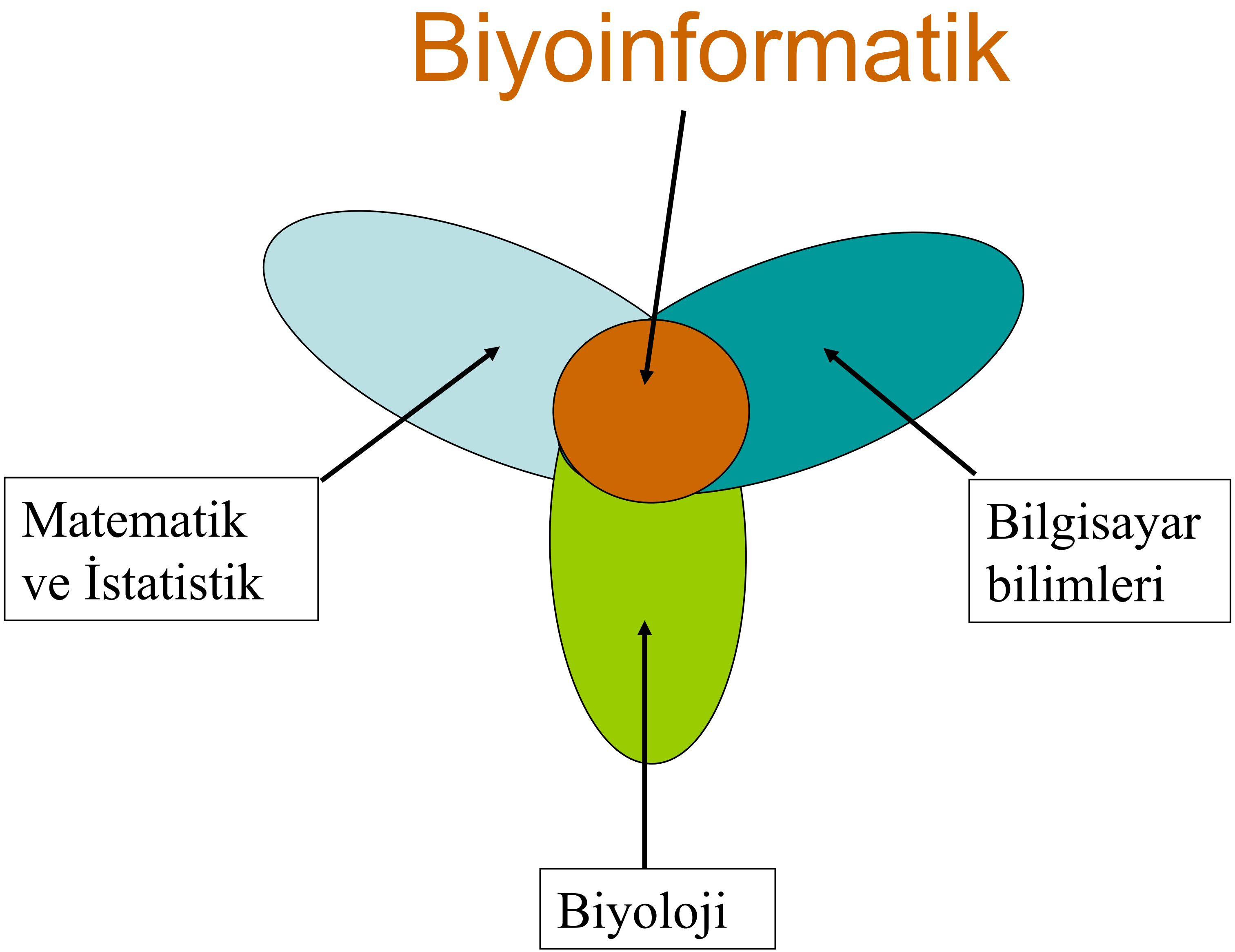
Biyoinformatik

- Bilgisayarların moleküler biyolojide kullanımı
- Moleküllerin üç boyutlu grafik temsili
- Moleküler dizilimler
- Üç boyutlu yapı moleküler veritabanları

- Kısa sürede yüksek miktarda veri üretilmesi
- Protein-protein ilişkileri
- Biyolojik olarak aktif moleküllerin araştırılması
- Bakteri, maya, insan, hayvan, bitki genom projelerinden elde edilen bilgiler

- Biyolojik problemlerin çözümünde bilişim teknolojilerinin kullanılması


Biyoinformatik
Matematik ve İstatistik
Bilgisayar bilimleri
Biyoloji

İki temel bilgi vardır:

- Genetik bilgi akışı:

  bir organizmanın DNAsı incelenir ve o organizma türünün oluşturduğu toplulukların karakteristik özelliklerine kadar olan bilgi akışı

- Deneysel bilgi akışı: biyolojik olaylar gözlenerek elde edilen bilgi açıklayıcı modeller ile tarif edilir, daha sonra bu modellerin doğruluğu yeni deneylerle test edilir.

Metodolojik çalışmalar

- DNA dizileme çalışmaları
- Protein dizi çalışmaları
- Makromoleküllerin yapılarının üç boyutlu dizilimleri
- Küçük moleküllerin ligandlarla ilişkilerinin araştılırılması
- Biyolojik bilginin paylaşımının kolaylaştırılması
- Bilgisayar ile otomize edilmiş veri analizi
- Etkileşimde bulunan gen ürünleri için veritabanları oluşturulması
- Biyolojik faaliyet süreçlerinin simülasyonu
- Büyük çaplı projelerden çıkan sonuçların değerlendirilmesi (genom projeleri)

Biyolojik çalışmalar

- Protein yapı ve fonksiyonunun belirlenmesi
- Her hangi bir biyolojik fonksiyonu artıran veya azaltan moleküllerin tasarlanması
- Karmaşık genetik fonksiyon veya regülasyonun tanımlanması
- Tıbbi ya da endüstriyel amaçlı yeni makro moleküller üretmek
- Genetik faktörlerin hastalık yatkınlığına etkilerinin ortaya çıkarmak

- Genetik bilgi:

  DNA $\longrightarrow$ protein

Biyoinformatik veritabanlarının kısa tarihçesi:

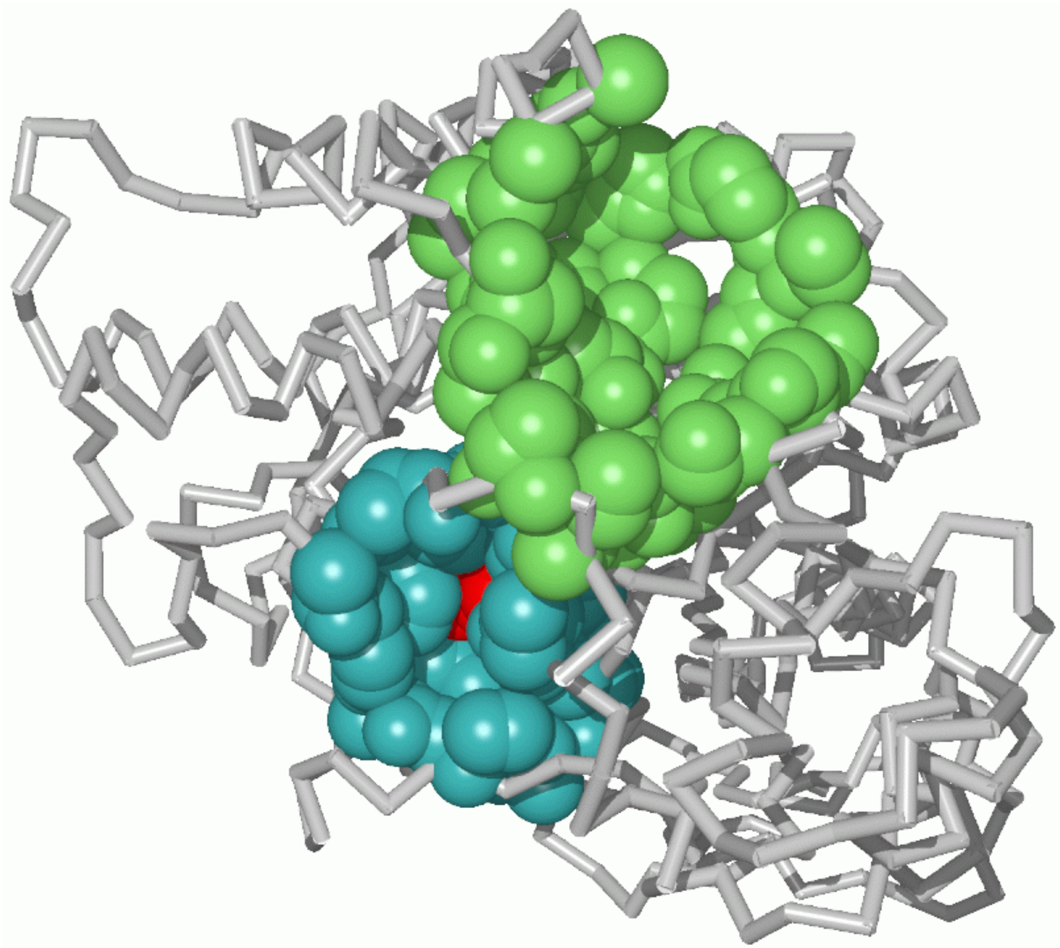
• ilk biyolojik veritabanı - Protein Identification Resource Margaret Dayhoff tarafından 1972'de kuruldu

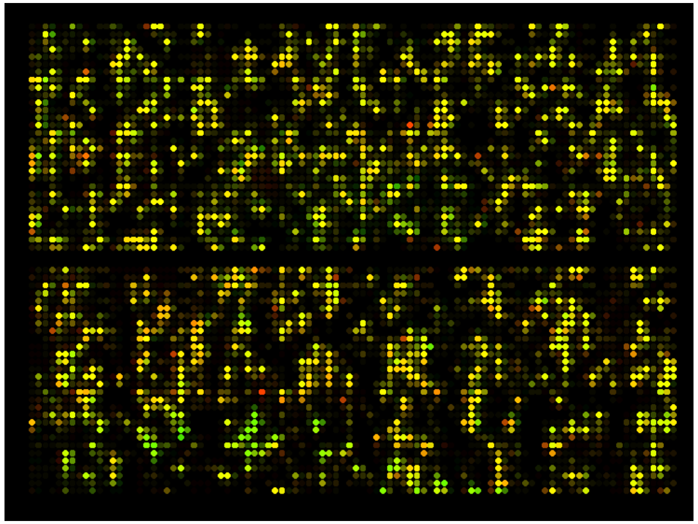
• Dayhoff ve arkadaşları protein ailelerini organize ettiler

•Protein dizilerinin karşılaştırılması fikri doğdu

• Şu an iki büyük protein bankası var : SwissProt, PIR International….

• İlk DNA veritabanı 1979'da kuruldu. Şu an çok sayıda veritabanı var: GenBank, EMBL, DDBJ…….



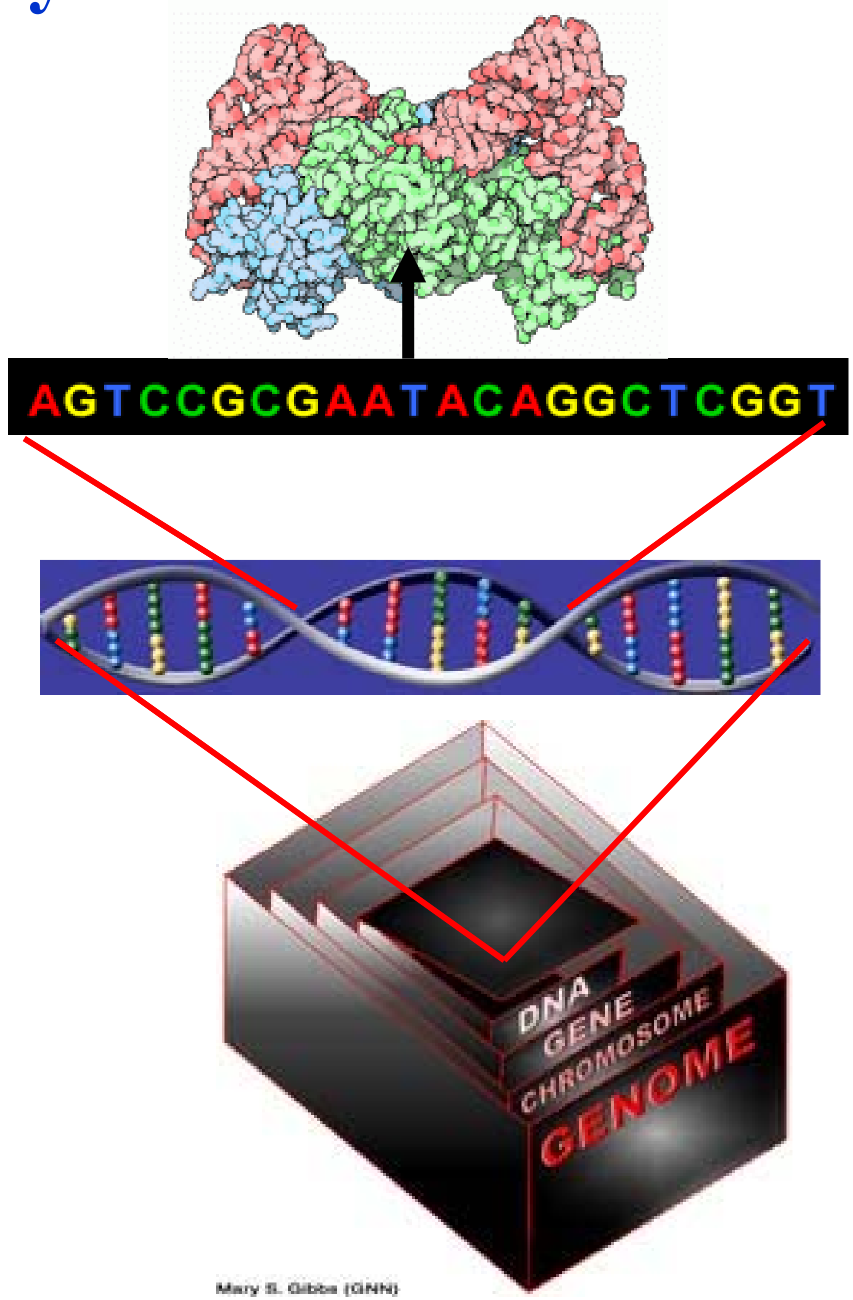


Gen ekspresyonu = Protein üretimi

Protein

Nukleotid dizisi

DNA



Biyolojik Çeşitlilik

Bakteriler

Mayalar

Meyva sineği

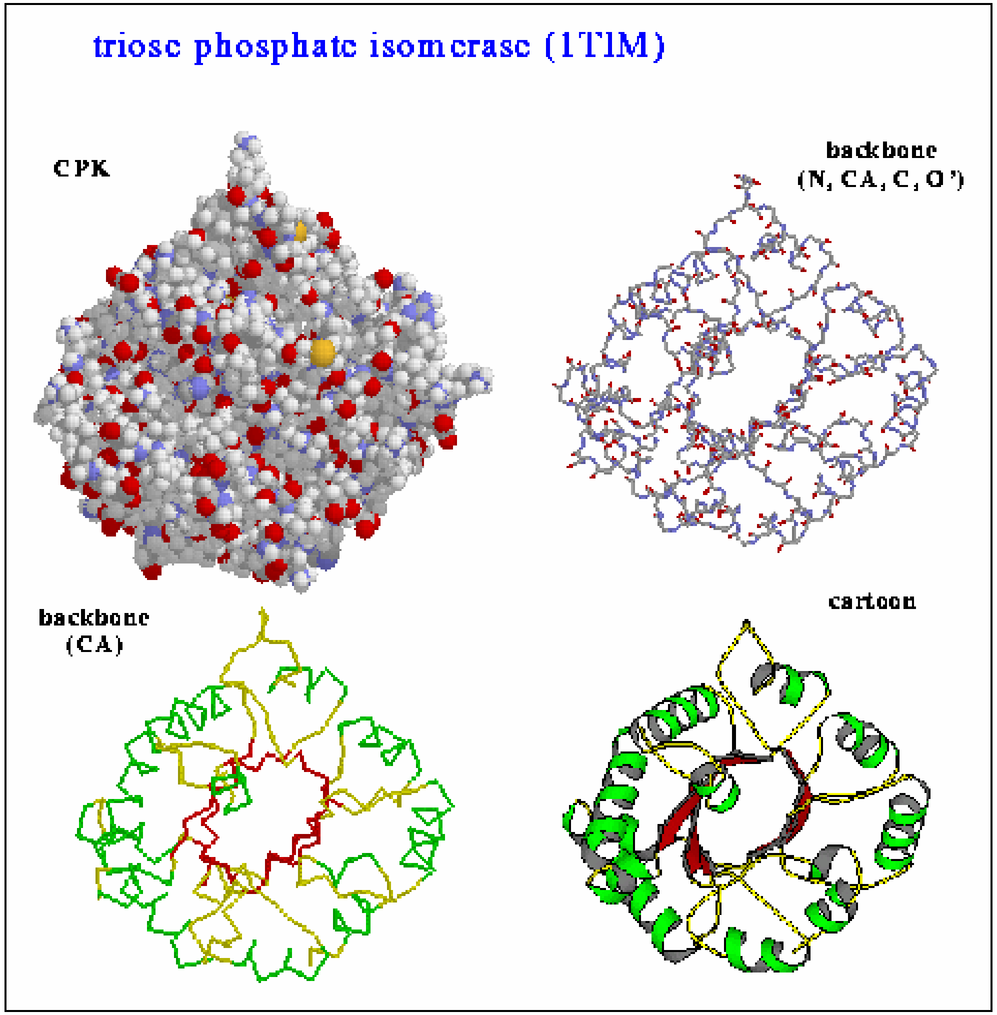
İnsan

Organism	Number of genes in the genome
Myscoplasma genitalium	517
Saccharomyces cerevisiae	6,275
Arabidopsis thaliana	~ 20,000
Caenorhabditis elegans	19,099
Haemophilus influenzae	1,743
Drosophila melanogaster	13,601
Neisseria meningitdis	2,158
Homo sapiens	~ 30,000

Protein dizisinden Yapısına

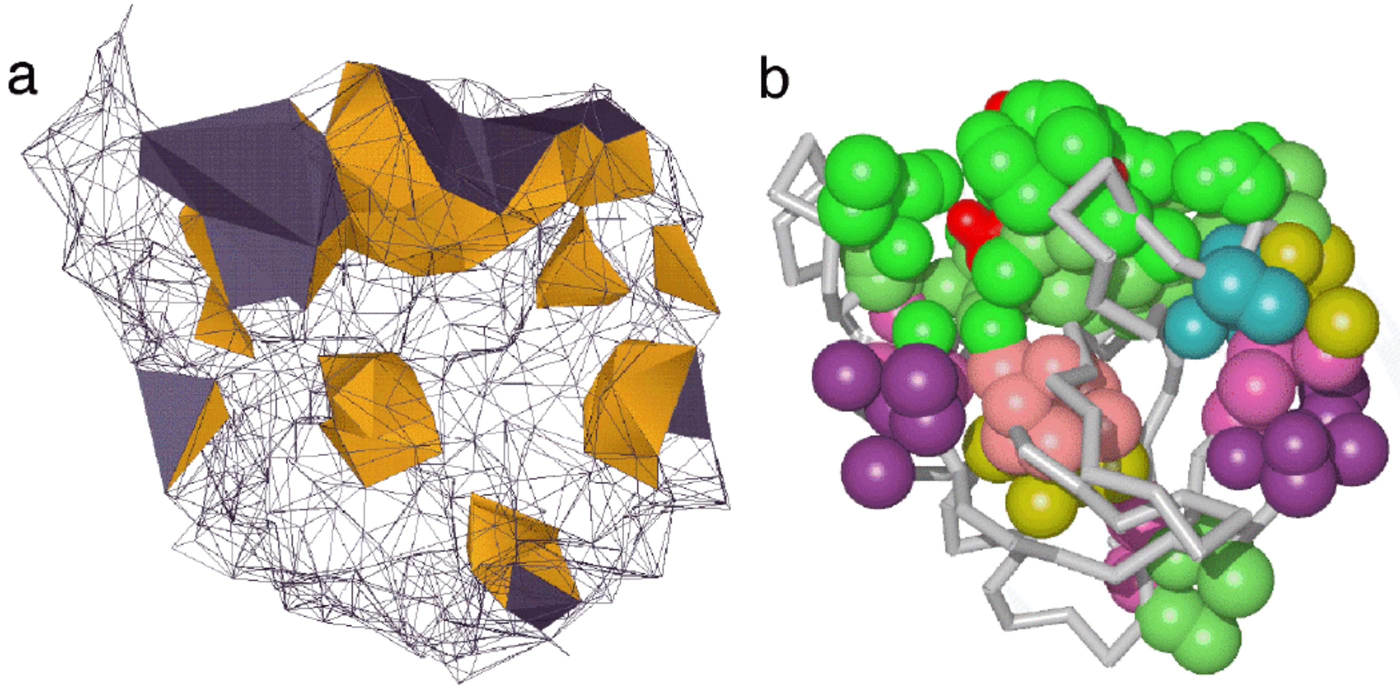
APRKFFVGGNWKMNGD
KKSLGELIHTLNGAKL
SADTEVVCGAPSIYLD
FARQKLDAKIGVAAQN
CYKVPKGAFTGEISPA
MIKDIGAAWVILGHSE
RRHVFGESDELIGQKV
AHALAEGLGVIACIGE
KLDEREAGITEKVVFE
QTKAIADNVKDWSKVV
LAYEPVWAIGTGKTAT
PQQAQEVHEKLRGWLK
SHVSDAVAQSTRIIYG
GSVTGGNCKELASQHD
VDGFLVGGASLKPEFV
DIINAKH

=



- Interdisipliner yaklaşım
  - Bilgisayar bilimleri, Matematik & İstatistik.
  - Moleküler biyoloji, Biyokimya & Tıp.

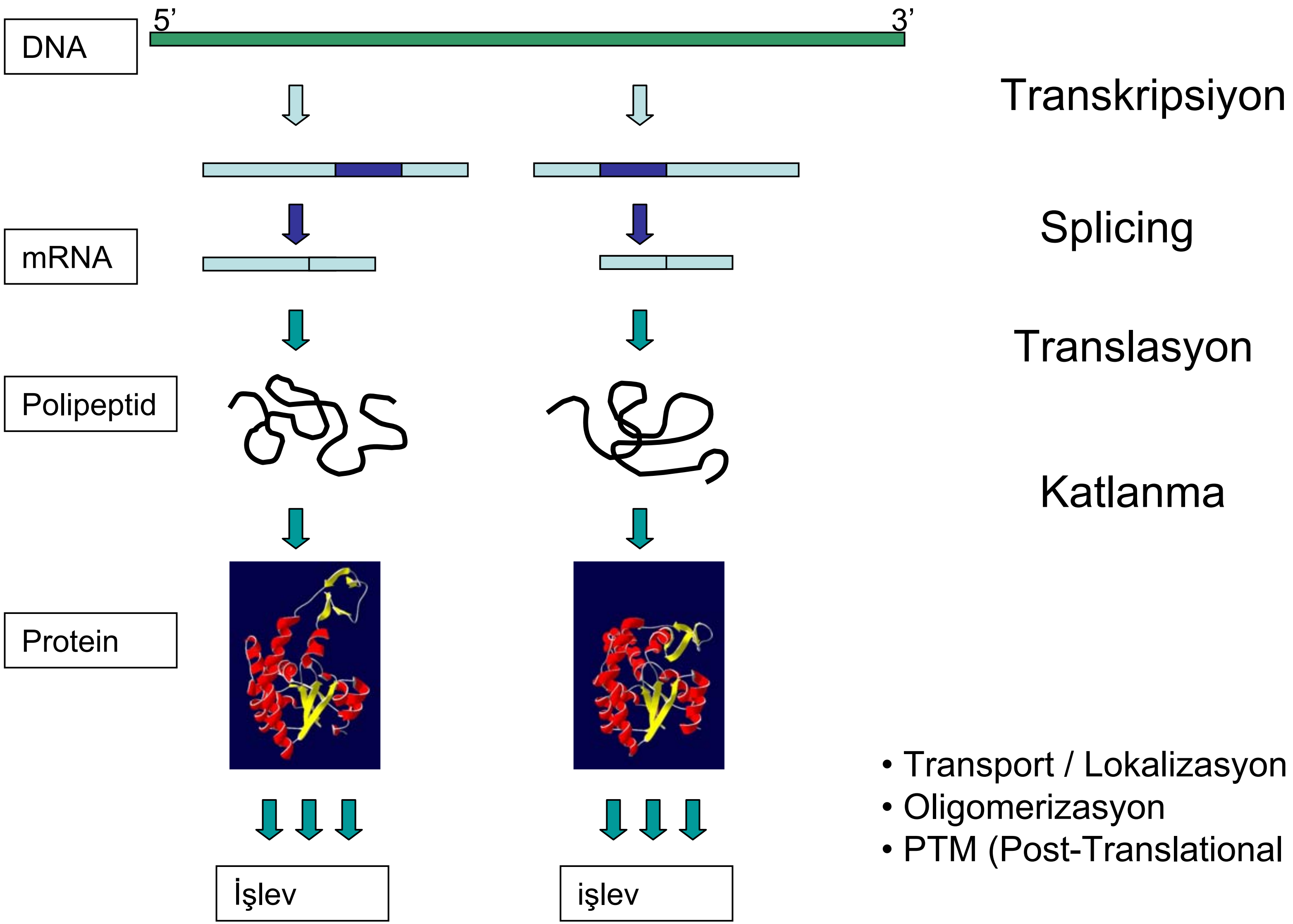
Pockets in Ribonuclease A



Biyoinformatik

- **Bazı Eş anlamlar:**
- Moleküler Biyoinformatik
- Computational Biology
- Biocomputing

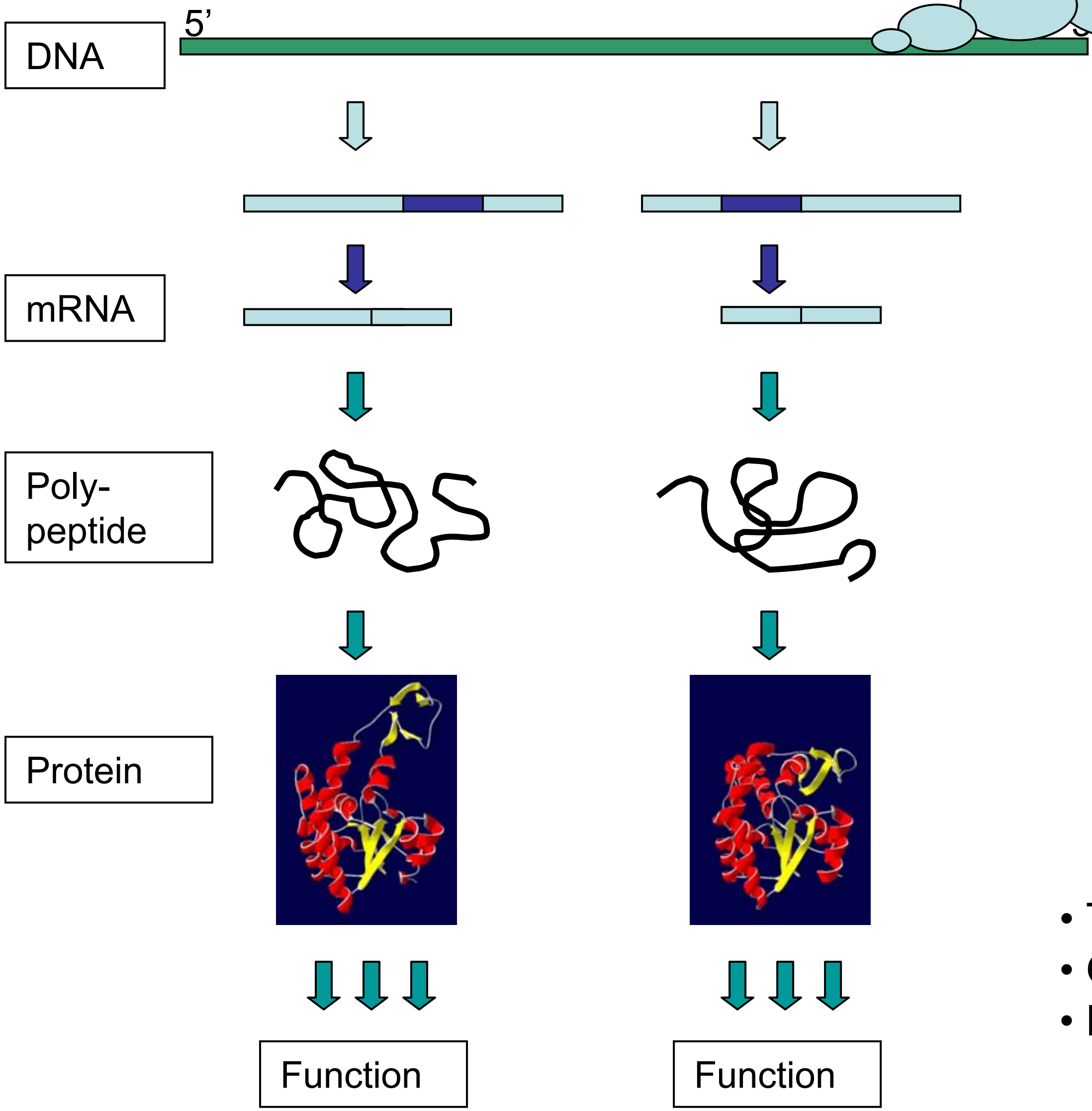
Biyoinformatiğe neden ihtiyacımız var?



- Transport / Lokalizasyon
- Oligomerizasyon
- PTM (Post-Translational Modification)

Biyoinformatiğe neden ihtiyacımız var?

Genom projelerinden gelen DNA dizilerinin saklanmaya ve organize edilmeye ihtiyacı var



Transcription

Splicing

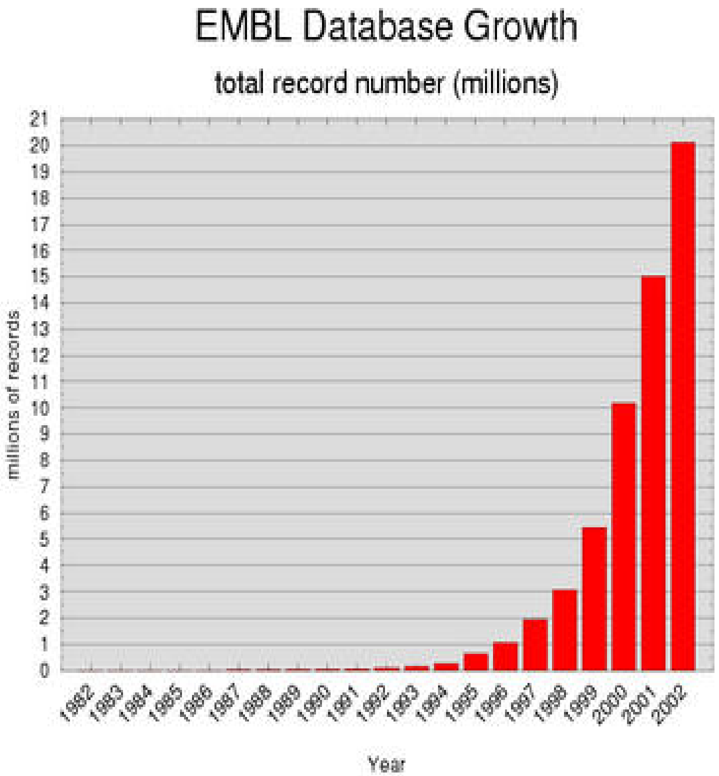
Translation

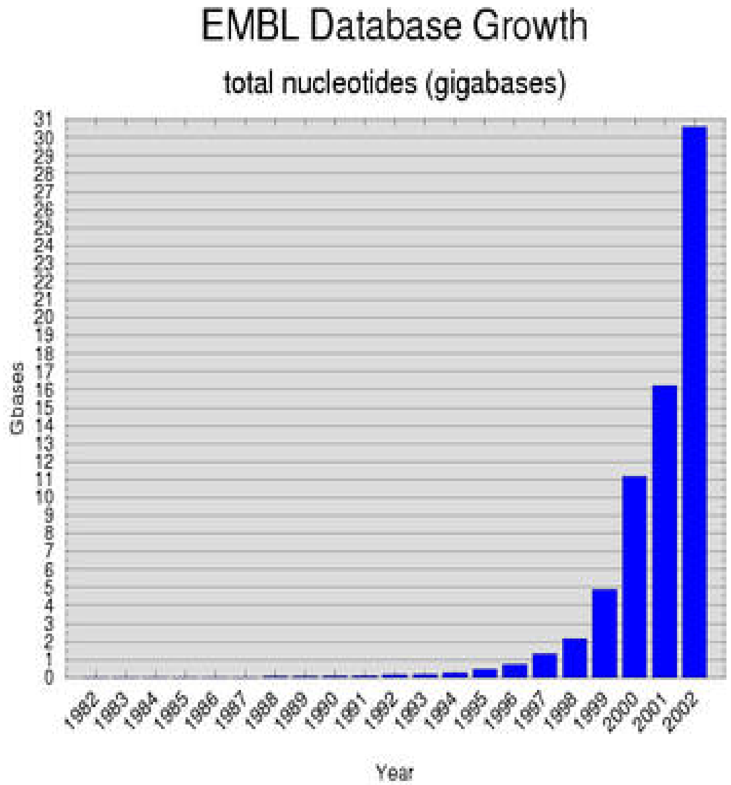
Folding

- Transport / Localization
- Oligomerization
- PTM (Post-Translational Modification)

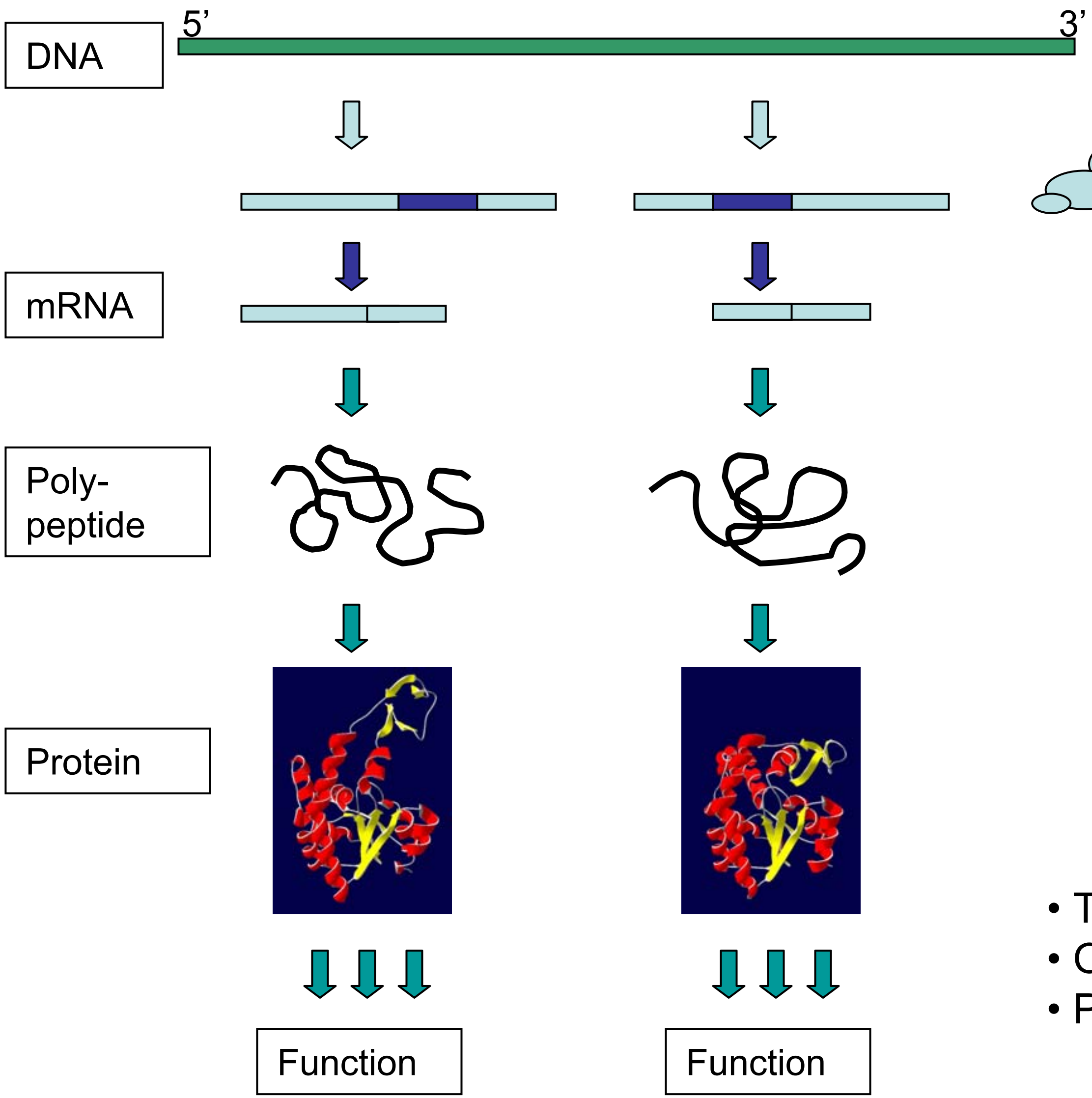
DNA veribankaları

- GenBank (USA) http://www.ncbi.nlm.nih.gov/Genbank/
- EMBL (Europe) http://www.ebi.ac.uk/embl/
- DDBJ (Japan) http://www.ddbj.nig.ac.jp/


EMBL Database Growth
total record number (millions)
millions of records
Year
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002


EMBL Database Growth
total nucleotides (gigabases)
Gbases
Year
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002

Biyoinformatiğe neden ihtiyacımız var?



Protein kodlayan bölgeleri nasıl bulacağız? intronları ve eksonları nasıl ayıracağız?

Splicing

Translation

Folding

- Transport / Localization
- Oligomerization
- PTM (Post-Translational Modification)

Prokaryotik genomlarda genler nasıl bulunur?

İçsel bilgi:

- Transkripsiyon sinyalleri
- kodon kullanımı
- GC – içeriği

Dışsal bilgi:

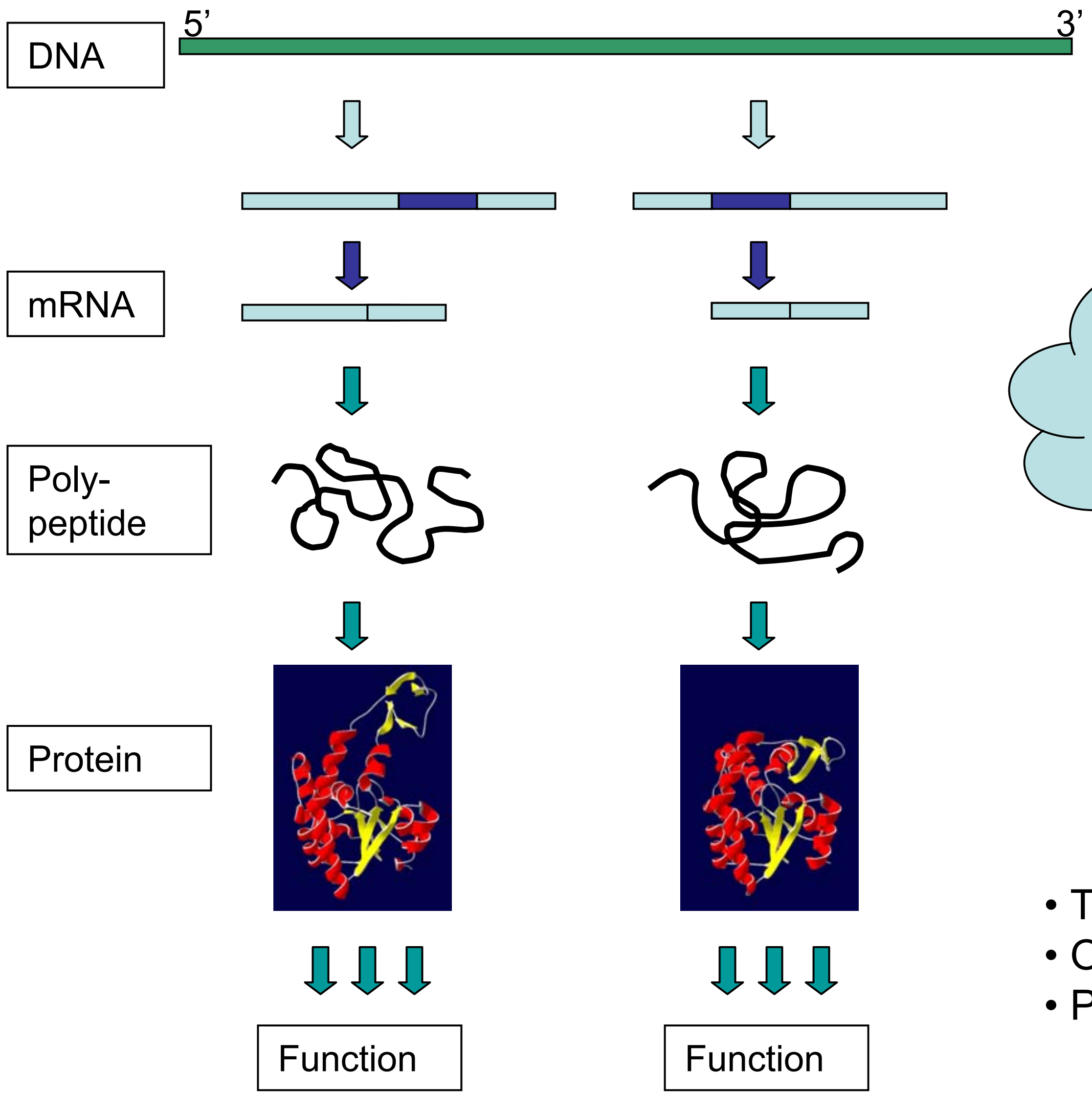
- Bilinen proteinlere benzerlik

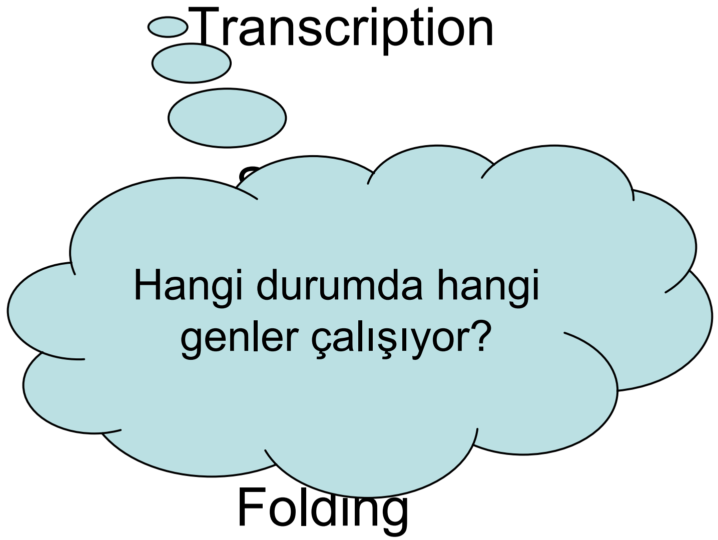
Ökaryotik genomlarda genler nasıl bulunur?

Bazı problemler:

- Genomun kompleksliği organizmanın kompleksliği ile orantılıdır
- 1-2 % kodlama yapan bölge
- Intron – exon yapıları
- Alternatif kırpılma (bütün genlerin ~30%)
- Pseudogenler

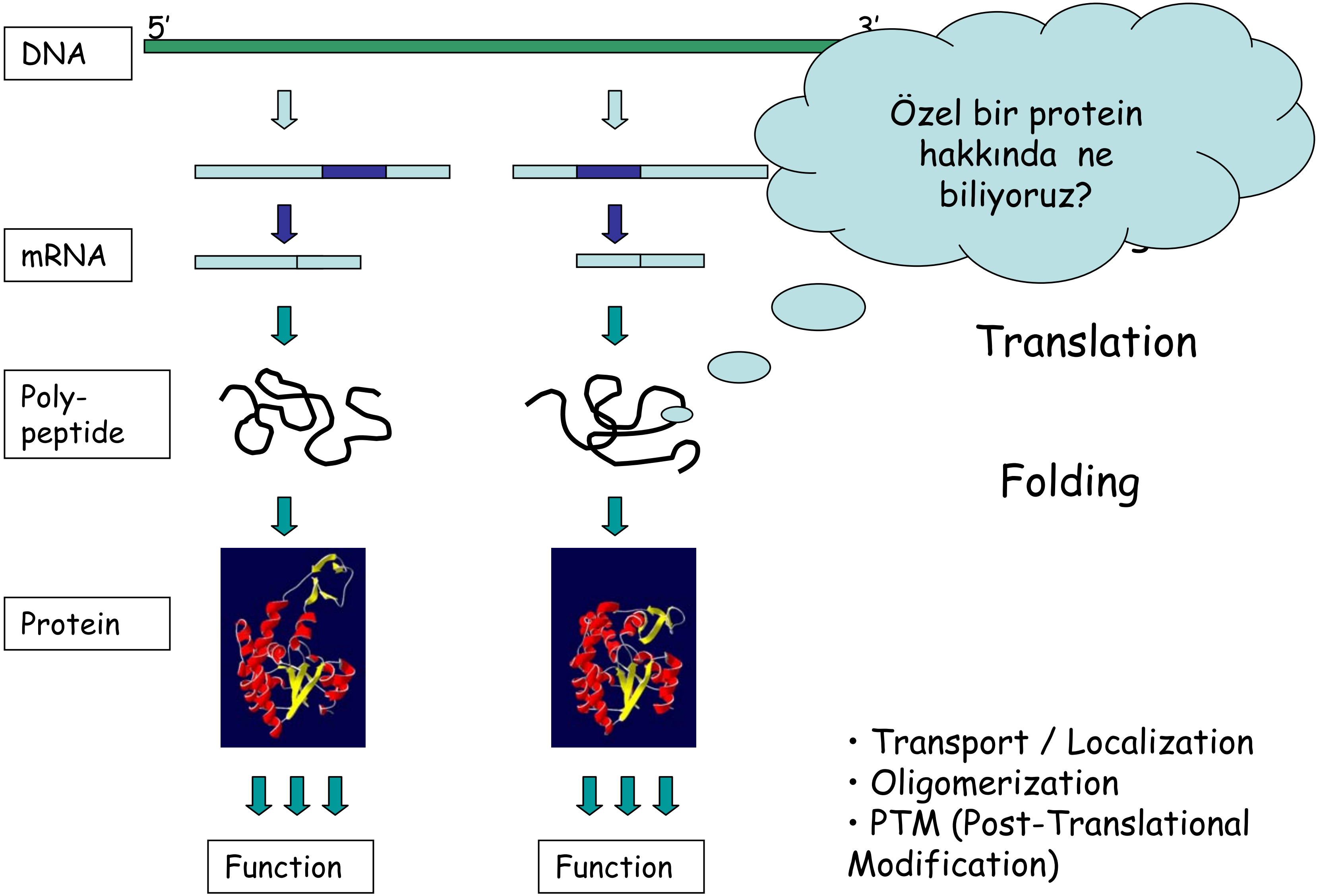
Biyoinformatiğe neden ihtiyacımız var?





- Transport / Localization
- Oligomerization
- PTM (Post-Translational Modification)

Biyoinformatiğe neden ihtiyacımız var?



Bir protein veritabanında ki minimal içerik

- diziler !!
- Accession number (AC)
- Taxonomik bilgi
- Referanslar
- Annotasyon
- Keywordler
- Dokumantasyon

SWISS-PROT/TrEMBL

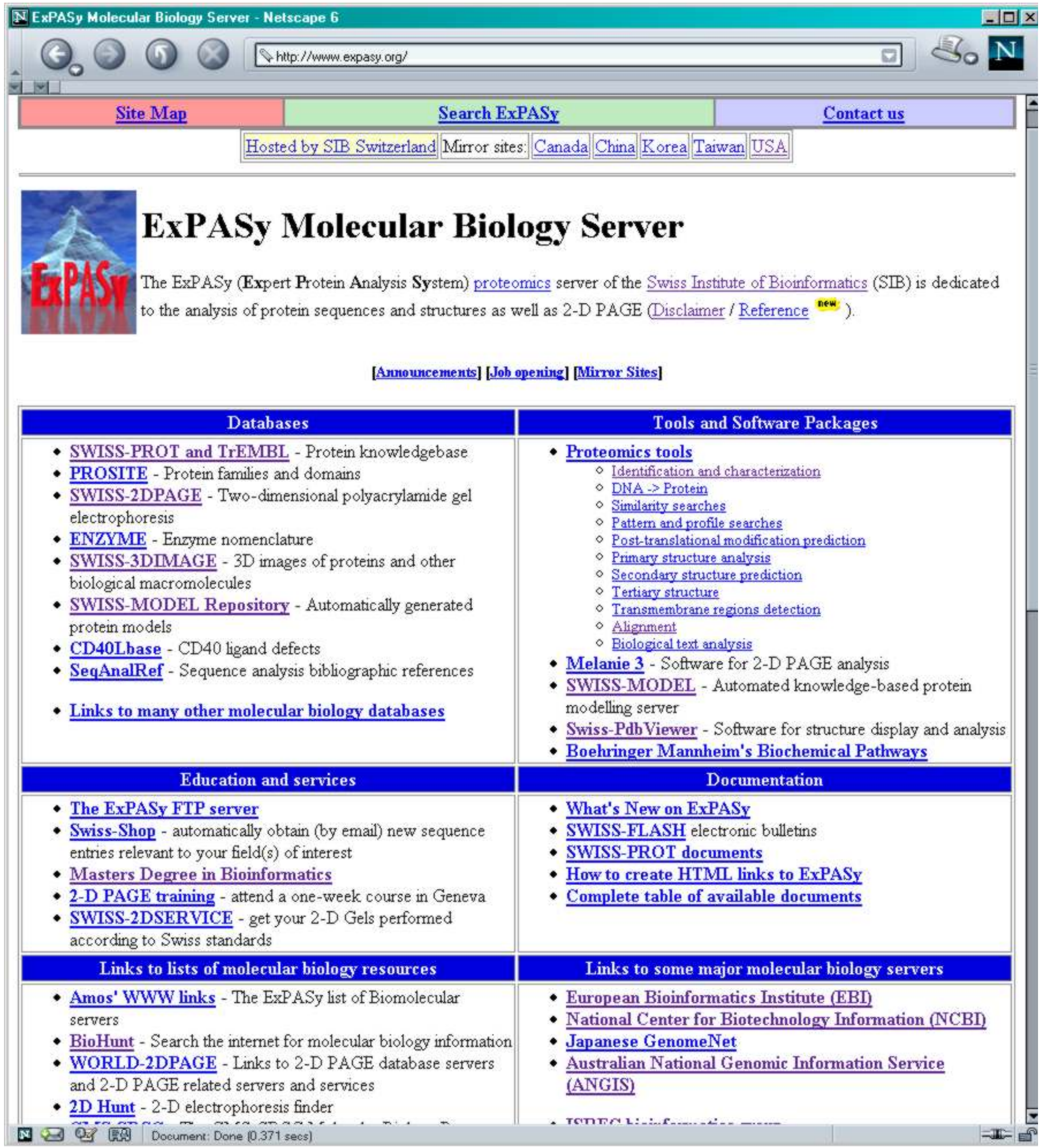
- the SIB (CH) ve EMBL/EBI (UK) arası ortaklık
- SWISS-PROT: protein dizisi veritabanı.
- TrEMBL

http://www.expasy.org/sprot/

ExPASy Web Server

ExPASy =

Expert
Protein
Analysis
System

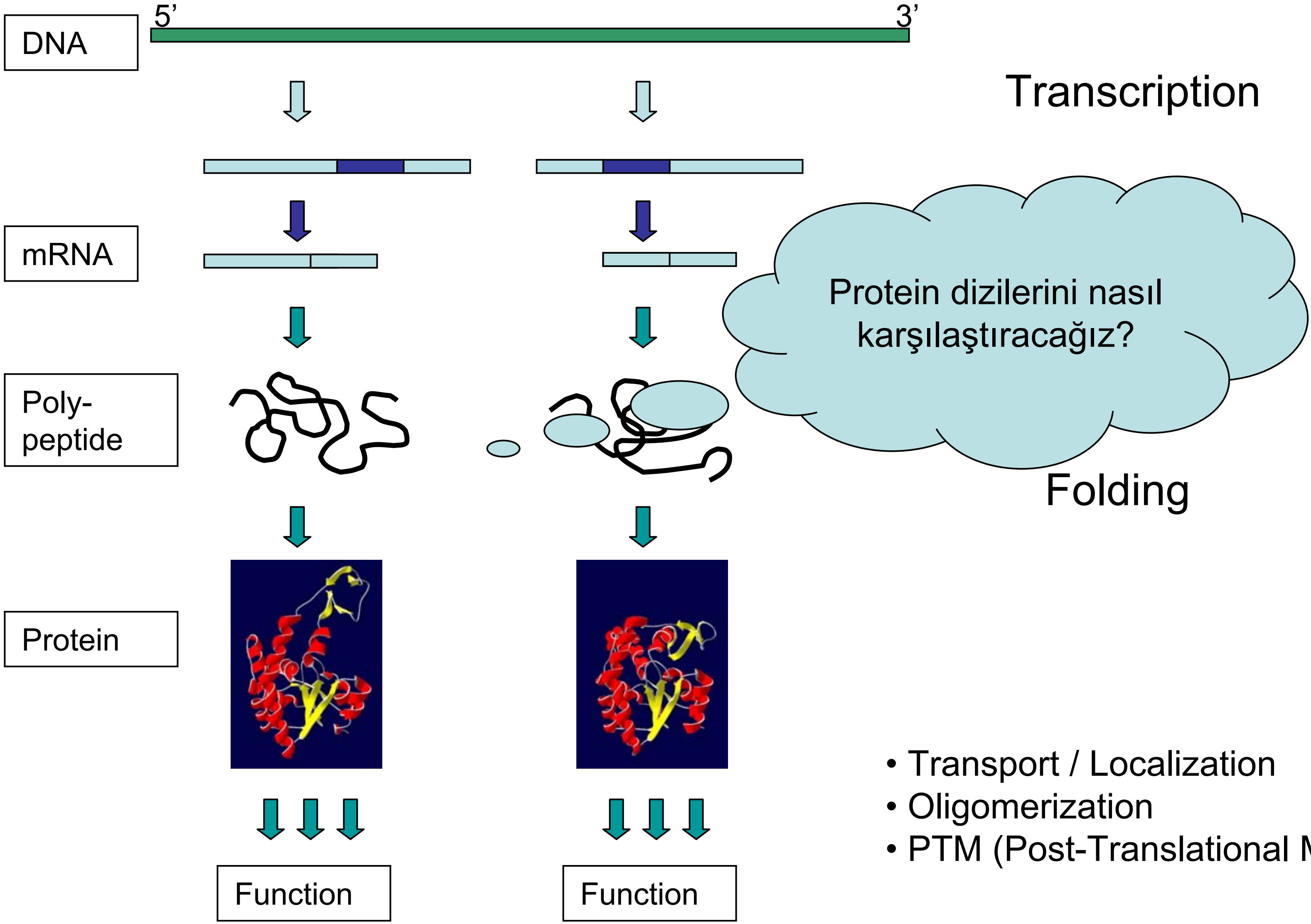


http://www.expasy.org/

moleküler biyoloji alanındaki bazı veritabanları...

AATDB, AceDb, ACUTS, ADB, AFDB, AGIS, AMSdb, ARR, AsDb, BBDB, BCGD, Beanref, Biolmage, BioMagResBank, BIOMDB, BLOCKS, BovGBASE, BOVMAP, BSORF, BTKbase, CANSITE, CarbBank, CARBHYD, CATH, CAZY, CCDC, CD4OLbase, CGAP, ChickGBASE, Colibri, COPE, CottonDB, CSNDB, CUTG, CyanoBase, dbCFC, dbEST, dbSTS, DDBJ, DGP, DictyDb, Picty_cDB, DIP, DOGS, DOMO, DPD, DPlnteract, ECDC, ECGC, EC02DBASE, EcoCyc, EcoGene, EMBL, EMD db, ENZYME, EPD, EpoDB, ESTHER, FlyBase, FlyView, GCRDB, GDB, GENATLAS, Genbank, GeneCards, Genline, GenLink, GENOTK, GenProtEC, **GermOnline**, GIFTS, GPCRDB, GRAP, GRBase, gRNAsdb, GRR, GSDB, HAEMB, HAMSTERS, HEART-2DPAGE, HEXAdb, HGMD, HIDB, HIDC, HlVdb, HotMolecBase, HOVERGEN, HPDB, HSC-2DPAGE, ICN, ICTVDB, IL2RGbase, IMGT, Kabat, KDNA, KEGG, Klotho, LGIC, MAD, MaizeDb, MDB, Medline, Mendel, MEROPS, MGDB, MGI, MHCPEP5 Micado, MitoDat, MITOMAP, MJDB, MmtDB, Mol-R-Us, MPDB, MRR, MutBase, MycDB, NDB, NRSub, 0-lycBase, OMIA, OMIM, OPD, ORDB, OWL, PAHdb, PatBase, PDB, PDD, Pfam, PhosphoBase, PigBASE, PIR, PKR, PMD, PPDB, PRESAGE, PRINTS, ProDom, Prolysis, PROSITE, PROTOMAP, RatMAP, RDP, REBASE, RGP, SBASE, SCOP, SeqAnaiRef, SGD, SGP, SheepMap, Soybase, SPAD, SRNA db, SRPDB, STACK, StyGene,Sub2D, SubtiList, SWISS-2DPAGE, SWISS-3DIMAGE, **SWISS-MODEL Repository**, SWISS-PROT, TelDB, TGN, tmRDB, TOPS, TRANSFAC, TRR, UniGene, URNADB, V BASE, VDRR, VectorDB, WDCM, WIT,

Biyoinformatiğe neden ihtiyacımız var?



- Transport / Localization
- Oligomerization
- PTM (Post-Translational Modification)

Dizi karşılaştırma

```
1: MYTAILORISRICH

2: MONTAILLEURESTRICHE
```

```
1: MY-TAIL--ORIS-RICH-
   ¦x ¦¦¦¦  x¦x¦ ¦¦¦¦
2: MONTAILLEURESTRICHE
```

Global Alignment

```
1:    TAILO      RICH
      ¦¦¦¦x      ¦¦¦¦
2:    TAILL      RICHE
```

Two Local Alignments

¦ = Identity
x = Mismatch
- = Insertion / Deletion

Multiple Sequence Alignment

(MSA)

Programlar:

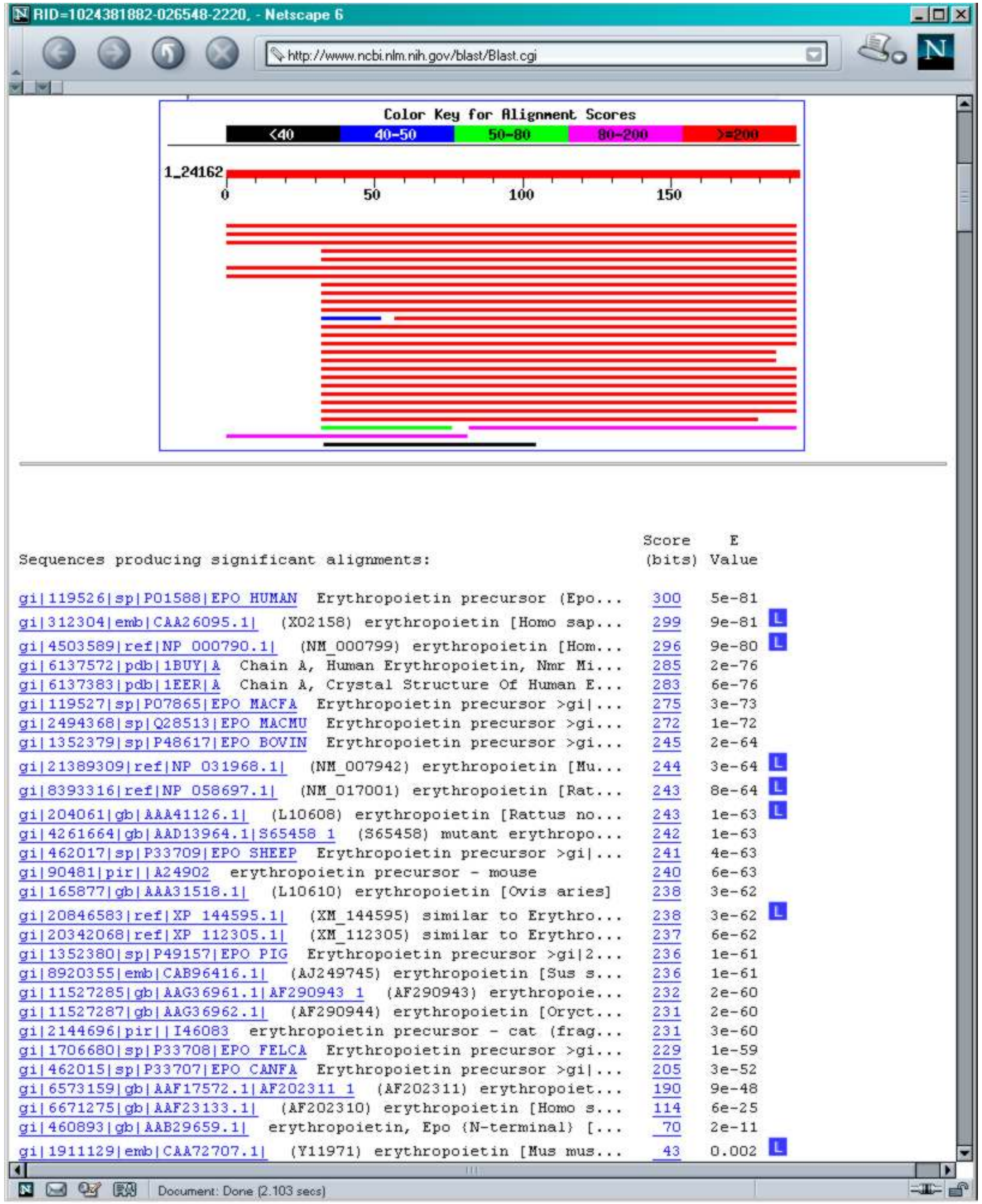
- CLUSTALW
- T_COFFEE
- MULTALIGN

```
HBA_CHICK   VL-SAADKNNVKGIFTKIAGHAEEYGAETLERMFTTYPPTKTYFPHF-DL    48
HBAD_CHICK  ML-TAEDKKLIQQAWEKAASHQEEFGAEALTRMFTTYPQTKTYFPHF-DL    48
HBPI_CHICK  AL-TQAEKAAVTTIWAKVATQIESIGLESLERLFASYPQTKTYFPHF-DV    48
HBB_CHICK   VHWTAEEKQLITGLWGKV--NVAECGAEALARLLIVYPWTQRFFASFGNL    48
HBE_CHICK   VHWSAEEKQLITSVWSKV--NVEECGAEALARLLIVYPWTQRFFASFGNL    48
HBRH_CHICK  VHWSAEEKQLITSVWSKV--NVEECGAEALARLLIVYPWTQRFFDNFGNL    48
MYG_CHICK   GL-SDQEWQQVLTIWGKVEADIAGHGHEVLMRLFHDHPETLDRFDKFKGL    49
               ....   .   ..*   . .. * * * *..  .* *   *  * ..

HBA_CHICK   SH-----GSAQIKGHGKKVVAALIEAANHIDDIAGTLSKLSDLHAHKLRV    93
HBAD_CHICK  SP-----GSDQVRGHGKKVLGALGNAVKNVDNLSQAMAELSNLHAYNLRV    93
HBPI_CHICK  SQ-----GSVQLRGHGSKVLNAIGEAVKNIDDIRGALAKLSELHAYILRV    93
HBB_CHICK   SSPTAILGNPMVRAHGKKVLTSFGDAVKNLDNIKNTFSQLSELHCDKLHV    98
HBE_CHICK   SSPTAIMGNPRVRAHGKKVLSSFGEAVKNLDNIKNTYAKLSELHCDKLHV    98
HBRH_CHICK  SSPTAIIGNPKVRAHGKKVLSSFGEAVKNLDNIKNTYAKLSELHCEKLHV    98
MYG_CHICK   KTPDQMKGSEDLKKHGATVLTQLGKILKQKGNHESELKPLAQTHATKHKI    99
            .      *.  .. ** .*.. . .  .. ..   .   *.. *    ..

HBA_CHICK   DPVNFKLLGQCFLVVVAIHHPAALTPEVHASLDKFLCAVGTVLTAKYR--   141
HBAD_CHICK  DPVNFKLLSQCIQVVLAVHMGKDYTPEVHAAFDKFLSAVSAVLAEKYR--   141
HBPI_CHICK  DPVNFKLLSHCILCSVAARYPSDFTPEVHAEWDKFLSSISSVLTEKYR--   141
HBB_CHICK   DPENFRLLGDILIIVLAAHFSKDFTPECQAAWQKLVRVVAHALARKYH--   146
HBE_CHICK   DPENFRLLGDILIIVLASHFARDFTPACQFAWQKLVNVVAHALARKYH--   146
HBRH_CHICK  DPENFRLLGNILIIVLAAHFTKDFTPTCQAVWQKLVSVVAHALAYKYH--   146
MYG_CHICK   PVKYLEFISEVIIKVIAEKHAADFGADSQAAMKKALELFRNDMASKYKEF   149
                . .... .   .* .   . ... .   .* .      .. **.

HBA_CHICK   ----   141
HBAD_CHICK  ----   141
HBPI_CHICK  ----   141
HBB_CHICK   ----   146
HBE_CHICK   ----   146
HBRH_CHICK  ----   146
MYG_CHICK   GFQG   153

Consensus length:  154; Identity  : 19 ( 12.3%); Similarity: 51 ( 33.1%)
Character to show that a position in the alignment is perfectly conserved: '*'
Character to show that a position is well conserved: '.'
```

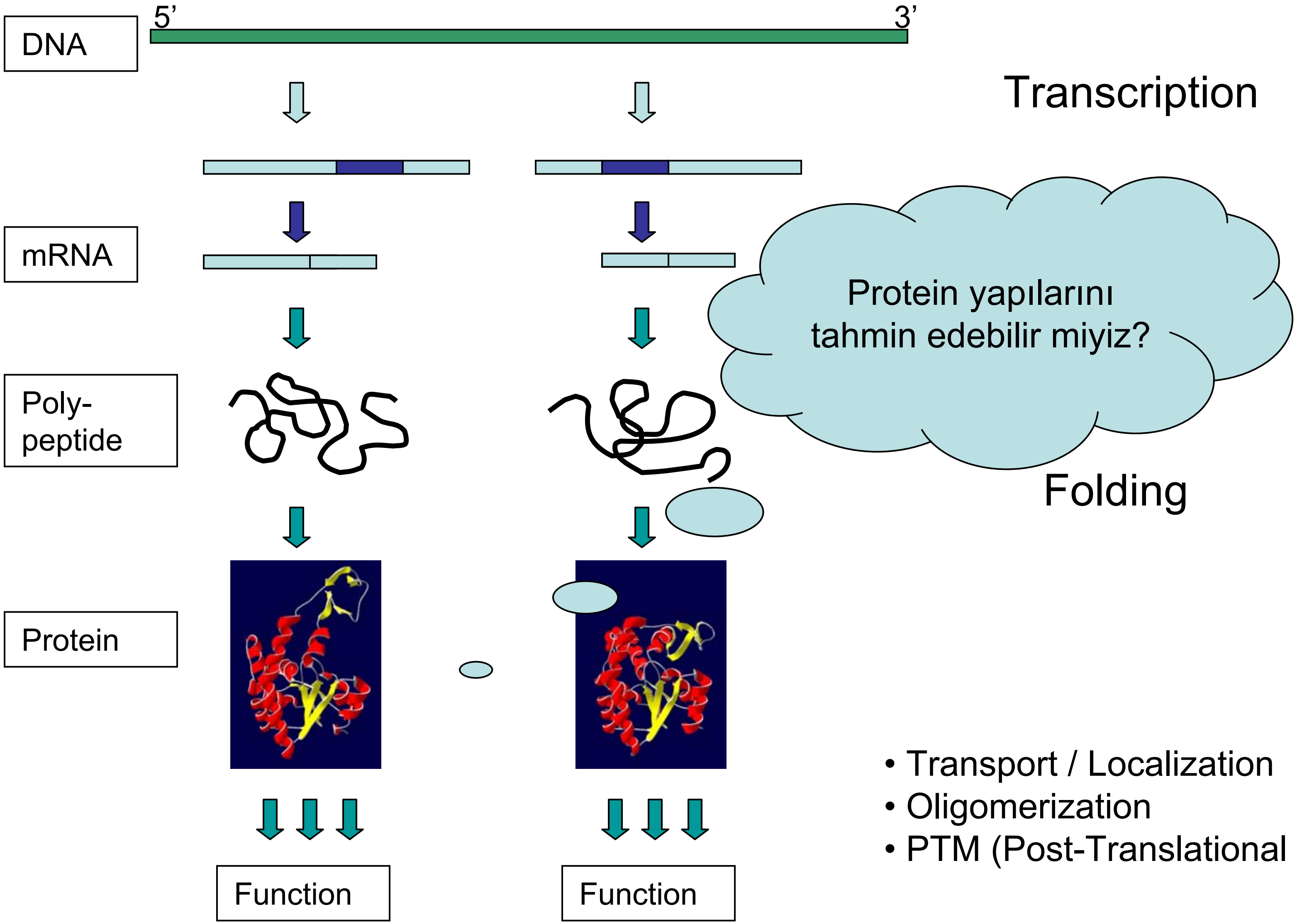
NCBI BLAST http://www.ncbi.nlm.nih.gov/blast/

BLAST:

Basic
Local
Alignment
Search
Tool



Biyoinformatiğe neden ihtiyacımız var?



- Transport / Localization
- Oligomerization
- PTM (Post-Translational Modification)

Protein Yapısı Modelleme

- *Ab initio* modeling
- Threading & Fold Recognition
- Homology Modeling

```
MNIFEMLRID EGLRLKIYKD TEGYYTIGIG
HLLTKSPSLN AAKSELDKAI GRNCNGVITK
DEAEKLFNQD VDAAVRGILR NAKLKPVYDS
LDAVRRCALI NMVFQMGETG VAGFTNSLRM
LQQKRWDEAA VNLAKSRWYN QTPNRAKRVI
TTFRTGTWDA YKNL
```