BMM 101
Biyomedikal Mühendisliğine Giriş

Biyoenformatik ve Protein Tasarımı

Dr. Ersin Emre Ören

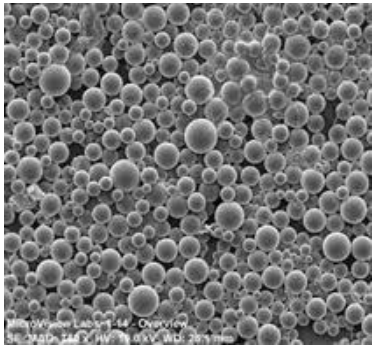
Biyonanodesign Laboratuvarı
Biyomedikal Mühendisliği Bölümü
Malzeme Bilimi ve Nanoteknoloji Mühendisliği Bölümü

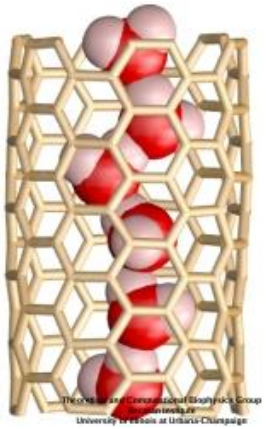
TOBB Ekonomi ve Teknoloji Üniversitesi
Ankara - TÜRKİYE

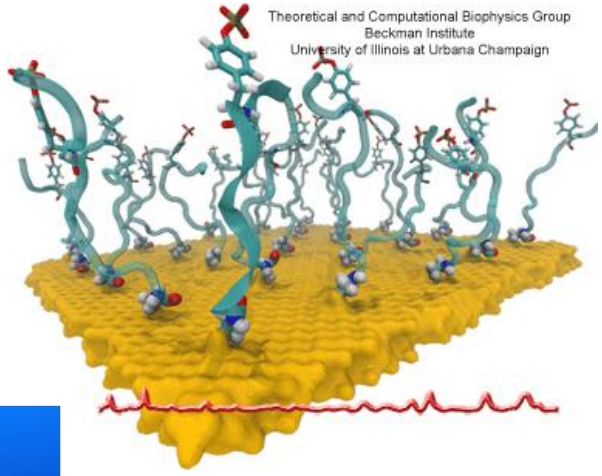
eeoren@etu.edu.tr
http://eeoren.etu.edu.tr

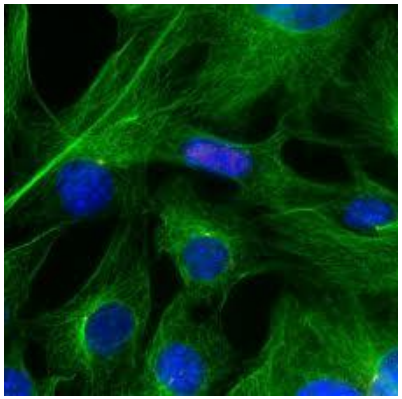
Bionanodesign Laboratory

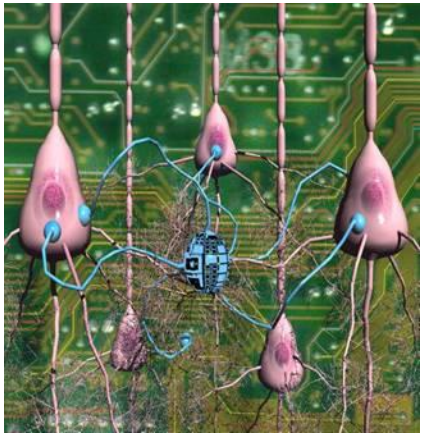
BİYOMEDİKAL MÜHENDİSLİĞİ - BİYONANOTEKNOLOJİ

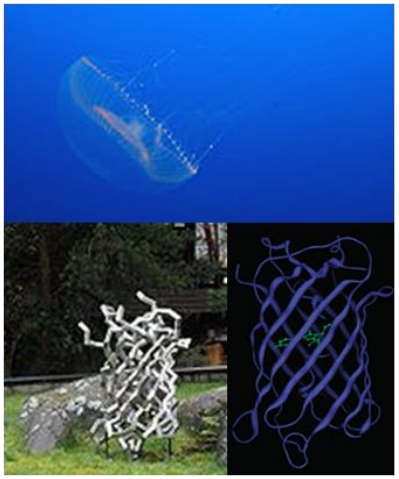


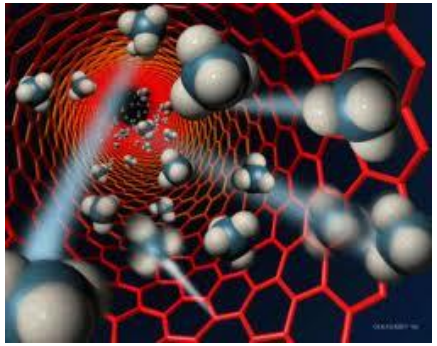


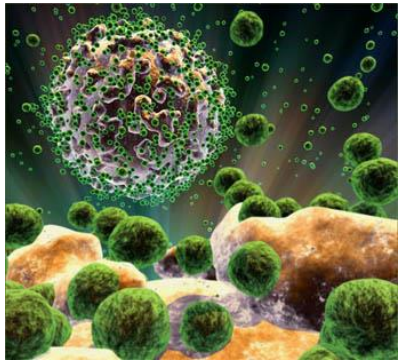


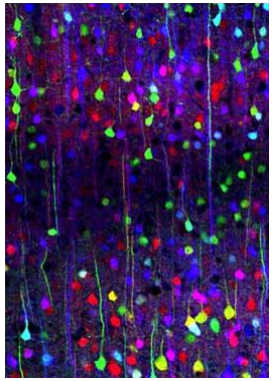


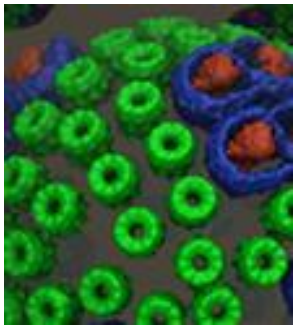




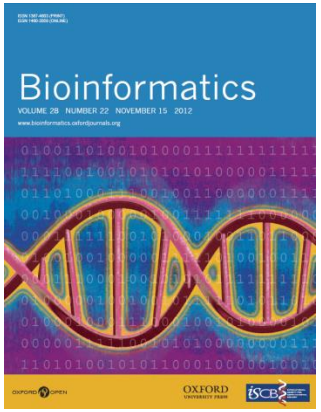


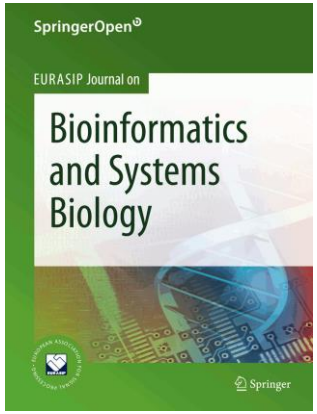






Biyoenformatik (Bioinformatics): Biyolojik verilerin toplanması, depolanması ve bu verilerin analizi için algoritmalar ve istatistik modeller geliştiren bir bilim dalıdır.



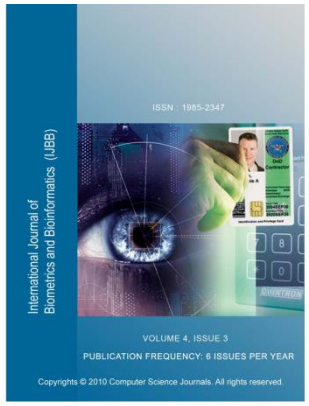


Algoritmalar, veri tabanları, web teknolojileri, yapay zeka, bilgi teknolojileri, yapısal biyoloji, modelleme ve simulasyon, istatistik.

BMM Ders Programı ve Biyoenformatik:

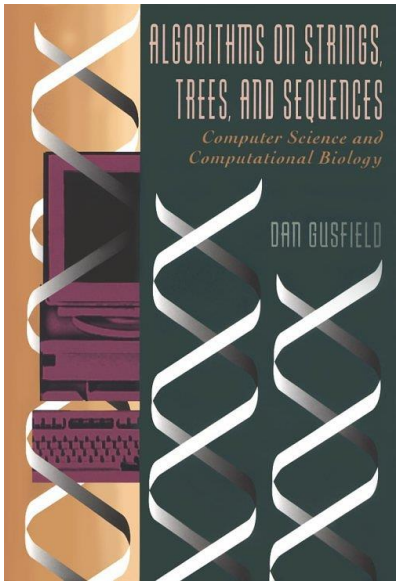
BMM 310 Biyomedikal Mühendisliğinde Sayısal Yöntemler } **Zorunlu**

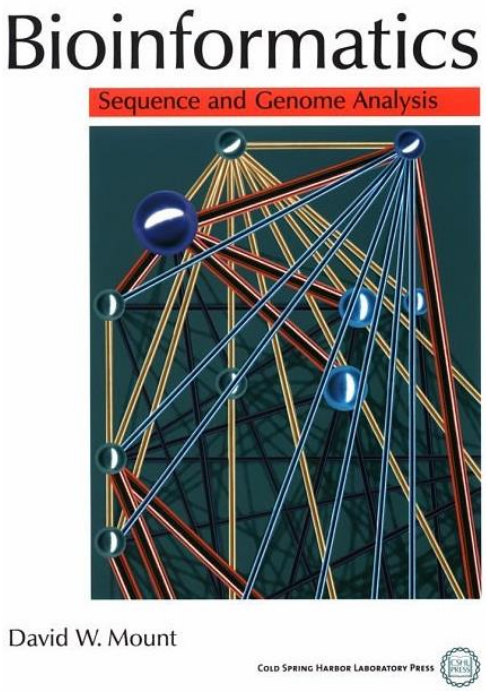
BMM 420 BMM'de Bilgisayar Uygulamaları
BMM 421 Biyoenformatik
BMM 422 Biyoistatistik
BMM 423 Biyobenzeşim ve Biyotasarım
BMM 424 Bilgisayar Destekli İlaç Tasarımı
} **Seçmeli**

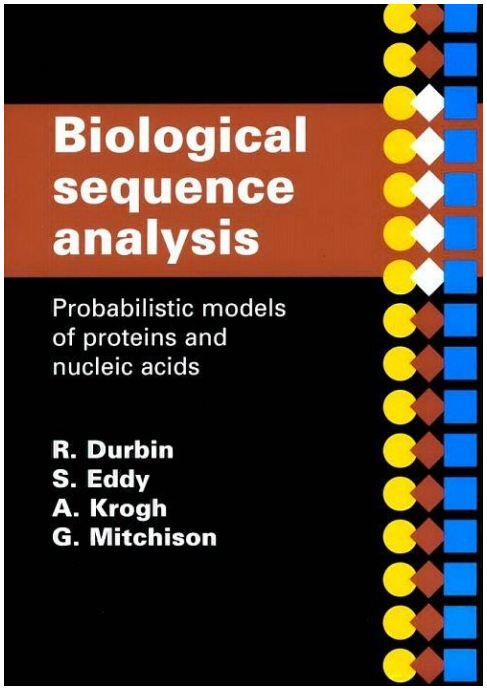


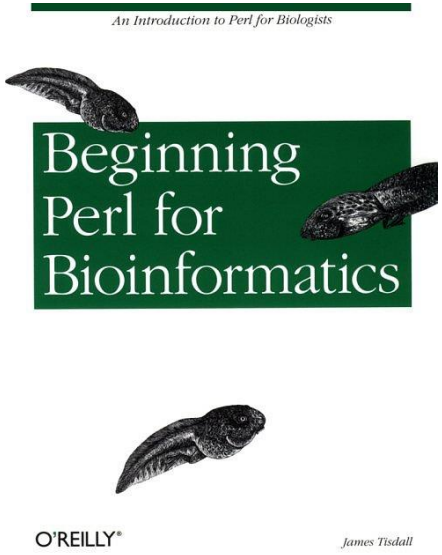
Referanslar:

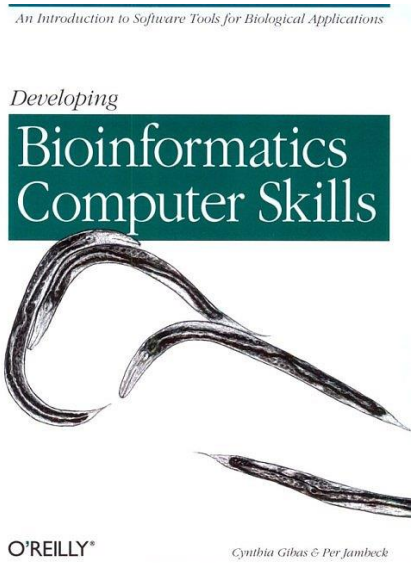
Biyoenformatik Kitapları:

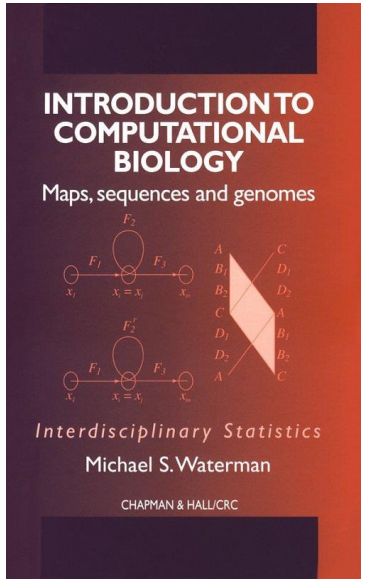












Diğer Referanslar:

- ISCB: http://www.iscb.org/
- NBCI: http://ncbi.nlm.nih.gov/
- http://www.bioinformatics.org/
- Journals (Bioinformatics)
- Conferences (ISMB, RECOMB, PSB…)

Online Referanslar:

Biyoenformatik (Bioinformatics)

BLAST	http://www.ncbi.nlm.nih.gov/BLAST/
	http://www.ebi.ac.uk/blastall
FASTA	http://www.fasta.bioch.virginia.edu
	http://www.ebi.ac.uk/fasta33/
DOTLET	http://www.isrec.isb-sib.ch/java/dotlet/Dotlet.html
DIALIN	http://bibiserv.techfak.uni-bielefeld.de/dialign/
CLUSTAL	http://www.ebi.ac.uk/clustalw/
TCOFFEE	http://igs-server.cnrs-mrs.fr/Tcoffee/tcoffee_cgi/index.cgi
GENEBEE	http://www.genebee.msu.su/services/malign_reduced.html
MAVID	http://baboon.math.berkeley.edu/mavid/

Veri Tabanları (Databases)

PDB	http://www.pdb.org	Yapı (Structure)
EBI-MSD	http://www.ebi.ac.uk/msd/	
SWISS-PROT	http://us.expasy.org/	Sekans (Sequence)
TrEMBL	http://www.ebi.ac.uk/trembl/	
PIR-PSD	http://pir.georgetown.edu/	
UniProt	http://www.expasy.uniprot.org/	

Online Referanslar:

Görüntüleme (Visualization)

WebMol	http://www.cmpharm.ucsf.edu/cgi-bin/webmol.pl
Protein Explorer	http://molvis.sdsc.edu/protexpl/frntdoor.htm
VMD	http://www.ks.uiuc.edu/Research/vmd
Kinemage	http://kinemage.biochem.duke.edu
MolSoft	http://www.molsoft.com
DeepView	http://ca.expasy.org/spdbv/

Moleküler Dinamik (Molecular Dynamics)

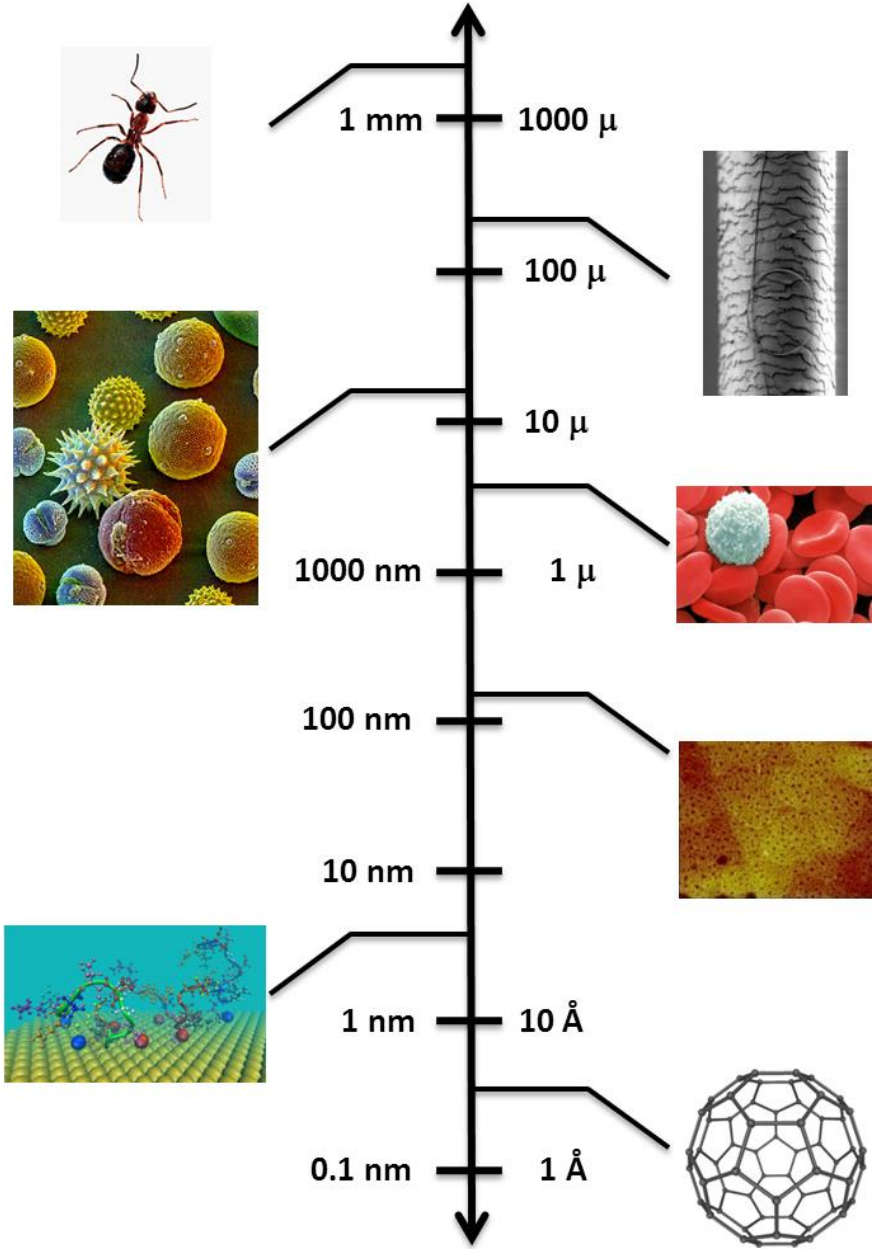
NAMD	http://www.ks.uiuc.edu/Research/namd/
GROMACS	http://www.gromacs.org/
Accelrys Insight	http://www.accelrys.com/products/insight/
UCSF Chimera	http://www.cgl.ucsf.edu/chimera/
CHARMM	http://www.ch.embnet.org/MD_tutorial/index.html
AMBER	http://amber.scripps.edu/
CHIME	http://www.umass.edu/microbio/chime/index.html
HyperChem	http://www.hyper.com/
TINKER	http://dasher.wustl.edu/tinker/

Nanoteknoloji, malzemelerin yaklaşık 1 ila 100 nanometrelik bir ölçekte gösterdikleri tamamen farklı özellikleri kullanarak yeni malzeme yapıları, cihazlar ve sistemler tasarlamak ve üretmek olarak tanımlanabilir.

Yani malzemelerin boyut ve şekillerinin nanometre (10^{-9} m) ölçeğinde kontrol edilmesi ile yeni aygıtların üretildiği disiplinler arası bir bilim ve mühendislik dalıdır.

Biyonanoteknoloji, nanoboyuttaki ilke ve tekniklerin, cansız maddeler ve biyosistemleri anlamak ve bunları ilaç sentezi, beynin işleyişinin anlaşılması, vücut parçası üretimi, tanı ve tedavi yöntemleri gibi tıbbi amaçlar için kullanılmasıdır.

Nanoteknolojinin, biyomedikal mühendisliği, biyoloji ve bilgi teknolojisileri ile entegrasyonu günümüzün ve yakın geleceğin en önemli araştırma konuları arasında gelmektedir.



Moleküler Biyolojinin Kısa Tarihi:

1665 Mikroskopik biyolojinin başlaması: Robert Hooke (1635-1703) organizmaların hücrelerden yapıldığını buldu

1865 Gregor Mendel, bezelyeler üzerindeki çalışmaları sonucunda kalıtımın temel ilkelerini geliştirdi. (baskın (dominant) – çekinik (recessive))

1869 Johann Friedrich Miescher DNA'yı bularak nüklein (nuclein) olarak adlandırdı.

1881 Edward Zacharias kromozomların nükleinden oluştuğunu gösterdi.

1899 Richard Altmann nüklein adını nükleik asit olarak değiştirdi.

1900 20 amino asitin tamamının kimyasal yapısı bulundu.

1902 Emil Hermann Fischer amino asitlerin bir araya gelerek proteinleri oluşturduğunu göstererek Nobel ödülünü kazandı.

1911 Thomas Hunt Morgan gen ve kromozomların kalıtım birimleri olduğunu gösterdi.

1911 Pheobus Aaron Theodore Lerene RNA'yı buldu.

R. Hooke

G. Mendel

J. F. Miescher

T. H. Morgan

Moleküler Biyolojinin Kısa Tarihi:

1941 George Beadle ve Edward Tatum genlerin proteinlerin üretilmesindeki rollerini ortaya çıkardılar.

1950 Edwin Chargaff Sitozin ile Guaninin ve Adeninin ile de Timinin eşleştiğini gösterdi.

1952 Alfred Hershey ve Martha Chase deneysel olarak DNA'nın genleri oluşturduğunu ve kalıtımdan sorumlu olduğunu gösterdiler.

1952-3 James D. Watson ve Francis H. C. Crick DNA'nın ikili sarmal adını verdikleri yapısını buldular.

1956 George Emil Palade showed the site of enzymes manufacturing in the cytoplasm is made on RNA organelles called ribosomes.

1970 Howard Temin ve David Baltimore ilk kısıtlayıcı enzimi izole ettiler.

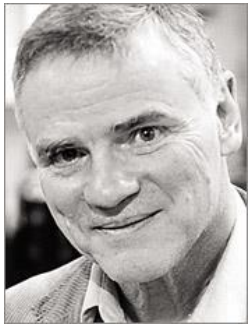
1986 Leroy Hood, otomatik sekanslama makinesini geliştirdi.

1986 Human Genome Initiative tanıtıldı.

1990 15 yıllık Human Genome Project ABD Kongresince başlatıldı.

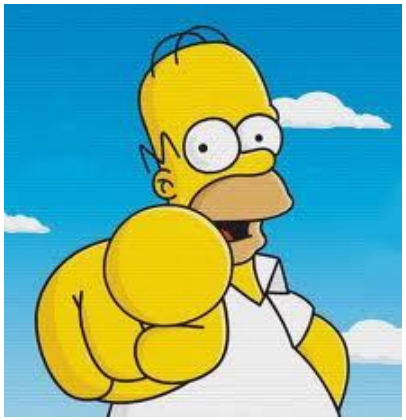
M. C. Chase

J. D. Watson & F. H. C. Crick



L. Hood

Genom Boyutu (Genome Size)



Homer Simpson
3.2 Gb

Marbled lungfish
130 Gb

Paris japonica
150 Gb

Polychaos dubium
670 Gb

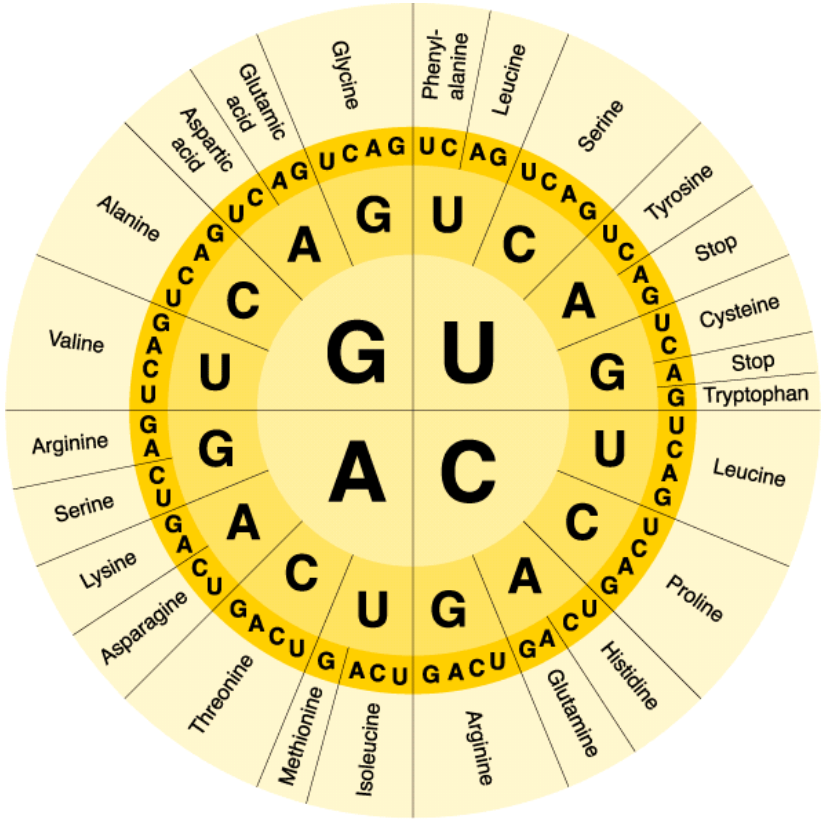
Dünya Gezegenindeki bütün hayat 3 kritik moleküle bağlıdır

Biyolojik Yapı Taşları: DNA, RNA, Protein

Yaşamın Kodu: DNA, RNA (A, T(U), C, and G)

Yaşamın Alfabesi: Amino Asitler (20)

Yaşamın Dili: Proteinler



62 aa'lık bir Protein kaç farklı şekilde oluşturulabilir?

$$20^{62} \cong 10^{80}$$

Protein, Yunanca “protas” kelimesinden gelir ve anlamı “en önemli”dir. "of primary importance"

Hayatın Makineleri: Canlı organizmaları karakterize eden her özellik proteinlerce belirlenir...

Protein foksiyonları

Kimyasal reaksiyonları Kataliz ederler (enzimler)

Birçok parçacığın taşınımı ve depolanmasında (e^- to makromoleküller)

Kimyasal enerjiyi mekanik enerjiye çevirirler (kas proteinleri actin ve myosin)

Organizmaları savunurlar (antibodies)

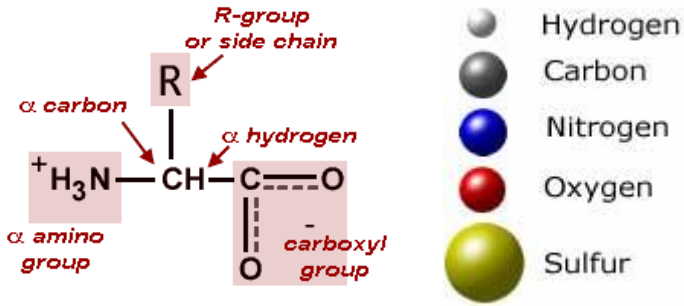
Gen expresyonunu kontrol ederler

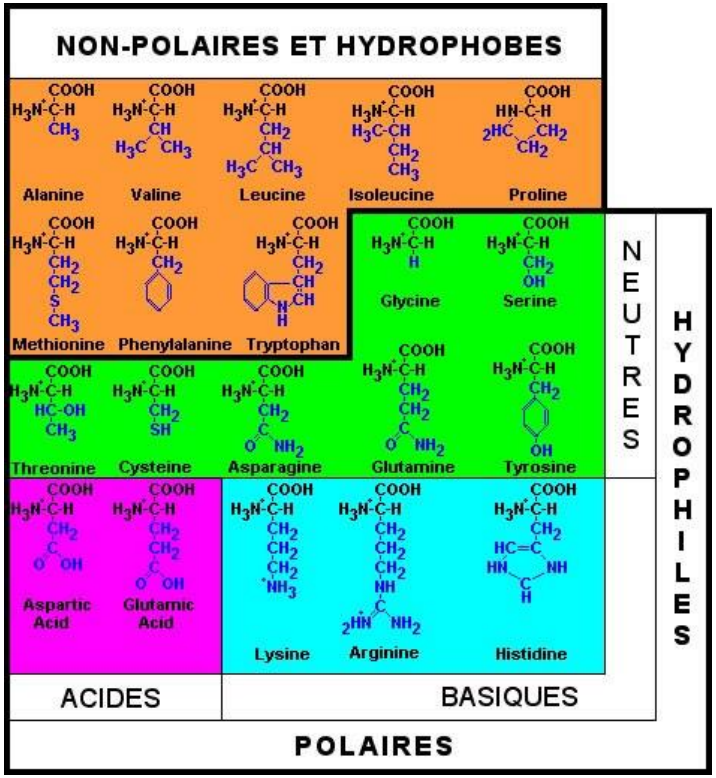
Duyular için gereklidir (görme, duyma ve diğerleri)

Yapı oluğumunda görev alırlar (saç, tırnak, diş, kemik)

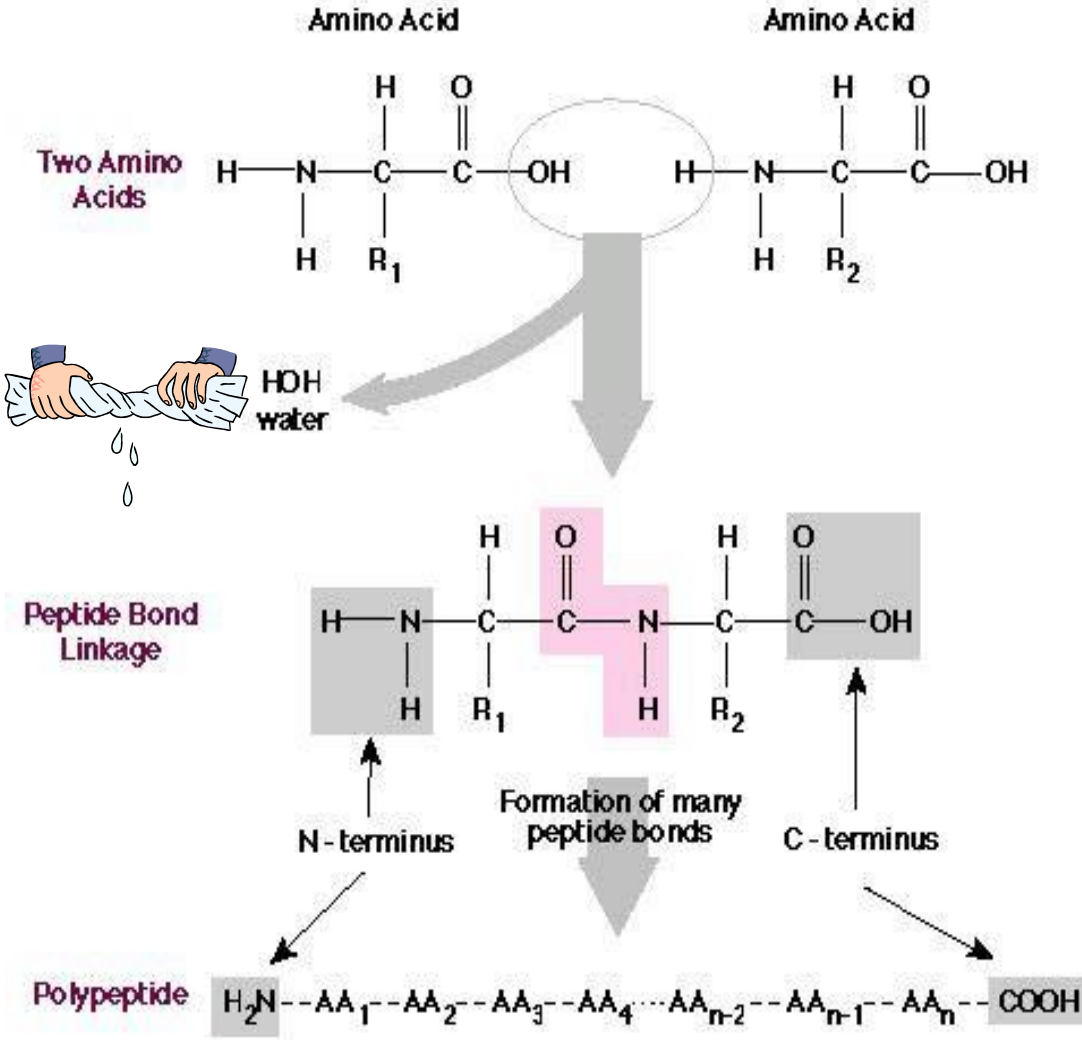
…

Amino Asitler:





Peptid Bağları:

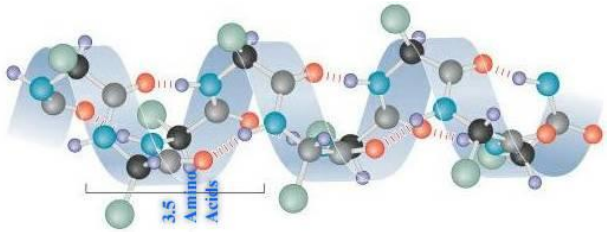


Birincil yapı: amino asit dizilimi…

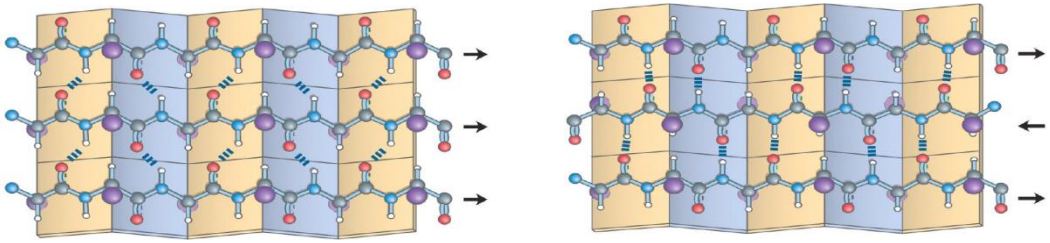
```
KVFGRCELAA AMKRHGLDNY RGYSLGNWVC AAKFESNFNT QATNRNTDGS
TDYGILQINS RWWCNDGRTP GSRNLCNIPC SALLSSDITA SVNCAKKIVS
DGNGMNAWVA WRNRCKGTDV
```

İkincil yapı: genel 3-boyutlu yapı tanımı...

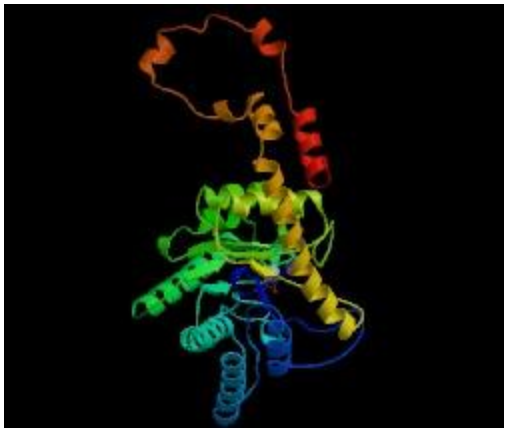
α-Helix



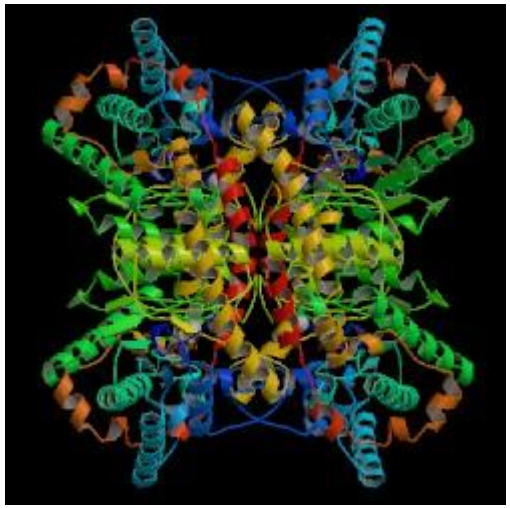
β-Sheets



Üçüncül yapı: proteinin bütün yapısı atomların pozisyonları…

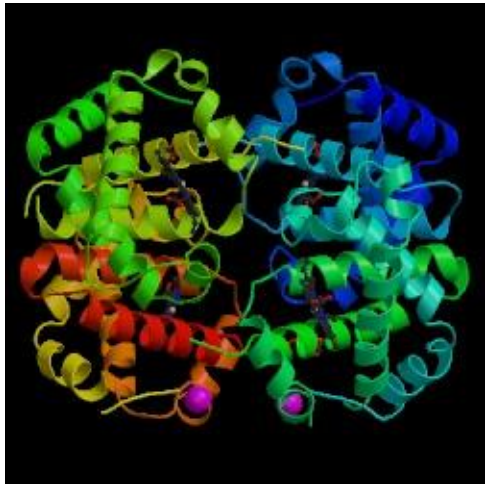
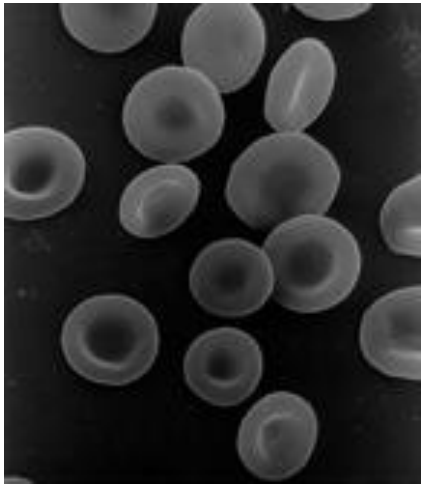


Dördüncül yapı: birden fazla proteinin oluşturduğu sistemin tamamının yapısı…

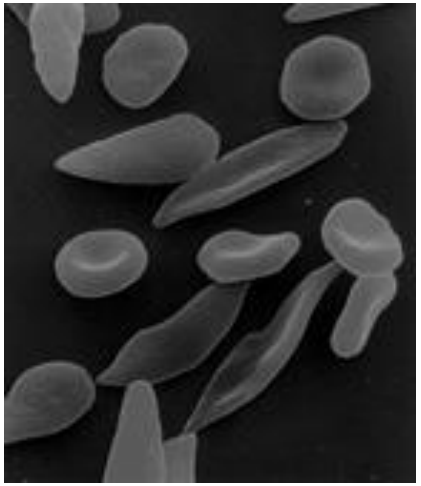


Orak Hücreli Anemi (Sickle Cell Anemia)

Hemoglobin: kanda solunum organından dokulara oksijen, dokulardan solunum organına ise karbondioksit ve proton taşıyan proteindir.

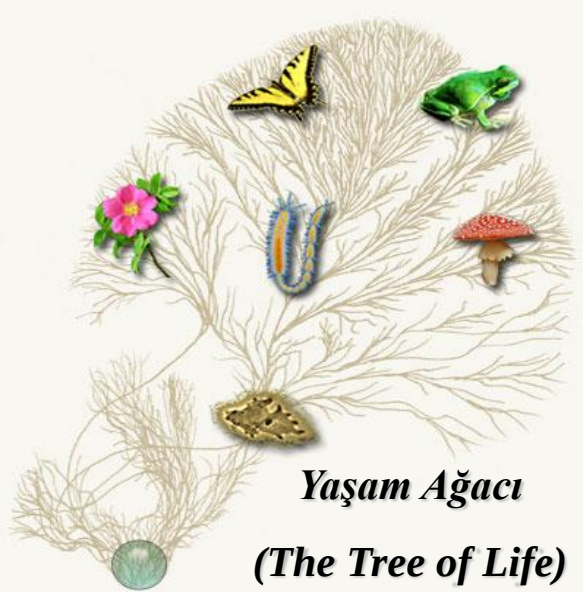
Sağlıklı Birey

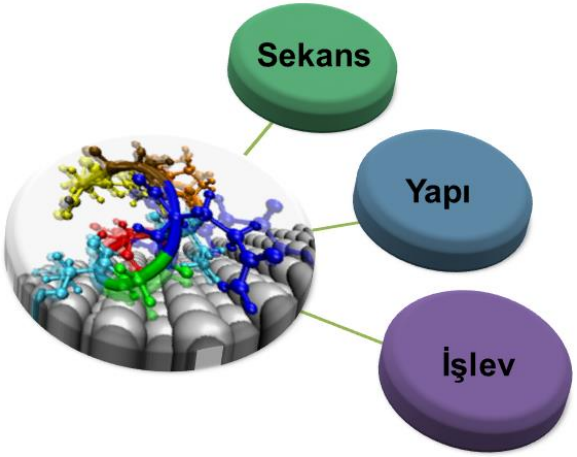


Hasta Birey

```
>gi|4504349|ref|NP_000509.1| beta globin [Homo sapiens]
MVHLTPE (V) EKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVK
AHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPV
QAAYQKVVAGVANALAHKYH
```

DNA sekansında T yerine A gelmesine neden olan bir mutasyon sonucu, üretilen hemoglobinde glutamik asit (GAG) yerine valin (GTG) bulunur…





Biology bilgi tabanlı bir bilimdir. Biyolojide verilerin analizi ve yorumlanması yeni gelen verinin daha önceki bilinen veriler ile karşılaştırılarak yapılır.

Bunu genellikle, bilinmeyen bir protein dizisinin daha önceden yapısını/işlevini bildiğimiz bir protein ile benzerliği olup olmadığına bakarak yaparız.

Biyolojideki en önemli ve aşılması gereken problemlerden birtanesi, eldeki bilgiyi en hızlı ve doğru şekilde nasıl kullanacağımız sorusudur.

Bioenformatik: biyolojik verilerin toplanması ve depolanması ve bu verilerin analizi için algoritmalar ve istatistik modeller geliştiren bir bilim dalıdır...

Farklı Alanlardaki Biyoenformatik Uygulamaları

Moleküler tıp

Kişiselleştirilmiş tıp

Önleyici tıp

Gen tedavisi

İlaç tasarımı

Microbial genome applications

Atık temizleme

İklim değişimi çalışmaları

Alternatif enerji kaynakları

Biyoteknoloji

Antibiyotik direnci

Adli analizler

Protein yapı analizi

Evrimsel çalışmalar

Tahıl geliştirme

Böcek resistansı

Besin kalitesinin arttırılması

Kuraklığa dayanıklı bitki geliştirme

Veterinerlik

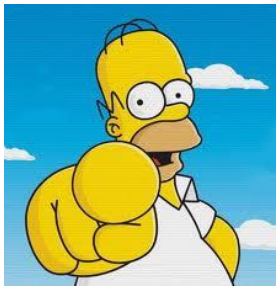
Karşılaştırmalı bilimler

Dünya Tarihinin En Büyük Doğal Felaketleri

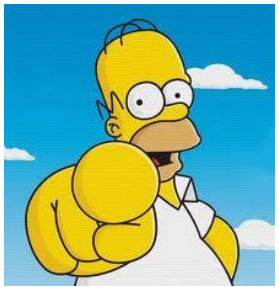
Rank	Death toll (estimate)	Event	Location	Date
1.	300,000,000 approx.	Smallpox	Worldwide	1900 to eradication.[7] Declared eradicated May 8, 1980.[8]
2.	200,000,000 ?	Measles	Worldwide	last 150 years[9]
3.	100,000,000 approx.	Black Death	Asia, Europe, Africa	1300s–1720s
4.	80,000,000–250,000,000	Malaria	Worldwide	20th century – present
5.	50,000,000–100,000,000	Spanish Flu	Worldwide	1918–1920
6.	40,000,000–100,000,000	Plague of Justinian	Asia, Europe, Africa	540–590
7.	40,000,000–100,000,000	Tuberculosis	Worldwide	20th century – present[9]
8.	30,000,000[10]	AIDS pandemic	Worldwide	1981–present
9.	12,000,000 ?	Third Pandemic of Bubonic Plague	Worldwide	1850s–1950s
10.	5,000,000	Antonine Plague	Roman Empire	165–180
11.	4,000,000	Asian Flu	Worldwide	1956–1958
12.	250,000 or more annually	Seasonal influenza	Worldwide	As of April 2009[11]

Grip (Influenza (flu)): kuşları ve memelileri etkiler.

Influenzavirus A: İnsanlardaki en etkili virustür, çok hızlı mutasyona uğrar ve yayılır.

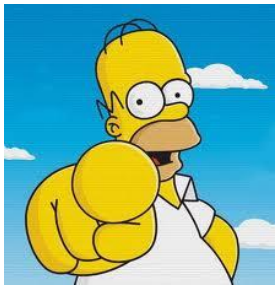


Influenzavirus B:



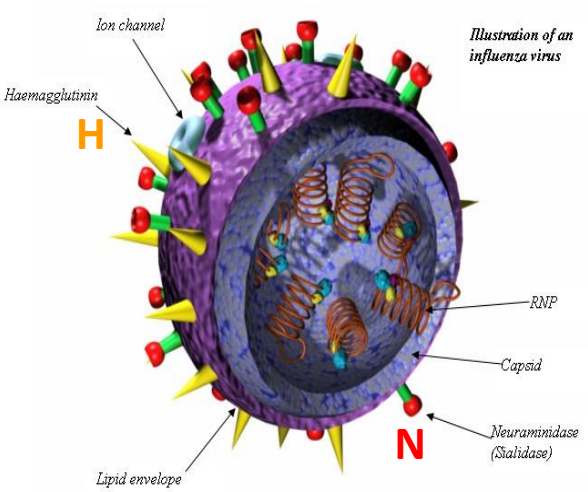
- Ölümler
- İlaç ve hasta bakım masrafları

Influenzavirus C:



- İş saati kaybı. ABD'de yıllık 75 milyon iş günü kaybı.

İlaç Tasarımı: Grip



Grip virüsleri, üzerlerinde taşıdıkları antijenlere (protein) göre adlandırılırlar.

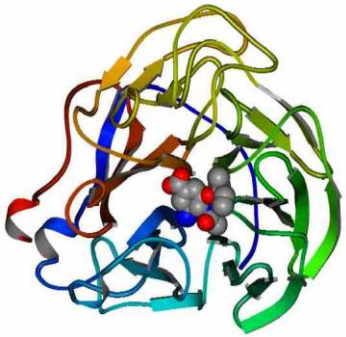
Grip virüslerinde 2 ana antijen bulunur = **N** (Neuraminidase) and **H** (Hemagglutinin).

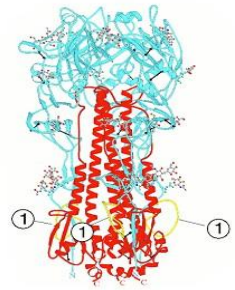
Farklı virüslerdeki **N** ve **H** antijenlerinin yapıları farklıdır.

Bilinen 9 **N** ve 16 **H** tipi mevcuttur.

Grip virüsü antijenlerinin rolü

N antijeni virüsün konak hücreden ayrılmasını ve başka hücrelere yayılarak çoğalmasını sağlar.



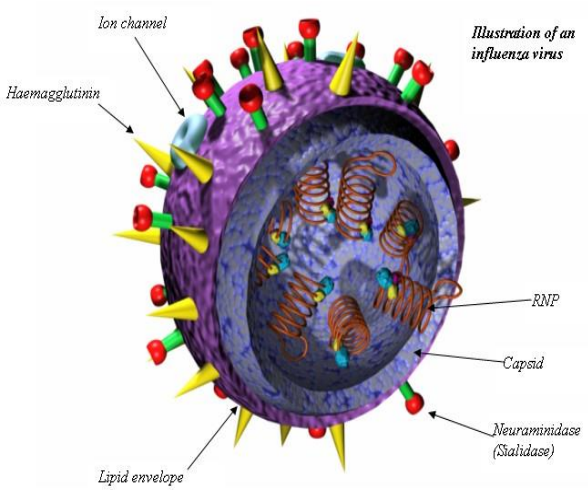


H antijeni ise bir anahtar gibi virüsün hücre duvarını aşarak hücreye girmesini sağlar; böylece virüs hücre içerisinde çoğalabilir.

Kuş Gribinde **H** antijeni kuşun bağırsak hücrelerini etkilerken; İnsan Gribindeki **H** antijeni ise insan akciğer hücrelerini etkiler.

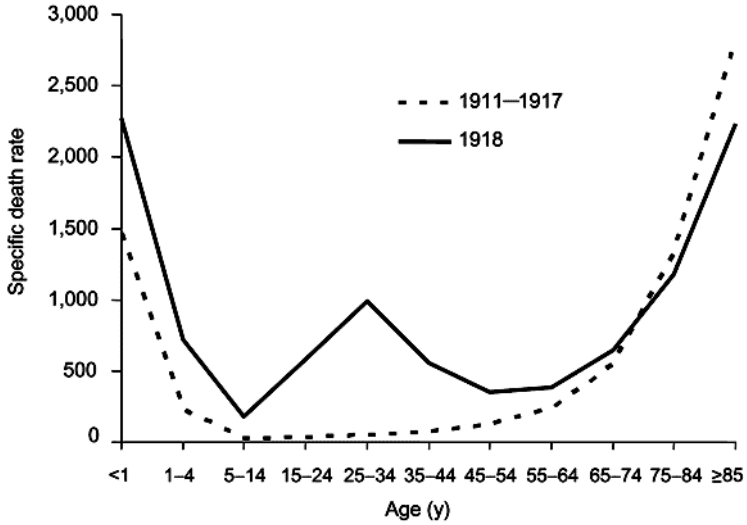
İlaç Tasarımı: Grip

H1N1: İspanyol Gribi - 1918, Domuz Gribi - 2009
H2N2: Asya Gribi - 1957
H3N2: Hong Kong Gribi - 1968
H5N1: Kuş Gribi - 2004
H1N2: İnsan, kuş ve domuzlardaki bölgesel salgınlar



İspanyol Gribi (Spanish Flu (1918)):

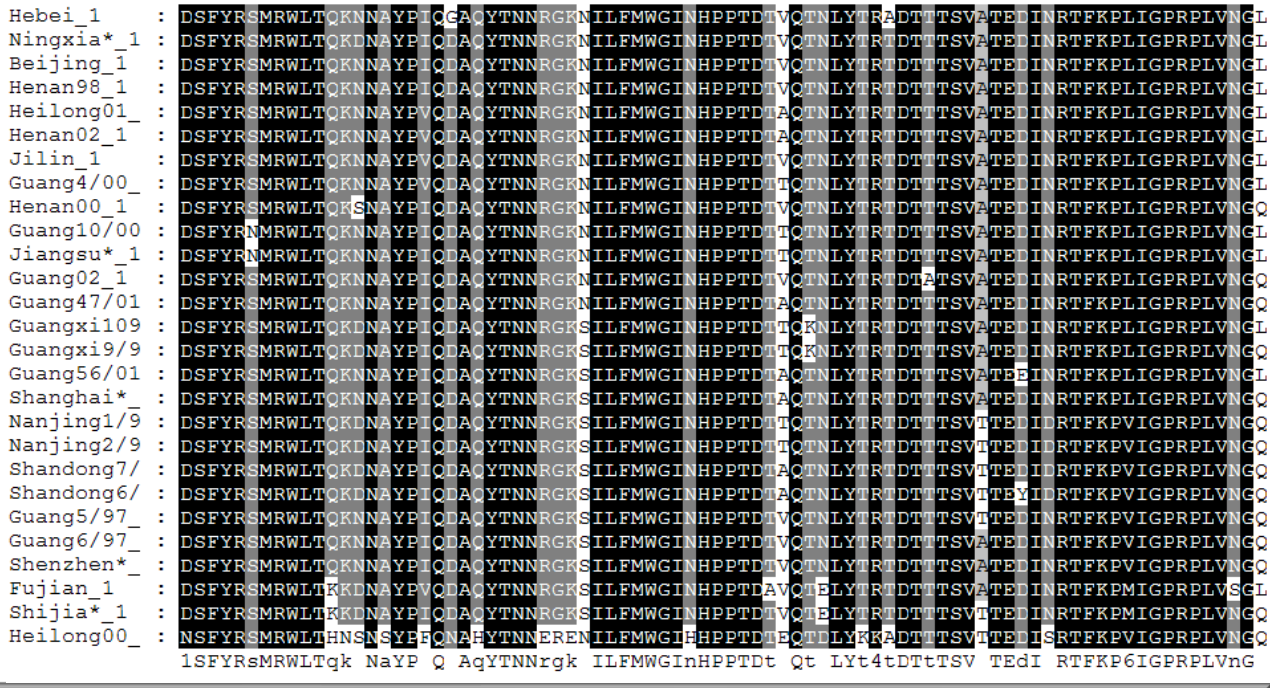
- 50 – 100 Milyon insan öldü...
- İnsanlık tarihinin en ölümcül doğal felaketi...



A/chicken/Fujian/411/2002/ (H3N2)

İlaç Tasarımı: Kuş Gribi

- İlaç tasarımında yapmamız gereken en önemli şey virüslerin işlevlerini durdurmak için virüs proteinlerinin aktif bölgelerine bağlanacak moleküller bulmaktır.
- Grip için ise amaç: grip virüsünün yüzeyindeki aktif bölgelere bağlanarak enzimin atkisini düşürmektir (neuraminidase, subtype N1: *virüsün çoğalarak başka hücreleri etkilemesine yardımcı olan bir enzimdir*).
- Biyoenformatik kullanılarak 300,000 ilaç molekülü kuş giribi virüsü H5N1'a karşı taranmış ve aralarından seçilen potensiyel moleküller virüsün aktif bölgesine bağlanma enerjilerine göre sıralanmıştır.
- Dünya çapındaki bir salgın için en büyük korku virüsün evrimleşmesi (+ mutasyona uğraması) sırasında geliştirilen bu ilaçlara karşı direnç kazanmasıdır.



İlaç Tasarımı: Kuş Gribi, Domuz Gribi

Gelecekte grip virüslerinin nasıl mutasyona uğrayacağını tahmin edebilirmiyiz ?

National Institute of Allergy and Infectious Diseases (NIAID)

http://www.niaid.nih.gov/dmid/genomes/mscs/influenza.htm

Influenza Genome Sequencing Project kapsamında araştırmacılar grip genomu hakkındaki bilgimizi arttırarak grip virüslerinin nasıl evrimleştiği, yayıldığı ve hastalığa neden olduğunu anlamaya çalışırlar.

H5N1 virüsünün hangi mutasyonlara uğraması durumunda insandan insana bulaşmaya başlayacağını bulabilirmiyiz?

H5N1 ⇒ Human Flu Viruses

Sekans Eşleme (Sequence Alignment)

Soru: Verilen herhangi iki protein sekansı birbiri ile ilişkili mi?
İki sekansı karşılaştırarak benzerlikleri bul…

Örnek: ***ALIGNMENT*** and ***LIGAMENT***

```
A L I G N M E N T
  | | |   | | | |
- L I G A M E N T
```

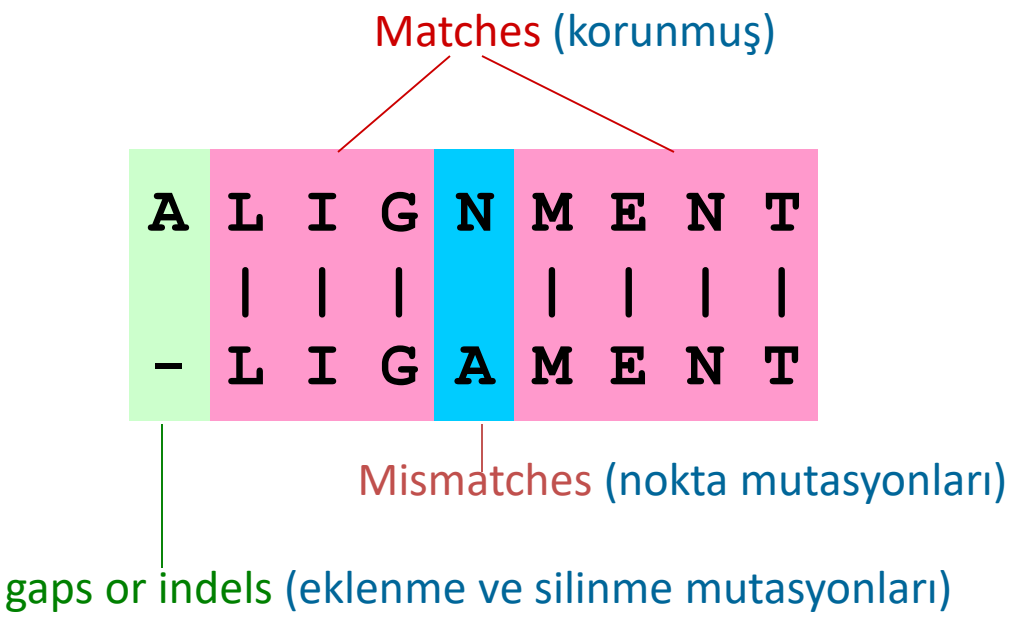
Benzerlik (Similarity) ve *Homoloji (Homology):* farklı anlamları olmasına rağmen, genelde aynı anlamda kullanılır.

Benzerlik (Similarity): iki veya daha fazla sekansın birbirleri ile olan ilişkisini tanımlamak için kullanılır.
Yüksek benzerlik biyolojik yapı ve/veya fonksiyonda ortak yönler olduğunu ifade eder.

Homoloji (Homology): ise evrimsel ortak bir geçmiş olduğu anlamına gelir.

Sekans Eşlemelerinin Yorumlanması:

Eğer, evrimsel olarak ortak bir ata paylaşan iki sekans arasında eşleme yapılırsa



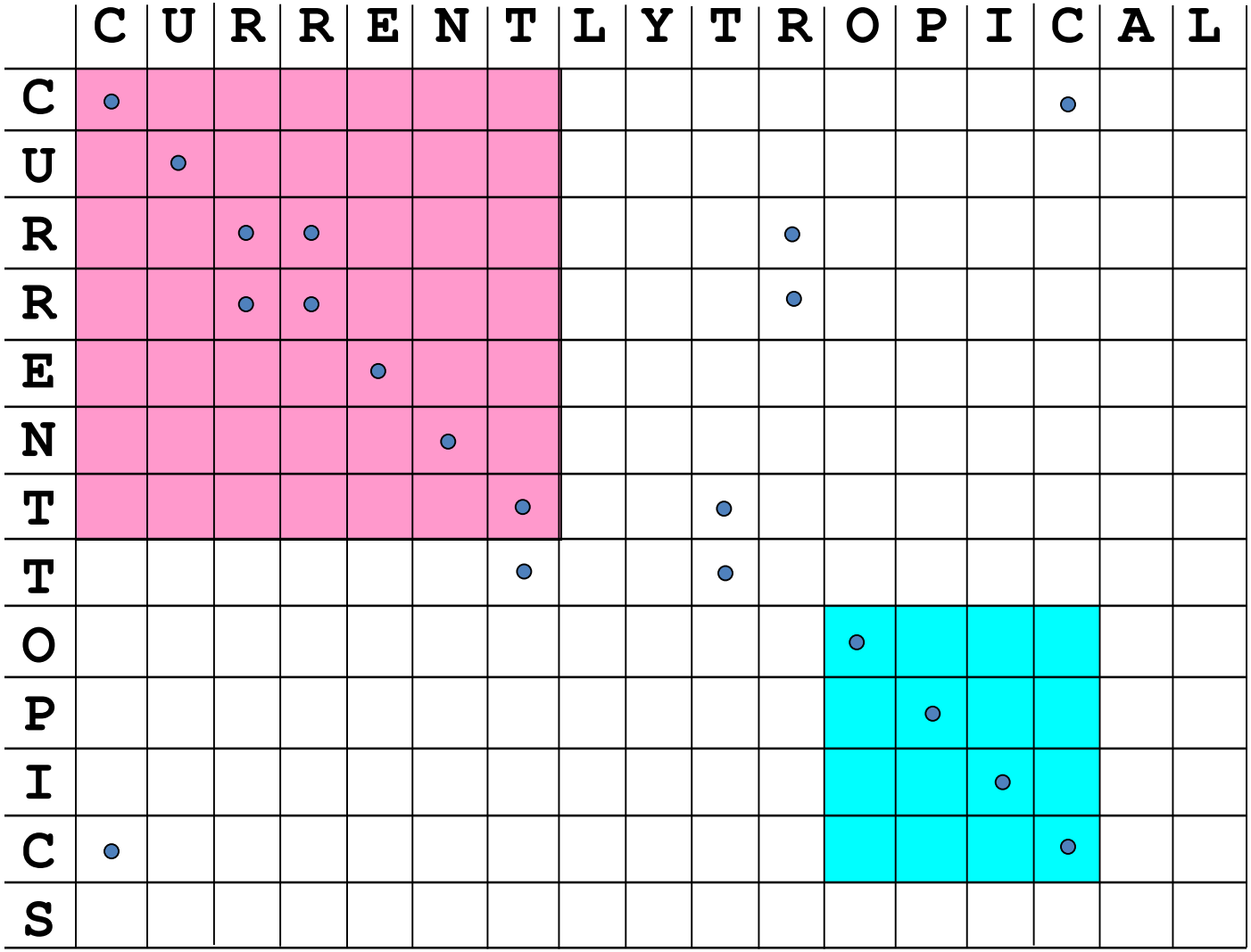
Temel Sorular

Verilen iki veya daha fazla sekans arasındaki en iyi eşlemeyi nasıl buluruz?

Bu eşlemenin iyi olup olmadığını nasıl ölçeriz?

Nokta Grafikleri (Dot Plots):

En temel ve basit yöntemdir ve görsel olarak ortak olan bölgelerin belirlenmesinde kullanılır.



- bölgesel eşlenmeler
- eklemeler
- silinmeler
- tekrarlar
- ters tekrarlar

AMA

Eşlenmenin kalitesi hakkında istatistiksel bir ölçüm vermez.

Nokta Grafiklerinde Arka Plan Gürültüsü

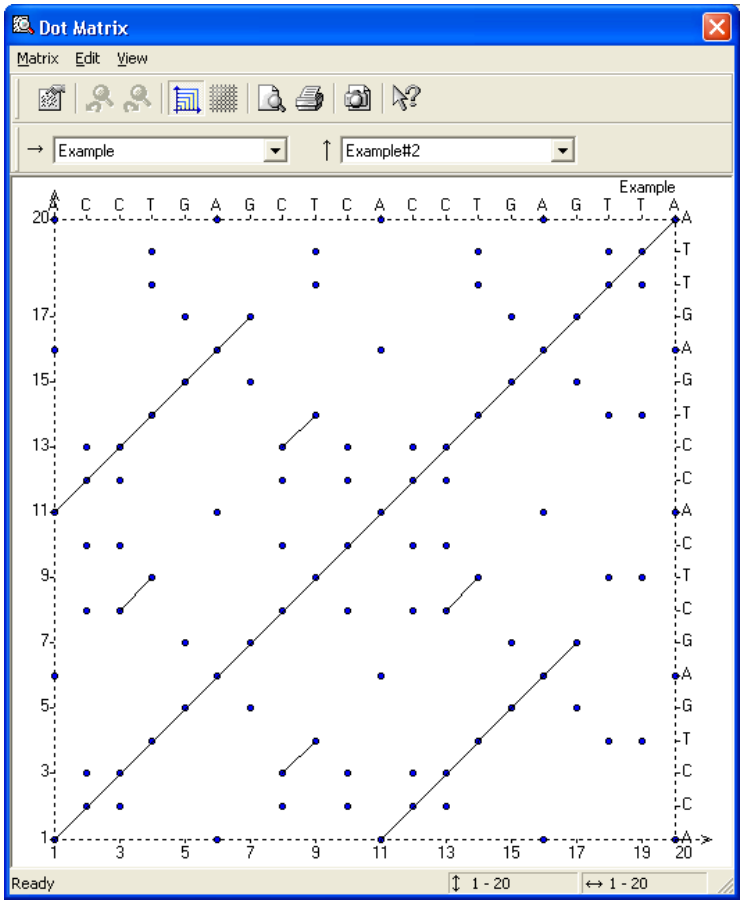
Nucleic Acids (DNA, RNA)

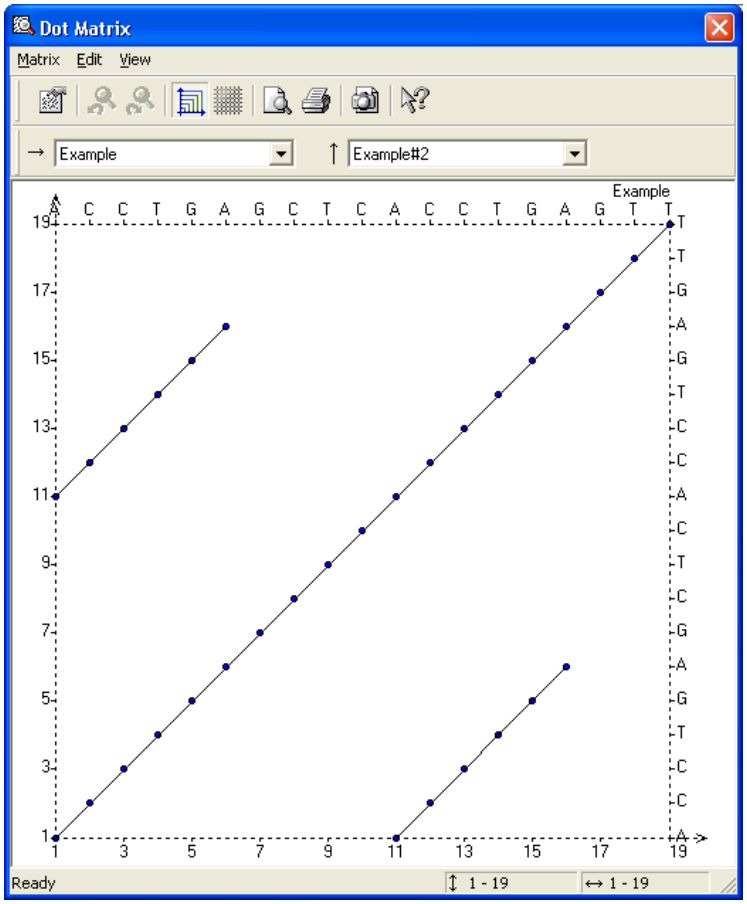
Nükleik asitler % 25 olasılıkla rastgele eşleşebilirler.

Arka Plan Gürültüsünün Azaltılması

Arka arkaya eşleşen nükleik asitlerin sayısı için belirli bir eşik değer seçilir.

Self alignment of ACCTGAGCTCACCTGAGTTA





Hangi eşleme daha iyi?

Eşleme 1

```
A - C - S S - A P T
|   |   |       | |
A T C S S A T - P T
```

Eşleme 2

```
A - C S S - A P T
|   | | |     | |
A T C S S A T P T
```

Olası ölçme yöntemleri:

uygunluk (match): +2
uyumsuzluk (mismatch): -1
ekleme/silme (indel) -2

Eşleme 1: 5 * 2 – 1 * 1 – 4 * 2 = 10 – 1 – 8 = 1

Eşleme 2: 6 * 2 – 1 * 1 – 2 * 2 = 12 – 1 – 4 = 7

Bu yöntem ile amino asitleri birbirlerinden ayrılamaz…

Skorlama Matrisleri (Substitution) Matrices (SM): Gözlemsel ağırlık değerleri (weighting factors)…

SM içerisindeki sayılar amino asitlerin birbirleri yerine proteinlerin yapı ve fonksiyolarında bozulma yapmadan geçebilme özelliklerini belirler.

SM sekans eşleşmelerinin değerlendirilmesinde ve en iyi eşleşmeyi bulmada en temel faktördür.

- ❖ Gözlenen amino asit dağılımları
- ❖ Türler arasındaki evrimsel bağlantı
- ❖ Amino asit özellikleri
  - ✓ Yük (Charge)
  - ✓ Boyut (Size)
  - ✓ Su severlik (Hydrophilicity)
  - ✓ Polarlık (Polarity)
  - ✓ Asitlik (Acidity)

	A	C	D	E	F	G	H →
A	4	0	-2	-1	-2	0	-2
C	0	9	-3	-4	-2	-3	-3
D	-2	-3	6	2	-3	-1	-1
E	-1	-4	2	5	-3	-2	0
F	-2	-2	-3	-3	6	-3	[illegible]
G	0	-3	-1	-2	-3	[illegible]	
H ↓	-2	-3	-1	0			

BLOSUM 62

Log Odds Ratio: Bir protein ailesi içerisinde amino asit i ve j'nin eşleşme olasılığı

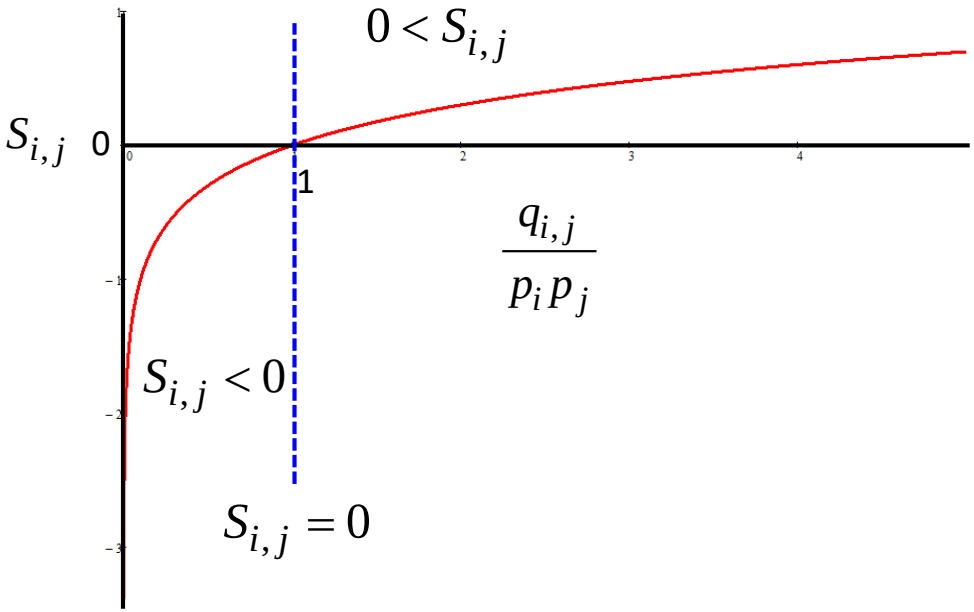
$$S_{i,j} = \log\left[\frac{q_{i,j}}{p_i p_j}\right]$$

Gözlenen olasılık (Observed probability)

Beklenen olasılık (Random probability)

Tüm proteinler içerisinde i ve j'nin bulunma olasılıkları…

Korunan ve Korunmayan Bölgeler



PAM: Point Accepted Mutation (Kabul Edilmiş Nokta Mutasyonları)

Dayhoff, M.O., Schwartz, R.M., Orcutt, B.C. A model of evolutionary change in proteins. In "Atlas of Protein Sequence and Structure" 5(3) M.O. Dayhoff (ed.), 345 - 352 (1978)

Bu matris birbirleri ile yakın ilişkili olduğu bilinen proteinler içerisindeki amino asit dağılımları gözlemlenerek ortaya çıkarılmıştır.

tüm eşleme
71 farklı protein grubu
1572 nokta mutasyonu

Bulunan sonuçlar kısa evrimsel geçmişi olan proteinler için geçerlidir.

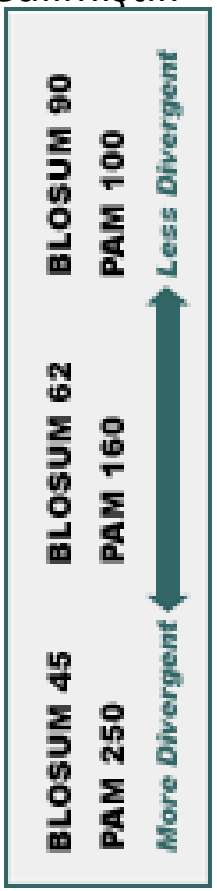
- Bütün bölgelerin eşit derecede mutasyona uğradığı kabul edilmiştir
- Bir amino asitin mutasyonunun çevresindeki amino asitlerden bağımsız olduğu kabul edilmiştir.

PAM matrisleri birbirleri ile ilgili sekansların benzerliklerini bulmada iyidir ama evrimsel geçmişi çok uzun olan proteinler arasındaki benzerlikleri bulmada kullanımı önerilmez…

BLOSUM: Block Substitution Matrices

Henikoff, S. and Henikoff, J. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA. 89(biochemistry): 10915 - 10919 (1992)

Bu skorlama matrisleri evrimsel olarak uzak proteinlerin, boşluksuz eşleştirilebilen bölümlerinin çoklu sekans eşleştirmesi yöntemi ile elde edilmiştir.



BLOSUM 62

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	*
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-3	-2	0	-2	-1	0	-4
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3	-1	0	-1	-4
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3	3	0	-1	-4
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3	4	1	-1	-4
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1	-3	-3	-2	-4
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2	0	3	-1	-4
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2	1	4	-1	-4
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3	-1	-2	-1	-4
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3	0	0	-1	-4
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3	-3	-3	-1	-4
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1	-4	-3	-1	-4
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2	0	1	-1	-4
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1	-3	-1	-1	-4
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1	-3	-3	-1	-4
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2	-2	-1	-2	-4
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2	0	0	0	-4
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0	-1	-1	0	-4
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3	-4	-3	-2	-4
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1	-3	-2	-1	-4
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4	-3	-2	-1	-4
B	-2	-1	3	4	-3	0	1	-1	0	-3	-4	0	-3	-3	-2	0	-1	-4	-3	-3	4	1	-1	-4
Z	-1	0	0	1	-3	3	4	-2	0	-3	-3	1	-1	-3	-1	0	-1	-3	-2	-2	1	4	-1	-4
X	0	-1	-1	-1	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1	-2	0	0	-2	-1	-1	-1	-1	-1	-4
*	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	1

AA not resolved

B: N or D
Z: Q or E
* : unknown

Glycine	G
Alanine	A
Valine	V
Leucine	L
Isoleucine	I
Phenylalanine	F
Proline	P
Serine	S
Threonine	T
Cysteine	C
Methionine	M
Tryptophan	W
Tyrosine	Y
Asparagine	N
Glutamine	Q
Aspartic acid	D
Glutamic Acid	E
Lysine	K
Arginine	R
Histidine	H

Hangi eşleme daha iyi?

Eşleme 1

```
A - C - S S - A P T
|   |   |       | |
A T C S S A T - P T
```

Eşleme 2

```
A - C S S - A P T
|   | | |     | |
A T C S S A T P T
```

Puanlama Yöntemi : BLOSUM 62 and Gap Penalty is 5

Eşleme 1:

```
A - C - S S - A P T
|   |   |       | |
A T C S S A T - P T
4-5+9-5+4+1-5-5+7+5 = 2
```

Eşleme 2:

```
A   -   C   S   S   -   A   P   T
|       |   |   |           |   |
A   T   C   S   S   A   T   P   T
4 - 5 + 9 + 4 + 4 - 5 - 0 + 7 + 5 = 23
```

Amino asitleri birbirinden ayrı ayrı değerlendirir…

Algoritmalar (Algorithms):

Sekans eşleme problemi için birçok farklı algoritmalar denenmiş ve halen birçok farklı program kullanılmakta ve geliştirilmektedir. Bunlar arasında kesin sonuçlar veren ama yavaş çalışan dinamik programlama (dynamical programming) teknikleri olduğu gibi büyük veri bankalarının yaranmasında kullanılan rastsal (heuristic) yöntemlerde vardır.

Global eşlemeler verilen sekanslar üzerindeki tüm amino asitleri eş zamanlı olarak eşlemeye çalışırlar ve sekanslar birbirleri ile benzer ve aminoasit sayıları yaklaşık olarak eşit ise kullanılır. Needleman-Wunsch algorithm.

Local eşlemeler ise verilen sekanslar birbirlerinden oldukça farklı ise ve belirli bölgelerde benzer motifler aranıyorsa kullanılır. Smith-Waterman algorithm.

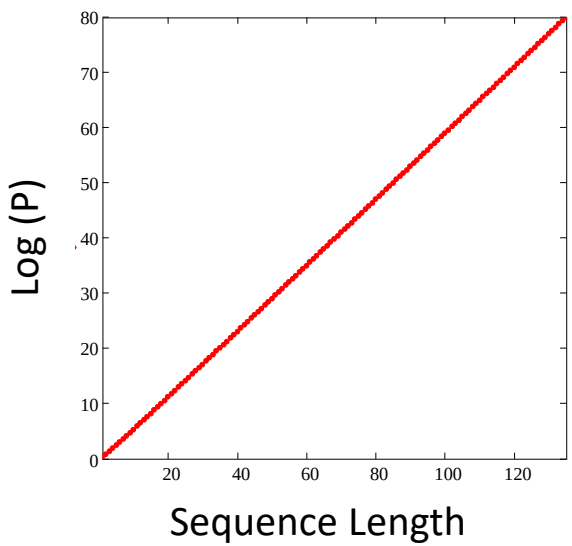
```
Global FTFTALILLAVAV
       F--TAL-LLA-AV
```

```
Local  FTFTALILL-AVAV
       --FTAL-LLAAV--
```

Eşlenecek sekanlar birbirlerine yeterince benzer ise her iki yöntem de aynı sonuçları verir.

Uzunlukları N olan iki sekans kaç farklı şekilde eşleşebilir:

$$P = \frac{2^{2N}}{\sqrt{\pi N}}$$



for N = 135

$P = 10^{80}$

Needleman-Wunsch Algoritması:

Needleman S. & Wunsch C. A general method applicable to the search for similarities in the amino acid sequence of two proteins, *J Mol Biol*. 48(3):443-53 (1970).

Dinamik programlama tekniği kısaca problemi küçük parçalara bölerek bu parçalar üzerinde bulduğu eşlemelerden tüm parçanın eşlemesini elde eder.

Needleman-Wunsch algoritması

Böl ve yönet stratejisi uygular:

- Problem daha küçük altproblemlere bölünür.
- Altproblemlerin çözümleri bulunur.
- Altproblemlerin çözümleri kullanılarak asıl problem için optimal çözüm bulunur.

Verilen iki sekans arasındaki mümkün olan tüm eşleşmeler göz önüne alınır.

Bu durumda iki farklı matris elde edilir: skorlama matrisi and geridönüş matrisi.

The Needleman-Wunsch algoritması 3 adımdan oluşur:

1. Skorlama matrisi oluşturulur
2. Skorlar hesaplanarak geridönüş matrisi oluşturulur
3. Geridönüş matrisi kullanılarak sekans eşlemesi yapılır

Linear Gap vs Affine gap Penalty

$$g(k) = -gk \qquad g(k) = -go - (k-1).ge \qquad for\ k = 1, 2, \ldots, n$$

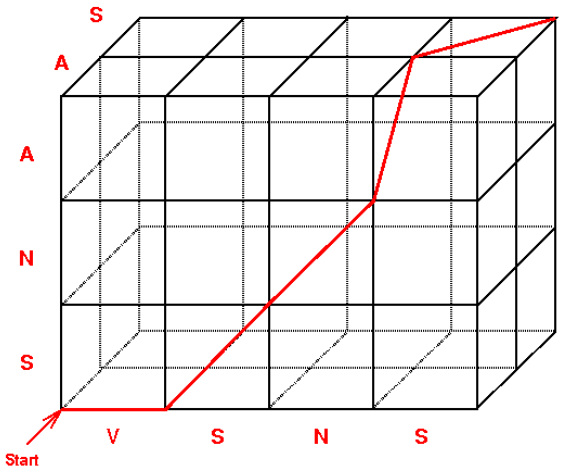
k : gap length
g : gap penalty
go : gap opening penalty
ge : gap extension penalty

Kontrol:

```
                SEND
Alignment 1: -AND score: -10-1+6+6 = +1
Alignment 2: A-ND score: +1-10+6+6 = +3 en iyi
Alignment 3: AN-D score: +1+0-10+6 = -3
Alignment 4: AND- score: +1+0-3-10 = -8
```

Needleman-Wunsch algoritmasının dezavantajları

Çoklu sekans eşleme (Multiple Sequence Alignments) yöntemleri için uygun değildir.




```
V S N _ S
_ S N A _
_ _ _ A S
```

Gerekli hesap sayısı: n^N

n sekanslar içerisindeki amino asit sayısı

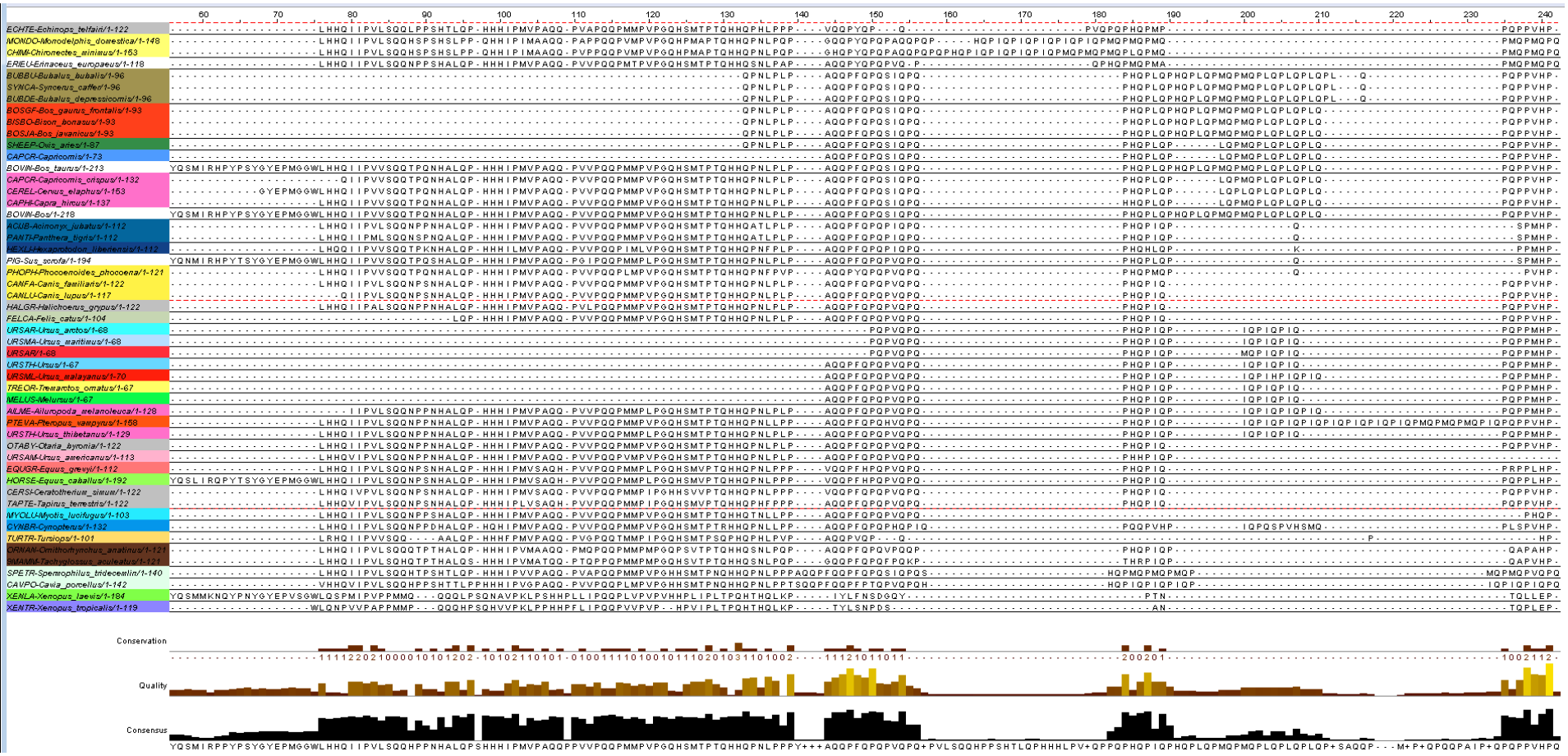
N sekans sayısı

```
Seq. 1: -PTS-AFFQWLPRDE--- 13
Seq. 2: -----AFFQWLPTTHILS 13
Seq. 3: APTSGGSQTWLPR----- 13
                   ***
```

Eşlenecek sekans sayısı artmaya başladıkça hızlı bir şekilde gereken hesaplama sayısı astronomik sayılara çıkarak pratik olarak hesaplanamaz hale gelir.

Çoklu sekans eşleme Multiple Sequence Alignment Methods: CLUSTAL W

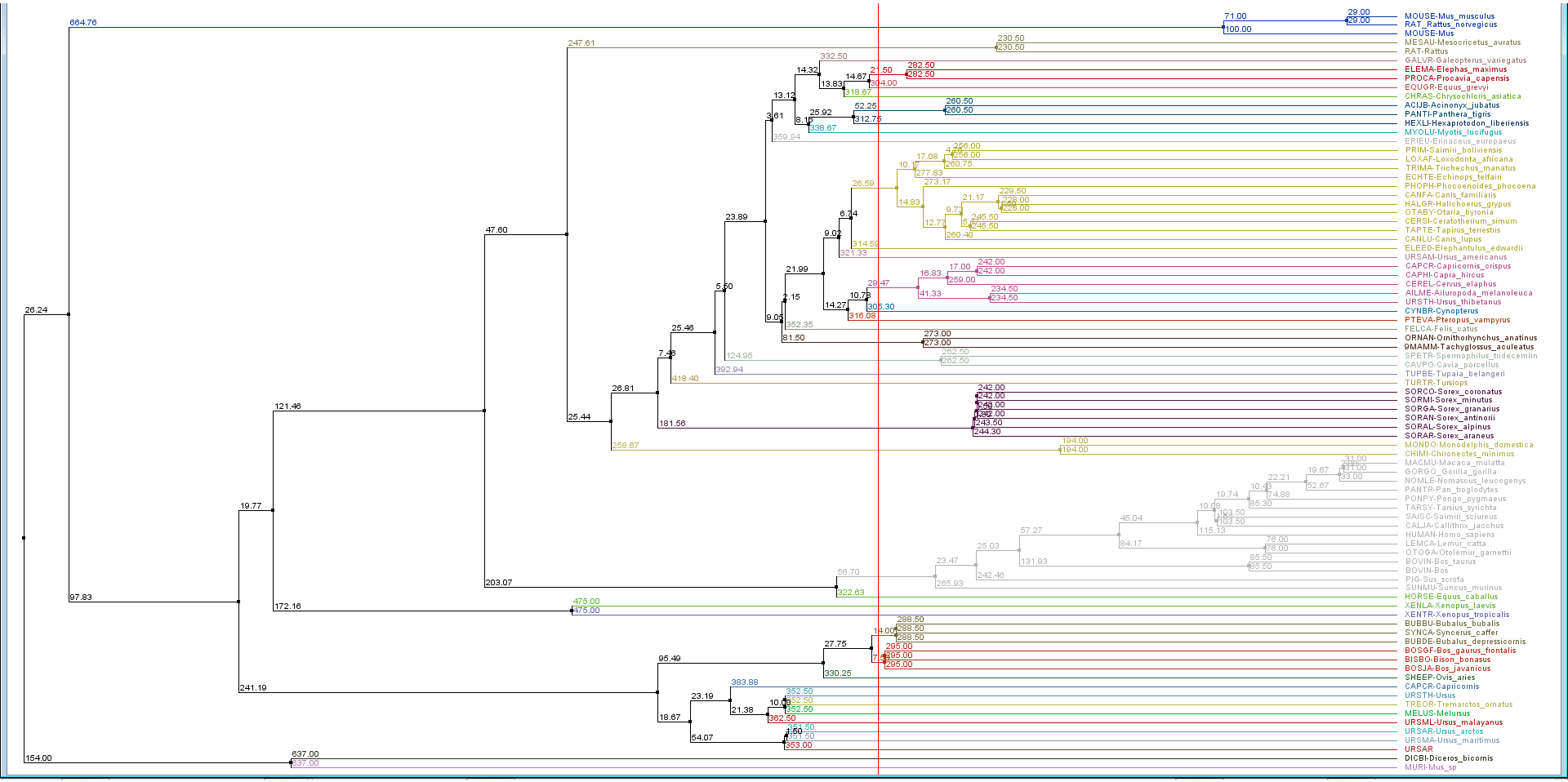
Örnek: Amelogenin Proteins



Filogenetik Ağaçlar (Phylogenetic Trees)

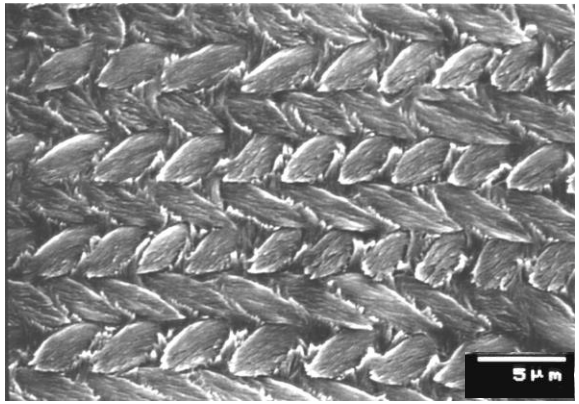
Ortak ataya sahip çeşitli türlerin veya diğer varlıklar arasındaki evrimsel ilişkileri gösteren grafiklerdir...

Örnek: Amelogenin

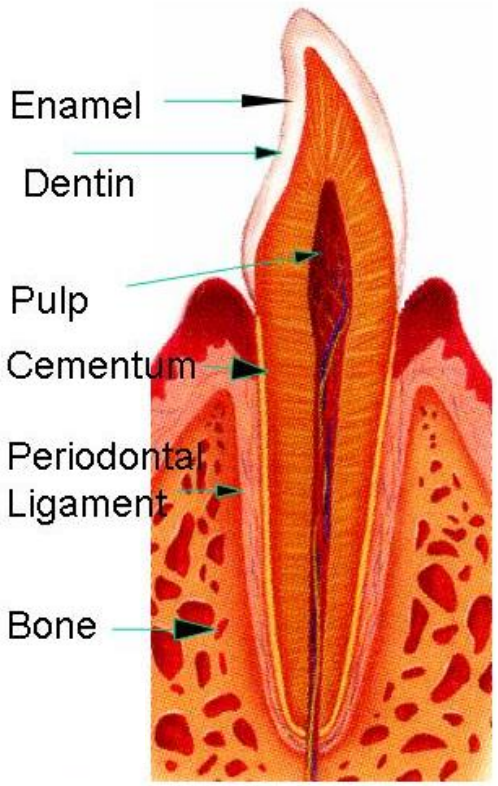


Hidroksiapatit:
Sert diş dokularının tamamı HA minerali içerir, dokular arasındaki fark mineralin 3 boyutlu organizasyonu yani nano ve mikro yapısındaki farklılıklardan kaynaklanır.

Amelogenin:
Diş minesinin oluşumu aşamasında sistemdeki proteinlerin % 90'ını oluşturur.



Amelogenin-directed enamel growth



AMELX Fonksiyonları

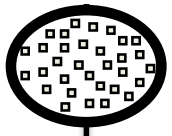
- Kalsiyum fosfat oluşumunda çekirdeklenmeyi sağlar (Tarasevich *et al*., 2007),
- Mineral büyümesini farklı kristal yüzeylere bağlanması ile kontrol eder. (Robinson *et al*., 1979),
- Hücreler arası sinyalleşmeyi kontrol eder. (Boabaid *et al*., 2004; Lacerda-Pinheiro *et al*., 2006; Le *et al*., 2007; Tompkins and Veis, 2002; Tompkins *et al*., 2005),
- ……

Protein Functional Domains (PFDs)

GEPIs

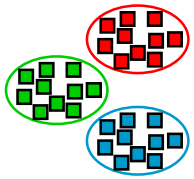
Selection

Phage or Cell Surface Display

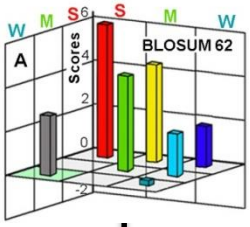


Characterization

Fluorescence Microscopy



Similarity Analysis



Scoring Matrix

Natural vs **Forced Evolution**

Material Specific Sequence Similarity Matrices (QUARTZ I, HA I, etc.)

Natural Proteins

Mineral Binding Proteins

Magnetite, Hydroxyapatite, Aragonite, Calcite, Silica, etc.

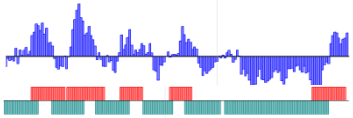
Sequence

Detection of Similarities

Sequence Comparison

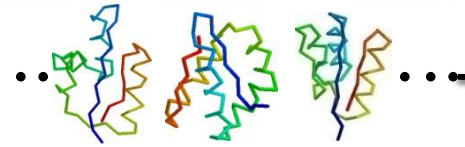
Silica Binding Peptides vs Silicateins

HAp bindign peptides vs Amelogenin



Structure Prediction

Generate → Minimize → Filter

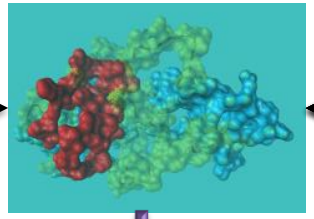


Inspection

Solvent Accessibility

Residue Functional Importance

Ion Binding Regions



MFS: Meta Functional Signatures

A. MFS Structural Signatures

B. MFS Sequence Signatures

SOAK: Ion Binding Domains

PFDs → **Characterization**

Binding & Biomineralization

GEPI vs Doğal Proteinler

Proteinlerin Benzerlik Analizi

Mineral	Protein 1	Protein 2	Protein 3
Magnetite	mms6	mms7	…
Hydroxyapatite	Amelogenin	Enamelin	…
Aragonite	PFMG1	AP7	…
Silica	Silafin	Silicatein	…

>Verilen bir protein sekansı için

MGTWILFACLLGTAFAMPLPPHPGHPGYINFSYEKSHSNAINI.....LPDLPLEAWPATDKTKREEVD

Strong

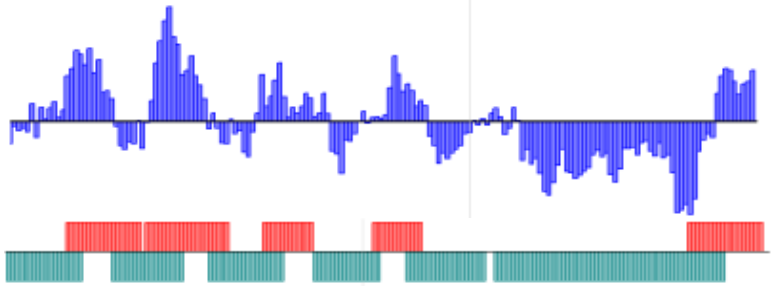
SPTKPTPPRSSQ
TSTNYWLYSSES
VPFQFKVTGDPL
AFSQLKGFYSRY
EFYTPTGLPPGR
HTVNRSMDVPGV
NTPAHANADFFD
ASGAKPWTSDLH
IPMTPSYDSHIL
HAPYKSHVWTEQ
AFAYRDNLSMHP
LLADTTHHRPWT
HWGEIPSRLSLP
LDTQFIKPPQKS
SVAALFRHVPGH
NGWWTASPGVPM

N: Protein sekansının uzunluğu
n: Peptit sekansının uzunluğu

} N – n + 1 Parça

HA I

Toplam Benzerlik Skoru



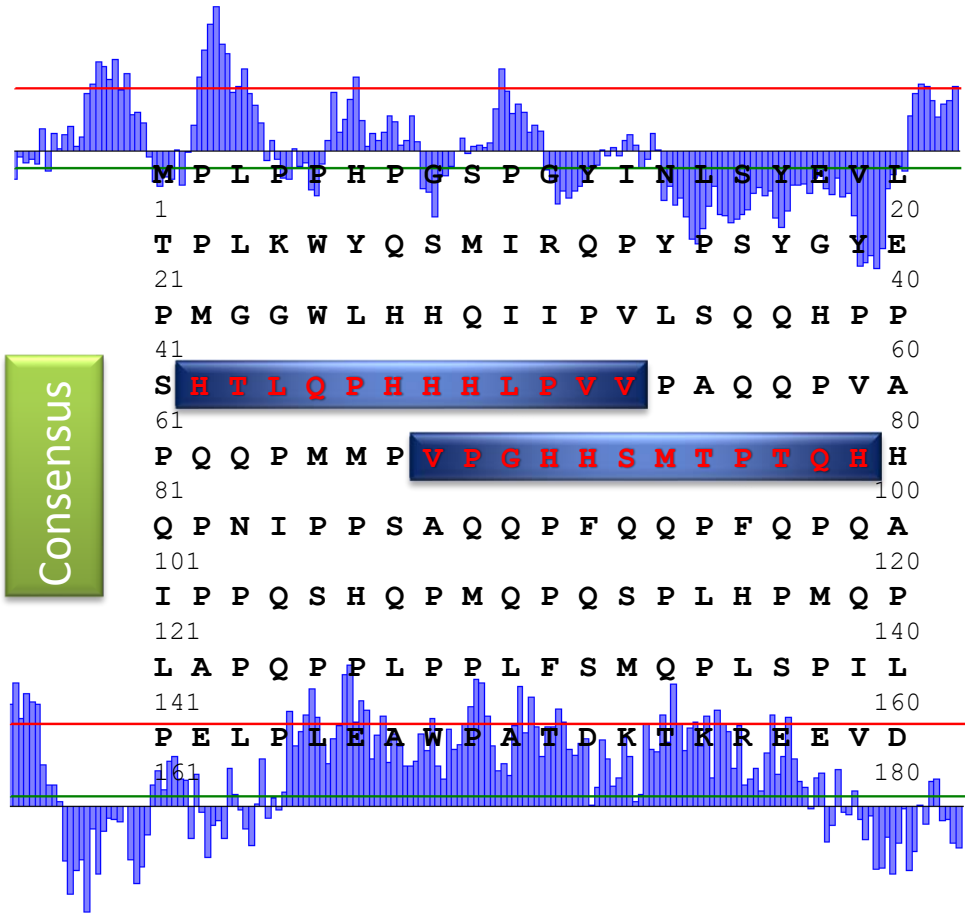
Ph. D. 12

```
M P L P P H P G S P G Y I N L S Y E V L
1                                      20
T P L K W Y Q S M I R Q P Y P S Y G Y E
21                                     40
P M G G W L H H Q I I P V L S Q Q H P P
41                                     60
S H T L Q P H H H L P V V P A Q Q P V A
61                                     80
P Q Q P M M P V P G H H S M T P T Q H H
81                                    100
Q P N I P P S A Q Q P F Q Q P F Q P Q A
101                                   120
I P P Q S H Q P M Q P Q S P L H P M Q P
121                                   140
L A P Q P P L P P L F S M Q P L S P I L
141                                   160
P E L P L E A W P A T D K T K R E E V D
161                                   180
```

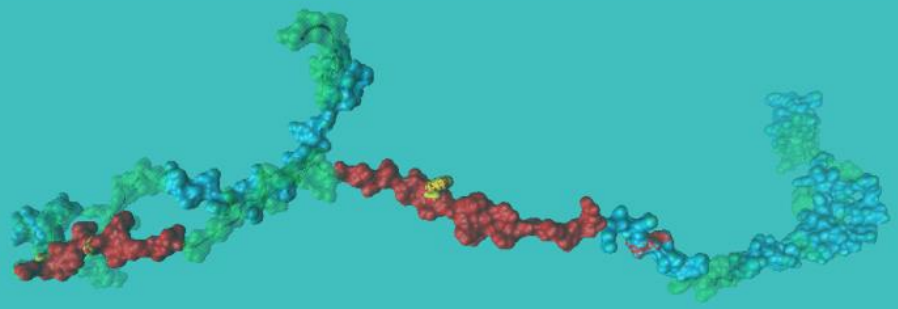
Ph. D. 7C

```
M P L P P H P G S P G Y I N L S Y E V L
1                                      20
T P L K W Y Q S M I R Q P Y P S Y G Y E
21                                     40
P M G G W L H H Q I I P V L S Q Q H P P
41                                     60
S H T L Q P H H H L P V V P A Q Q P V A
61                                     80
P Q Q P M M P V P G H H S M T P T Q H H
81                                    100
Q P N I P P S A Q Q P F Q Q P F Q P Q A
101                                   120
I P P Q S H Q P M Q P Q S P L H P M Q P
121                                   140
L A P Q P P L P P L F S M Q P L S P I L
141                                   160
P E L P L E A W P A T D K T K R E E V D
161                                   180
```

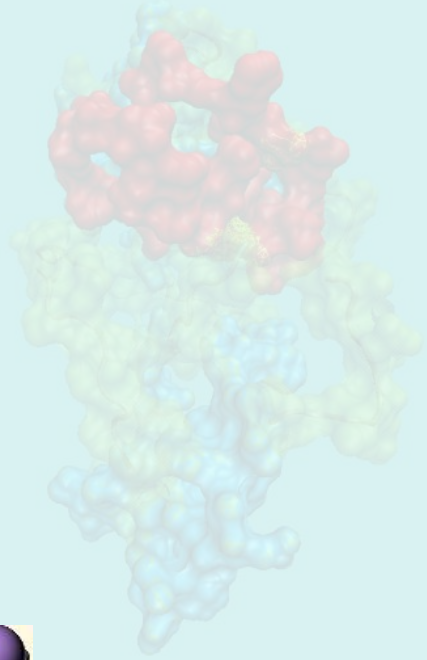
Consensus

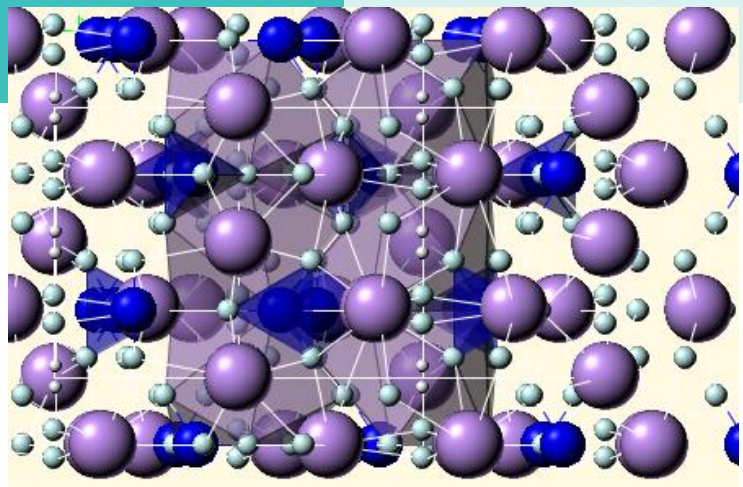


Linear Models



Globular Models





Ca

P

O

H

$Ca_{10}(PO_4)_6(OH)_2$

SOAK

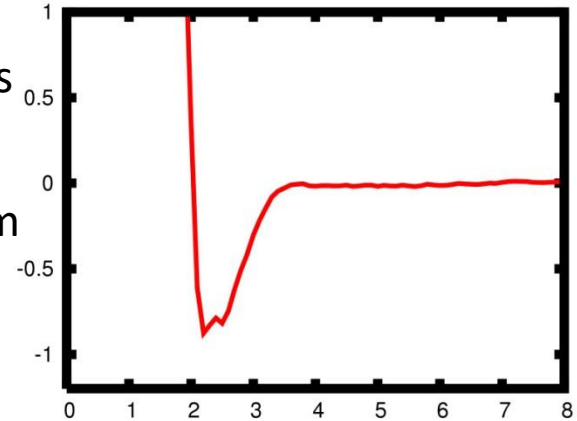
A novel method of accurately predicting metal ion-binding pockets using 3D space and charge distributions across solvent-exposed areas on a protein structure (SOAK; (Cheng et al., *in preparation*)).

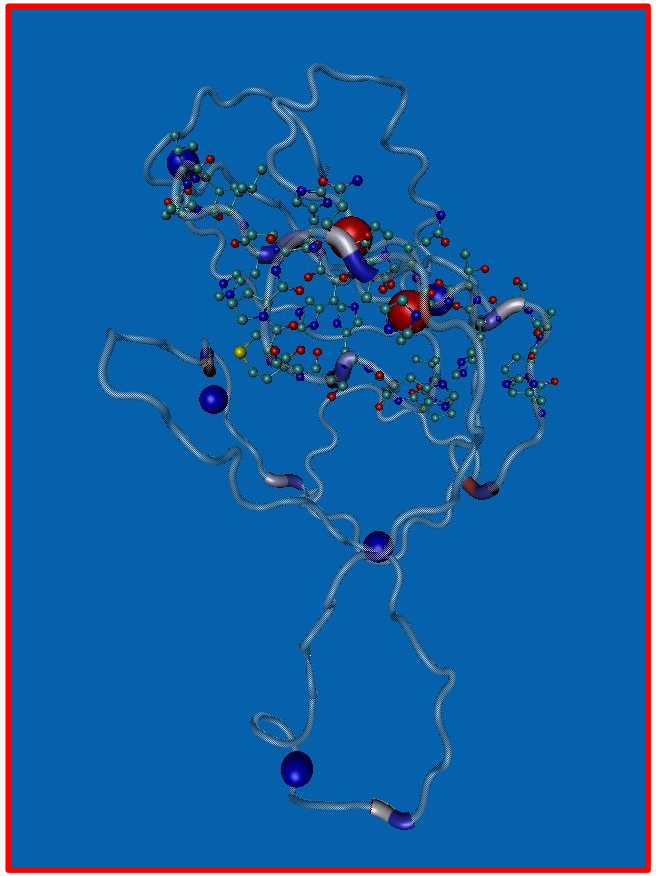
http://protinfo.compbio.washington.edu/soak/

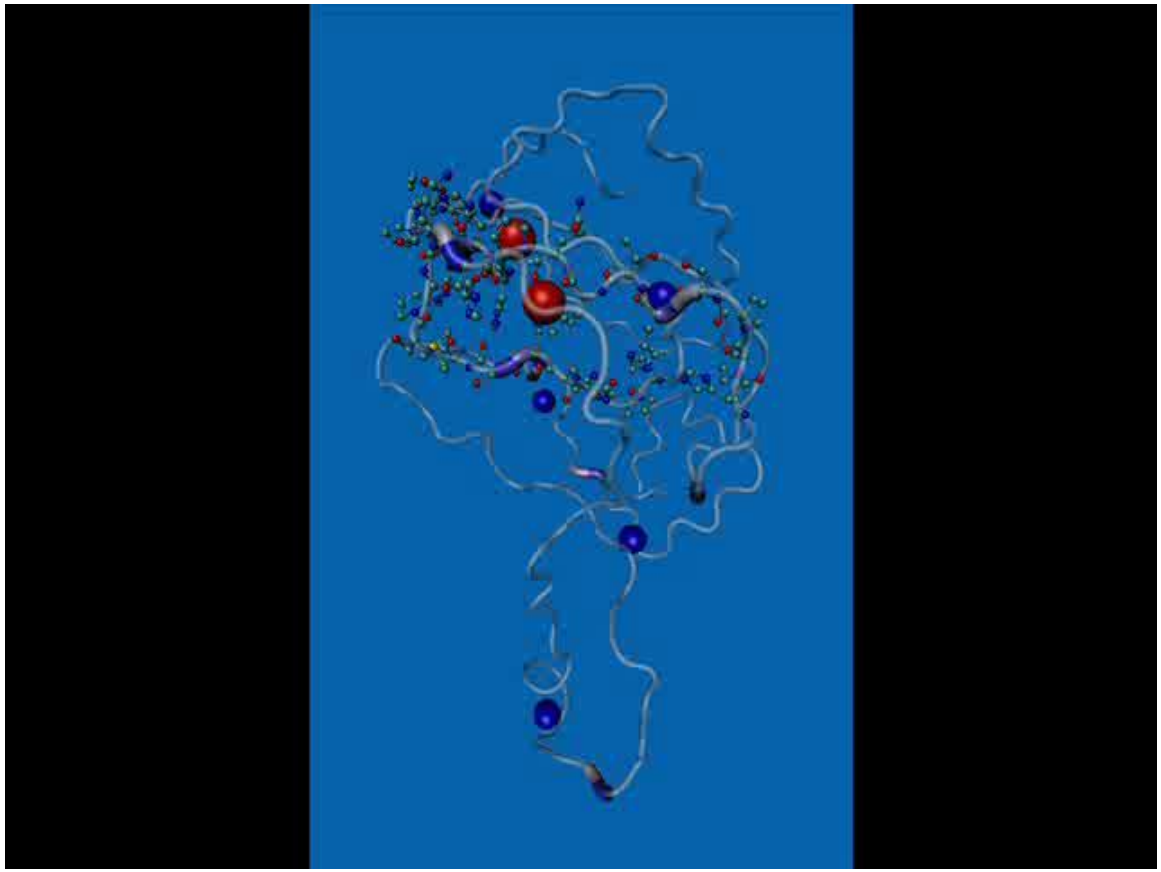
Crystal Structures

↓

Bayesian Theorem

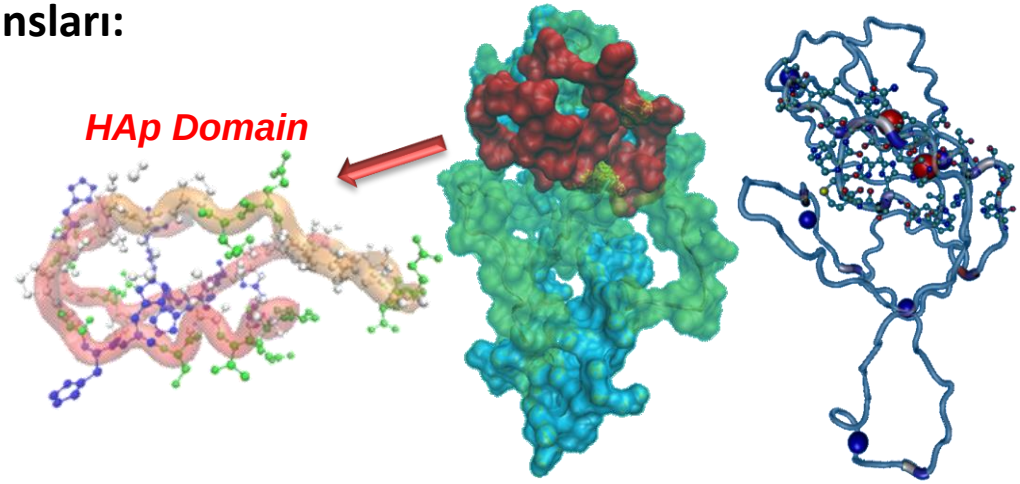


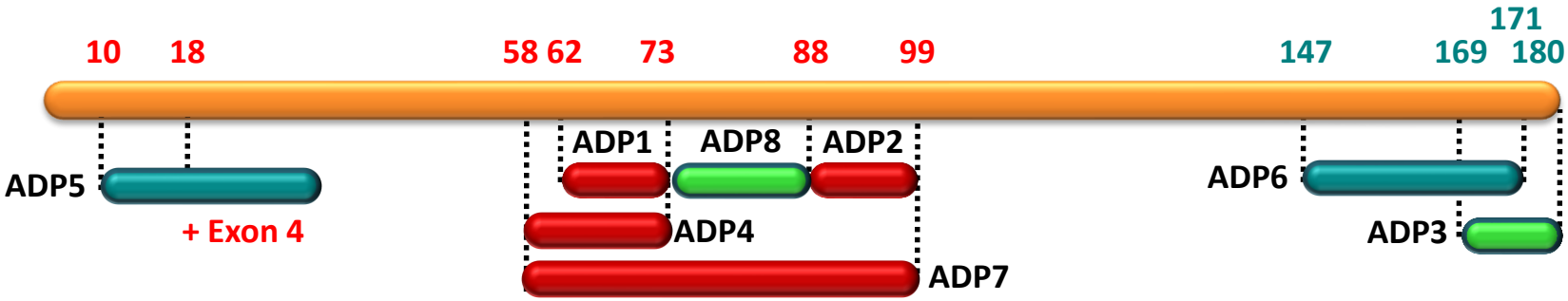




Fare amelogeninden elde edilen peptit sekansları:

ADP1	HTLQPHHHLPVV	(12 aa)
ADP2	VPGHHSMTPTQH	(12 aa)
ADP3	WPATDKTKREEVD	(13 aa)
ADP4	HPPSHTLQPHHHLPVV	(16 aa)
ADP5	PGYINLSYEKSHSQAINTDRTA	(22 aa)
ADP6	LPPLFSMQPLSPILPELPLEAWPAT	(25 aa)
ADP7	HPPSHTLQPHHHLPVVPAQQPVAPQQPMMPVPGHHSMTPTQH	(42 aa)
ADP8	PAQQPVAPQQPMMP	(14 aa)



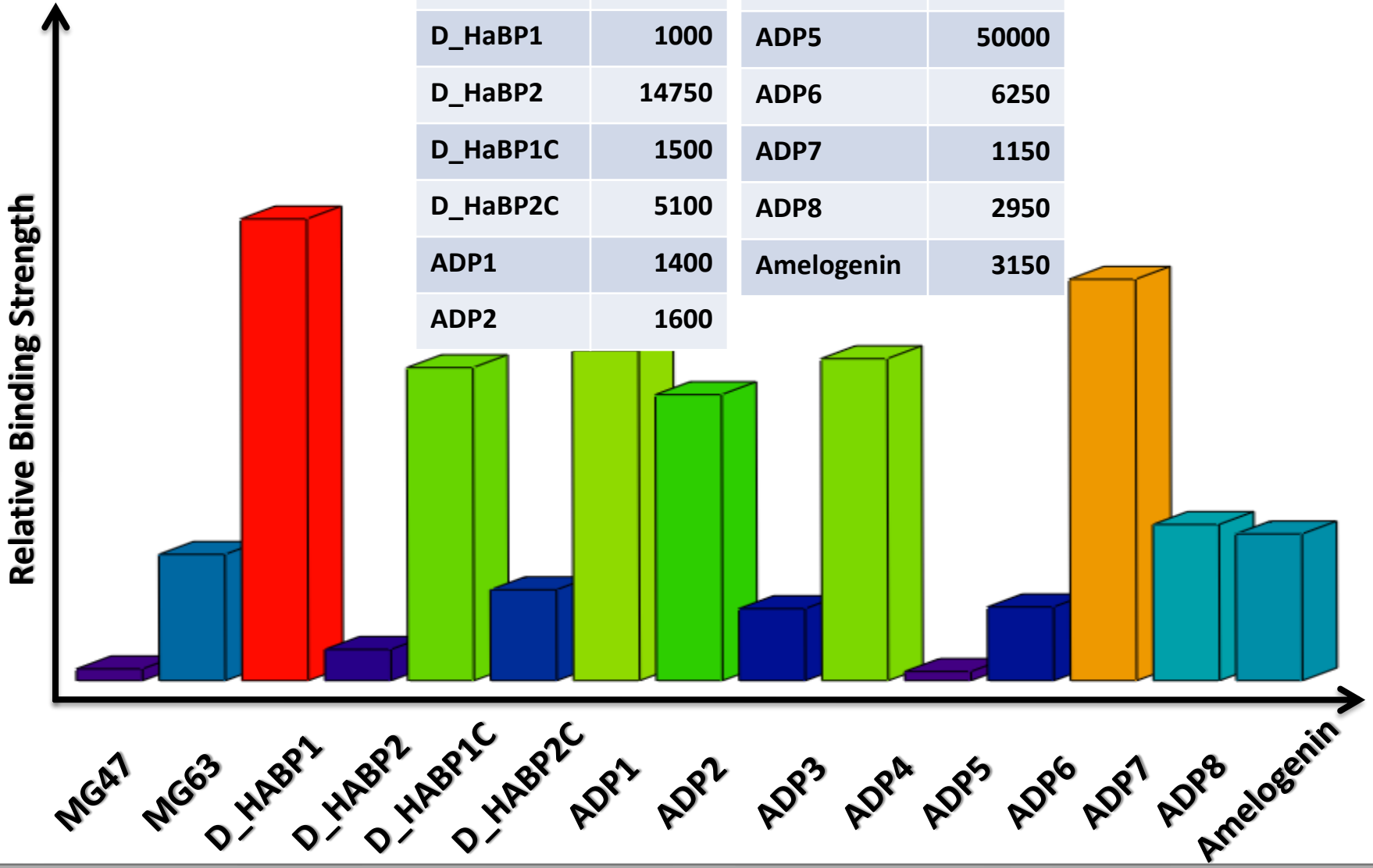


Mouse amelogenin M180 From *M. L. Snead Lab*, Univ. of Southern California, LA, CA, USA.

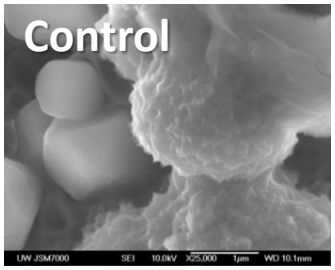
QCM: Bağlanma Analizi

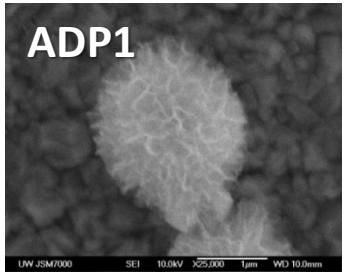
	KD (nM)
MG47	3650
MG63	38950
D_HaBP1	1000
D_HaBP2	14750
D_HaBP1C	1500
D_HaBP2C	5100
ADP1	1400
ADP2	1600

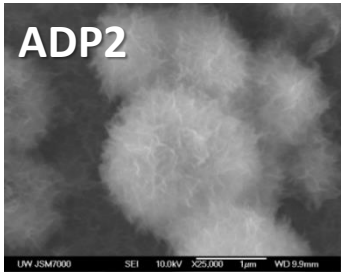
	KD (nM)
ADP3	6400
ADP4	1400
ADP5	50000
ADP6	6250
ADP7	1150
ADP8	2950
Amelogenin	3150

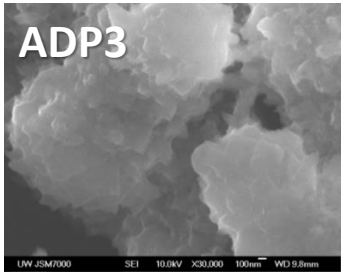


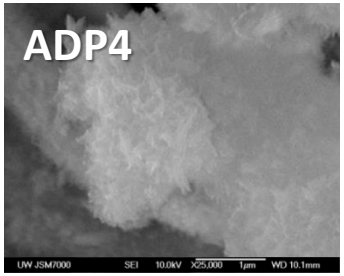
Biyominerilizasyon

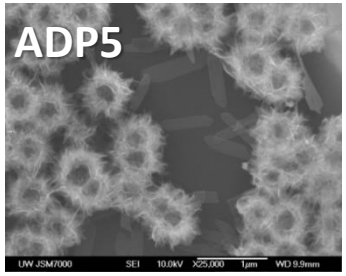


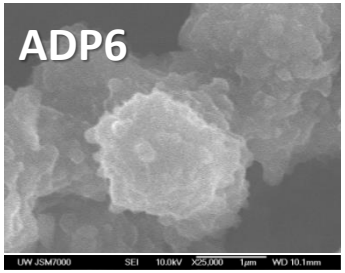


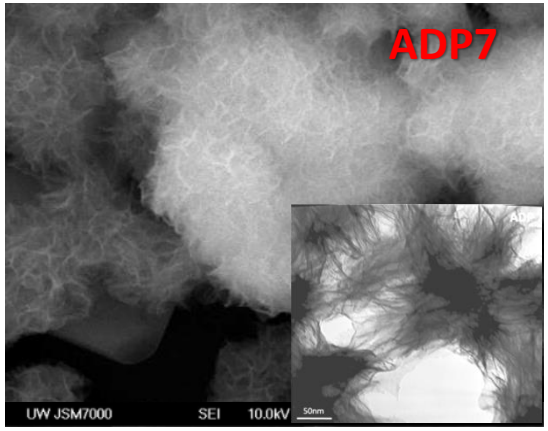


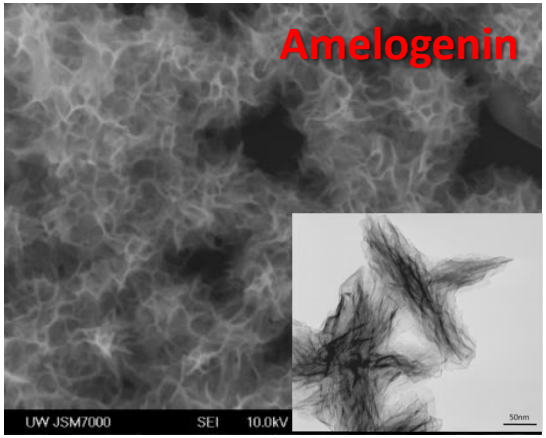


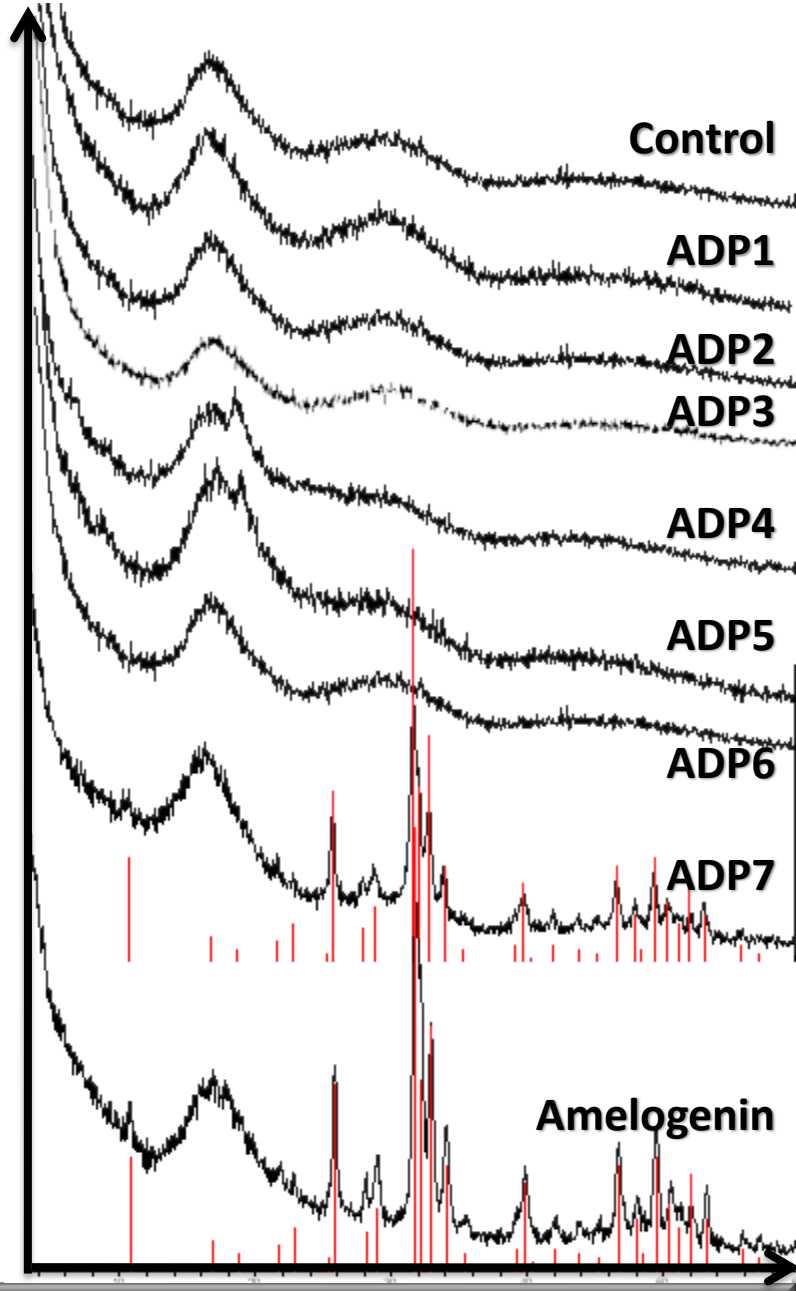












BMM 205 Dersinde görüşmek üzere

Başarılar!