

## Exhaustive Search

- Kapsamlı arama algoritmaları tasarlamak için az çaba gerektirir, ancak ilgi duyulan birçok sorun için kabul edilebilir sürelerde makul büyüklükteki girdileri işleyemez.
- Bu soruna rağmen, kapsamlı arama veya kaba kuvvet algoritmaları genellikle daha verimli algoritmalar tasarlamamanın ilk adımıdır.

## Kısıtlama Haritalama

- Hamilton Smith, 1970 yılında HindII kısıtlama enziminin DNA moleküllerini parçaladığını keşfetti.
- GTGCAC veya GTTAAC sekansları, uzun bir molekülü bir dizi kısıtlama fragmanına böler.
- Kısa bir süre sonra, DNA moleküllerindeki kısıtlama alanlarının haritaları veya kısıtlama haritaları, belirli genetik işaretleyicilerin konumunun daraltılmasına yardımcı olarak moleküler biyolojide bir alanı haline geldi.

## Kısıtlama Haritalama

- Bir organizmanın genomik DNA dizisi biliniyorsa, HindII için bir kısıtlama haritasının oluşturulması, genomdaki tüm GTGCAC ve GTTAAC oluşumlarını bulmak anlamına gelir.
- İlk bakteri genomu, kısıtlama enzimlerinin keşfedilmesinden yirmi beş yıl sonra sıralandığı için, uzun yıllar boyunca biyologlar, genomların DNA dizisi hakkında önceden bilgi sahibi olmadan genomlar için kısıtlama haritaları oluşturmaya zorlandı.

## DNA Dizilerindeki Düzenleyici Motifler

- İnsanlar gibi meyve sinekleri de bakteri ve diğer patojenlerin neden olduğu enfeksiyonlara karşı hassastır.
- Meyve sinekleri, insanlar kadar gelişmiş bir bağışıklık sistemine sahip olmasalar da, sinek genomunda genellikle uykuda olan, ancak organizma enfekte olduğunda bir şekilde devreye giren küçük bir bağışıklık genleri kümesine sahiptirler.
- Bu genler açıldığında, patojeni yok eden, genellikle enfeksiyonu iyileştiren proteinler üretirler.

## DNA Dizilerindeki Düzenleyici Motifler

- Yapılan deneyler ile Meyve sineği genomundaki birçok bağışıklık geninin, genlerin başlangıcının üst kısmında bulunan TCGGGGATTCC'yi anımsatan dizelere sahip olduğu ortaya çıktı.
- Bağlanma alanları olarak adlandırılan bu kısa diziler, bağışıklığı ve diğer genleri etkinleştiren düzenleyici motiflerin önemli örnekleridir.
- Transkripsiyon faktörleri olarak bilinen proteinler bu motiflere bağlanarak RNA polimerazı aşağı akım genleri kopyalamaya teşvik eder.
- Motif bulma, bu tür motifleri, motiflerin nasıl görüldüğüne dair herhangi bir ön bilgi olmadan keşfetme problemidir.

## DNA Dizilerindeki Düzenleyici Motifler

- Sinek enfeksiyonu deneyi, bir genomdan bir dizi sekans verildiğinde, şaşırtıcı bir şekilde sıkça meydana gelen kısa alt dizileri bulabilen bir algoritma gerektirir.
- Motif bulmaya yönelik popüler bir yaklaşım, sık veya nadir kelimelerin DNA'daki düzenleyici motiflere karşılık gelebileceği varsayımına dayanıyor.
- Bir kelime beklenenden çok daha sık geçerse, o zaman bir tür "sinyal" olma olasılığı daha yüksektir ve sinyalin biyolojik anlamını bulmak çok önemlidir.
- Bu yaklaşım, olası tüm kalıpları numaralandırmaya ve aralarından en sık (veya istatistiksel olarak en anlamlı olanı) seçmeye dayanır.

## DNA Dizilerindeki Düzenleyici Motifler

- Şekilde rastgele oluşturulan yedi adet 32 nükleotidlik DNA dizisini göstermektedir.

```
CGGGGCTGGGTCGTCACATTCCCCTTTGATA
TTTGAGGGTGCCCAATAACCAAAGCGGACAAA
GGGATGCCGTTTGACGACCTAAATCAACGGCC
AAGGCCAGGAGCGCCTTTGCTGGTTCTACCTG
AATTTTCTAAAAAGATTATAATGTCGGTCCTC
CTGCTGTACAACCTGAGATCATGCTGCTTCAAC
TACATGATCTTTTGTGGATGAGGGAATGATGC
```

(a) Seven random sequences.

## DNA Dizilerindeki Düzenleyici Motifler

- Aynı dizilerin rastgele pozisyonlarına  $l=8$  uzunluğundaki gizli model  $P = \text{ATGCAACT}$  eklenmiş hali şekilde yer almaktadır.

```
CGGGGCTATGCAACTGGGTCGTCACATTCCCCTTTGATA
TTTGAGGGTGCCCAATAAATGCAACTCCAAAGCGGACAAA
GGATGCAACTGATGCCGTTTGACGACCTAAATCAACGGCC
AAGGATGCAACTCCAGGAGCGCCTTTGCTGGTTCTACCTG
AATTTTCTAAAAAGATTATAATGTCGGTCCATGCAACTTC
CTGCTGTACAACCTGAGATCATGCTGCATGCAACTTCAAC
TACATGATCTTTTGTGGATGAGGGAATGATGC
```

(b) The same DNA sequences with the implanted pattern ATGCAACT.

## DNA Dizilerindeki Düzenleyici Motifler

- P modelinin ne olduğunu veya her dizide nereye yerleştirildiğini bilmediğinizi varsayalım.
- DNA dizilerini analiz ederek P'nin ne olduğunu tespit edebilir misiniz?

```
CGGGGCTATGCAACTGGGTCGTCACATTCCCCTTTTCGATA
TTTGAGGGTGCCCAATAAATGCAACTCCAAAGCGGACAAA
GGATGCAACTGATGCCGTTTGACGACCTAAATCAACGGCC
AAGGATGCAACTCCAGGAGCGCCTTGTCTGGTTCTACCTG
AATTTTCTAAAAAGATTATAATGTCGGTCCATGCAACTTC
CTGCTGTACAACCTGAGATCATGCTGCATGCAACTTTCAAC
TACATGATCTTTTGATGCAACTTGGATGAGGGAATGATGC
```

(c) Same as (b), but hiding the implant locations. Suddenly this problem looks difficult to solve.

## DNA Dizilerindeki Düzenleyici Motifler

- Her bir l-mer dizisinin kaç kez ortaya çıktığını basitçe sayabiliriz.
- Örnekte yalnızca  $7 \cdot (32 + 8) = 280$  nükleotid olduğundan, eklenmiş model dışındaki herhangi bir 8-mer'in birden fazla görülmesi olası değildir.
- Şekil (c) 'de tüm 8-mer oluşumlarını saydıktan sonra, 8-merlerin çoğunun numunede sadece bir kez görünmesine rağmen, numunede şüpheli bir şekilde çok sayıda görünen bir 8-mer olduğunu gözlemleyeceğiz.
- Bu aşırı temsil edilen 8-mer, bulmaya çalıştığımız P modelidir.

## DNA Dizilerindeki Düzenleyici Motifler

- Yukarıdaki basit eklenti kalıplarımızın aksine, DNA bazı nükleotid pozisyonlarında mutasyonlara izin vererek daha yaratıcı bir düzenleyici motif kavramı kullanır.

```
CGGGGCTATcCAgCTGGGTCGTCACATTCCCCTTTTCGATA
TTTGAGGGTGCCCAATAAagGCAACTCCAAAGCGGACAAA
GGATGgAtCTGATGCCGTTTGACGACCTAAATCAACGGCC
AAGGAaGCAACcCCAGGAGCGCCTTGTCTGGTTCTACCTG
AATTTTCTAAAAAGATTATAATGTCGGTCCtTGgAACTTC
CTGCTGTACAACCTGAGATCATGCTGCATGccAtTTTCAAC
TACATGATCTTTTGATGgcACTTGGATGAGGGAATGATGC
```

(d) Same as (b), but with the implanted pattern ATG-CAACT randomly mutated in two positions; no two implanted instances are the same. If we hide the locations as in (c), the difficult problem becomes nearly impossible.

## DNA Dizilerindeki Düzenleyici Motifler

- İmplant edilen P paterninin mutasyona uğramasına izin verildiğinde, 8-mer sayısı paterni göstermediğinden P'nin yeniden yapılandırılması daha karmaşık hale gelir.
- Aslında, ATGCAACT dizisi şekil (d) 'de görünmüyor bile, ancak onun yedi mutasyona uğramış versiyonu, birinci sekansta 8. pozisyonda, ikinci sekansta 19. pozisyonda, üçüncü sırada 3, dördüncü sırada 5, Beşincide 31, altıncıda 27 ve yedide 15. sırada yer almaktadır.

## DNA Dizilerindeki Düzenleyici Motifler

- Motif bulma problemini açık bir şekilde formüle etmek için, "motif" ile ne demek istediğimizi tam olarak tanımlamamız gerekir.
- Bir motifi temsil etmek için tek bir dizgeye güvenmek, gerçek biyolojik dizilerde desenin varyasyonunu temsil etmekte genellikle başarısız olur.
- Bir motifin daha esnek bir temsili bir profil matrisi kullanır.

## DNA Dizilerindeki Düzenleyici Motifler

- Ayrıntılı arama ile nasıl bulunur?
- The Motif Finding Problem

```

CGGGGTTATCCAGCTGGGTGCTCACATTCCTT...
TTTGAGGGTGCCCAATAAGTSCACTCCAAAGCGGACAAA
GGATGACCTGATGCCCTTGACGACCTA...
AAGGAGACACCCAGGAGCGCCTTTGCTGG...
AATTTTCTAAAAAGATTATAATGTCGGTCCCTGGAACTTC
CTGCTGTACAACTGAGATCATGCTGCATGCCAATTCACAC
TACATGATCTTTTGATGCGCACTTGGATGAGGGAATGATGC

```

(a) Superposition of the seven highlighted 8-mers from figure 4.2 (d).

Alignment		A	T	C	C	A	G	C	T
		G	G	G	C	A	A	C	T
		A	T	G	G	A	A	C	C
		A	T	G	G	A	A	C	C
		A	T	G	G	A	A	C	C
		A	T	G	G	A	A	C	C
		A	T	G	G	A	A	C	C
Profile	A	5	1	0	0	5	5	0	0
	T	1	5	0	0	0	1	1	6
	G	1	1	6	3	0	1	0	0
	C	0	0	1	4	2	0	6	1
Consensus		A	T	G	C	A	A	C	T

(b) The alignment matrix, profile matrix and consensus string formed from the 8-mers starting at positions  $s = (8, 19, 3, 5, 31, 27, 15)$  in figure 4.2 (d).

**Figure 4.3** From DNA sample, to alignment matrix, to profile, and, finally, to consensus string. If  $s = (8, 19, 3, 5, 31, 27, 15)$  is an array of starting positions for 8-mers in figure 4.2 (d), then  $Score(s) = 5 + 5 + 6 + 4 + 5 + 5 + 6 = 42$ .