gerçekleştirdikleri skor matrix optimizasyonu çalışmalarını hakkında verdikleri bilgiler için teşekkür ederiz.

7. Kaynakça

- [1] David C. Trudgian and Zheng Rong Yang, Substitution Matrix Optimisation for Peptide Classification, In Proceedings of EVOBIO 2007, Valencia, Spain, Lecture Notes In Computer Science, vol. 4447, pp. 291-300, 2007.
- [2] Oren EE, Tamerler C, Sahin D, Hnilova M , Seker UOS, Sarıkaya M, Samudrala R, A novel knowledge-based approach to design inorganic-binding peptides, Bioinformatics, 23 (21), 2816-2822, 2007
- [3] Hong C, Tewfik A, Designing a Searching Space for Pairwise Global Sequence Alignments Over Multiple Scoring Systems, EUSIPCO 2007
- [4] Needleman, S.B. and Wunsch, C.D. (1970) A general method applicable to search for similarities in amino acid sequence of 2 proteins. J. Mol. Biol., 48, 443-453.
- [5] Smith, T.F. and Waterman, M.S. (1981) Identification of common molecular subsequences. J. Mol. Biol., 147, 195-197.
- [6] Henikoff, S. and Henikoff,J.G. (1992) Amino-acid substitution matrices from protein blocks. Proc. Natl Acad. Sci. USA, 89, 10915-10919.
- [7] Dayhoff,M.O., Schwartz R. M. And Orcutt B.C. (1978) A model of evolutionary change in proteins. In Dayhoff,M.O. (ed.), Atlas of Protein Sequence and Structure. Vol. 5, (Suppl. 3). National Biomedical Research Foundation, Washington DC, pp. 345-352.
- [8] Gonnet,G.H., Cohan M. A. And Benner S. A. (1992) Exhaustive matching of the entire protein-sequence database. Science, 256, 1443-1445
- [9] Y. Yu and J. C. Worrone and S. F. Altschul, “The compositional adjustment of. amino acid substitution matrices”, PNAS, vol. 100, pp. 15688-15693, 2003
- [10] H.-G. Beyer and H.-P. Schwefel. Evolution Strategies: A Comprehensive Introduction. Journal Natural Computing, 1(1):3-52, 2002.
- [11] Aygun, E. and Cataltepe, Z. (2009), hazırlanma aşamasında.