

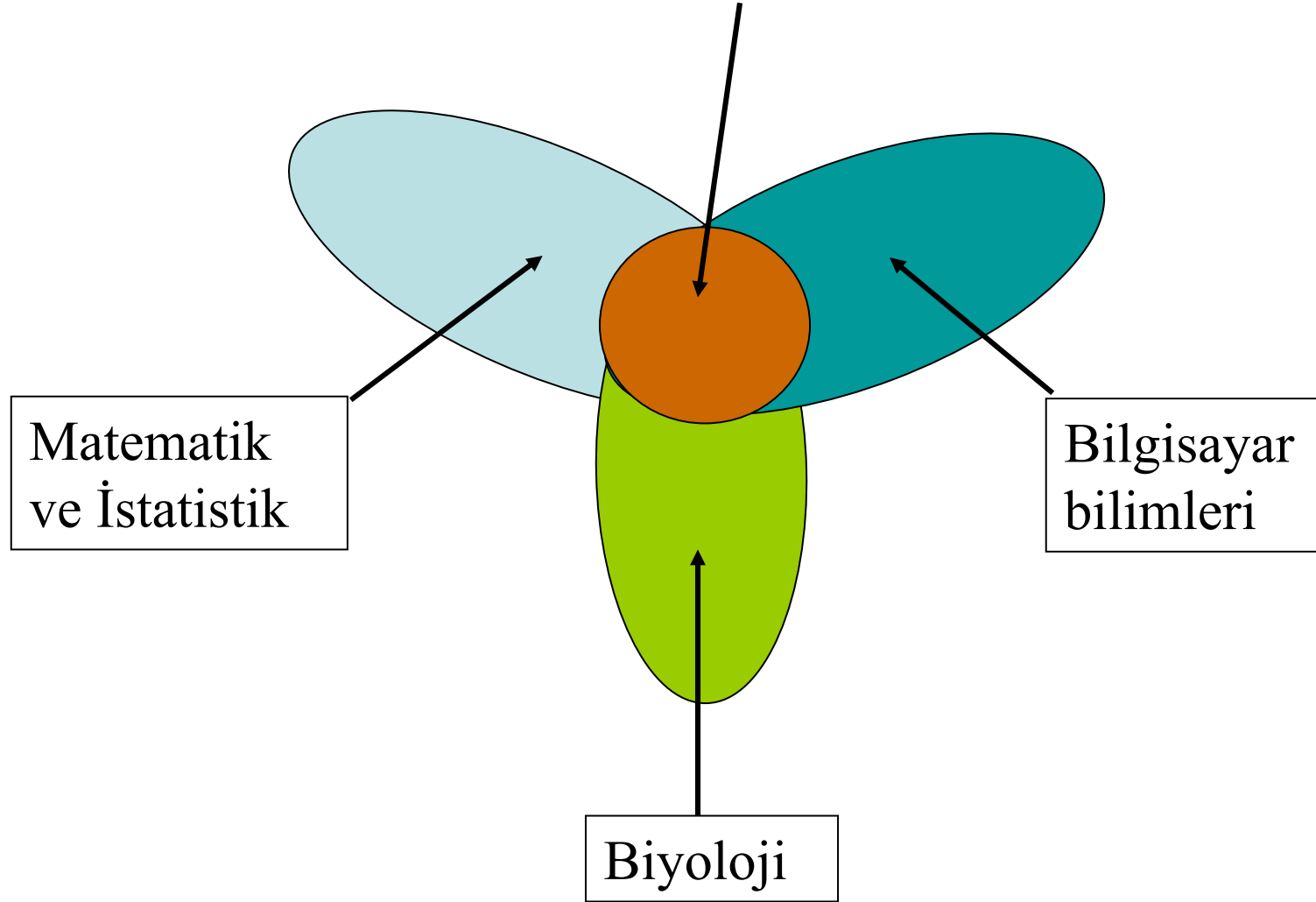
Biyoinformatik

- 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



İki temel bilgi vardır:

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

- Deneyssel 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 araştırılması
- 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 alıřmalar

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

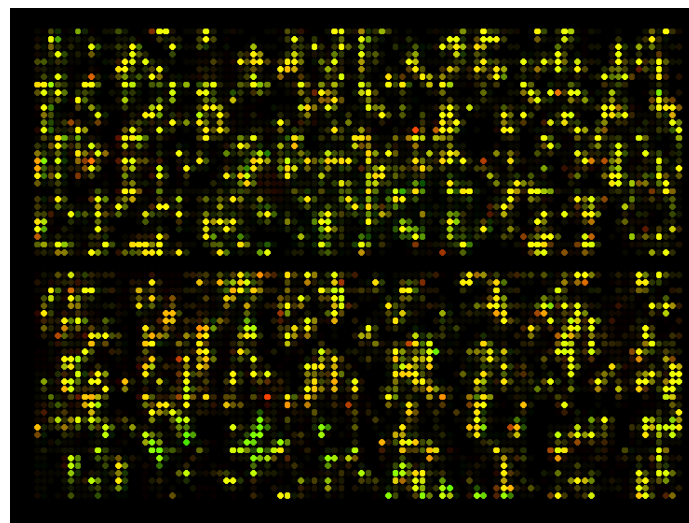
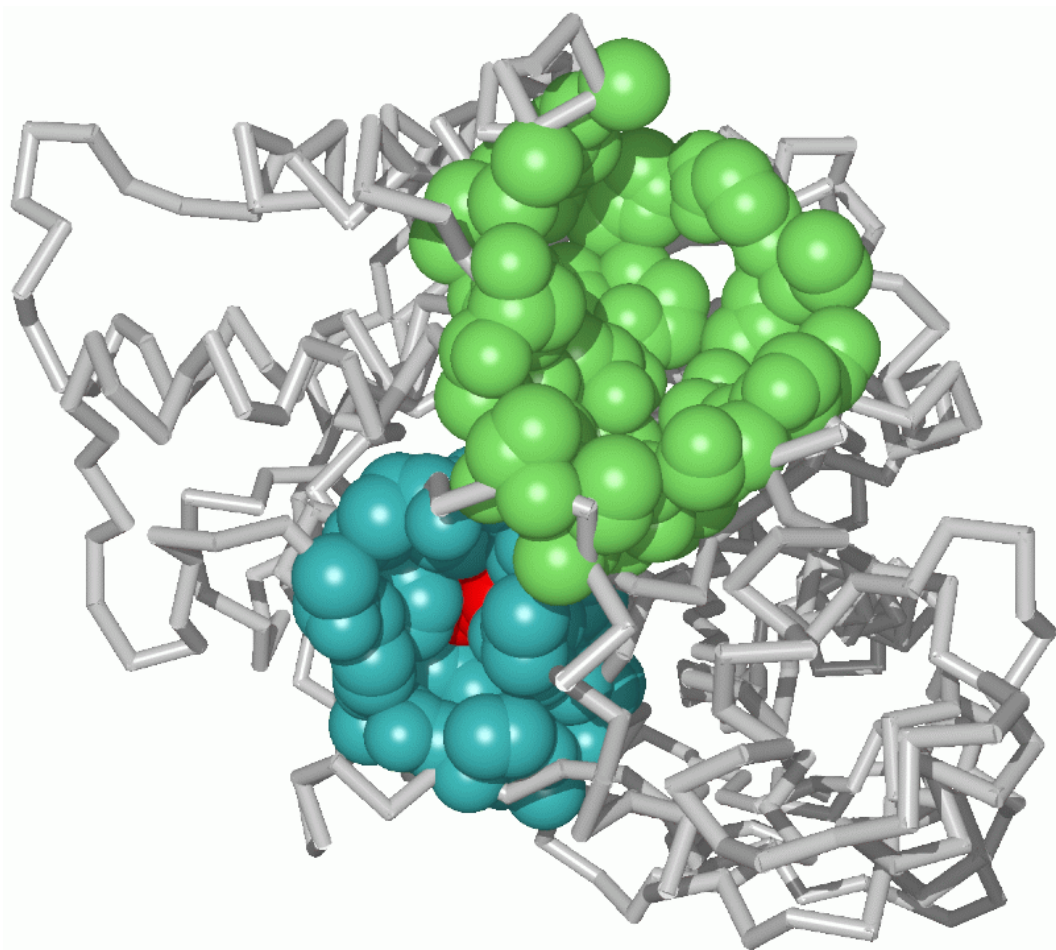
- Genetik bilgi:

DNA  protein

Biyoinformatik veritabanlarının kısa tarihçesi:

- ilk biyolojik veritabanı - **P**rotein **I**dentification **R**esource 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.....

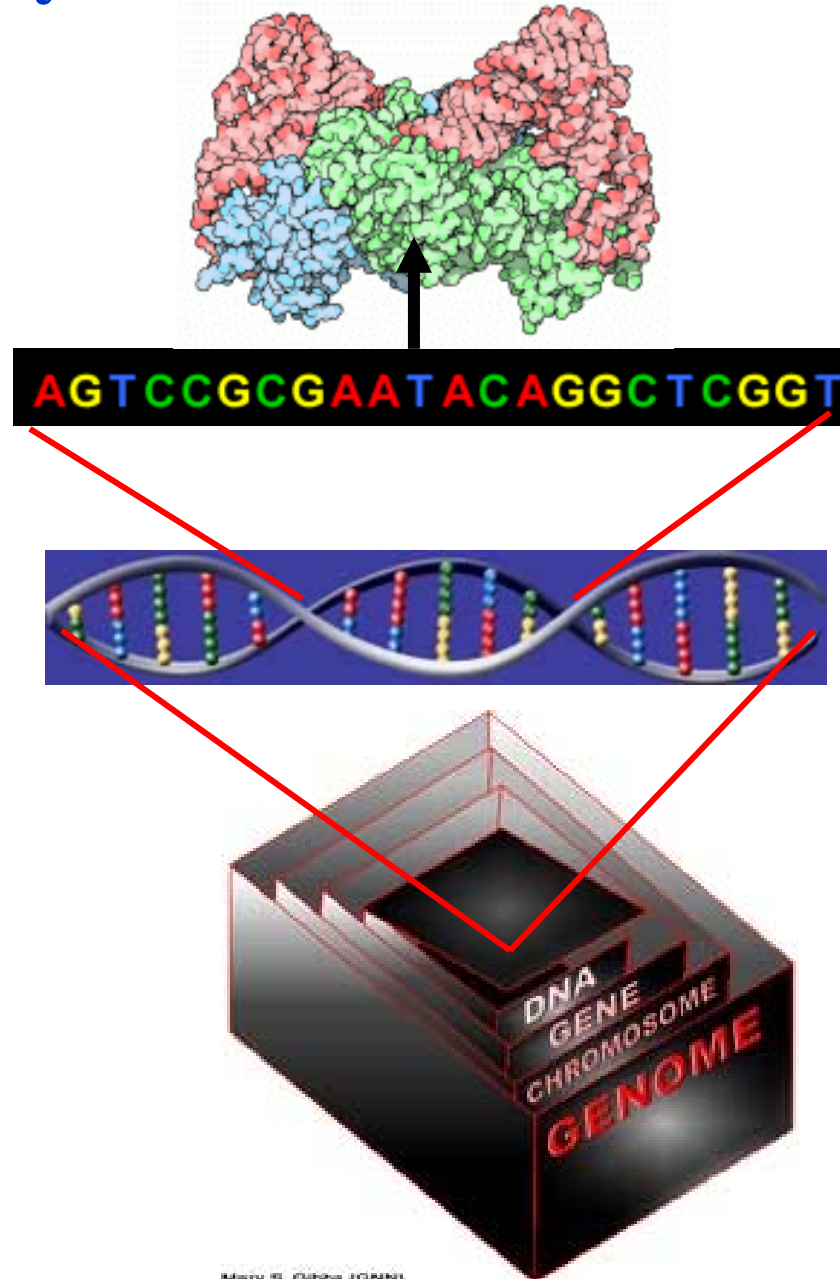


Gen ekspresyonu = Protein üretimi

Protein

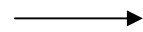
Nukleotid
dizisi

DNA



Biyolojik Çeşitlilik

Bakteriler



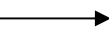
Organism

Number of
genes in
the genome

*Mycoplasma
genitalium*

517

Mayalar



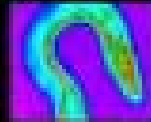
*Saccharomyces
cerevisiae*

6,275



*Arabidopsis
thaliana*

~ 20,000



*Caenorhabditis
elegans*

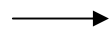
19,099



*Haemophilus
influenzae*

1,743

Meyva
sineği



*Drosophila
melanogaster*

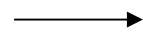
13,601



*Neisseria
meningitidis*

2,158

İnsan



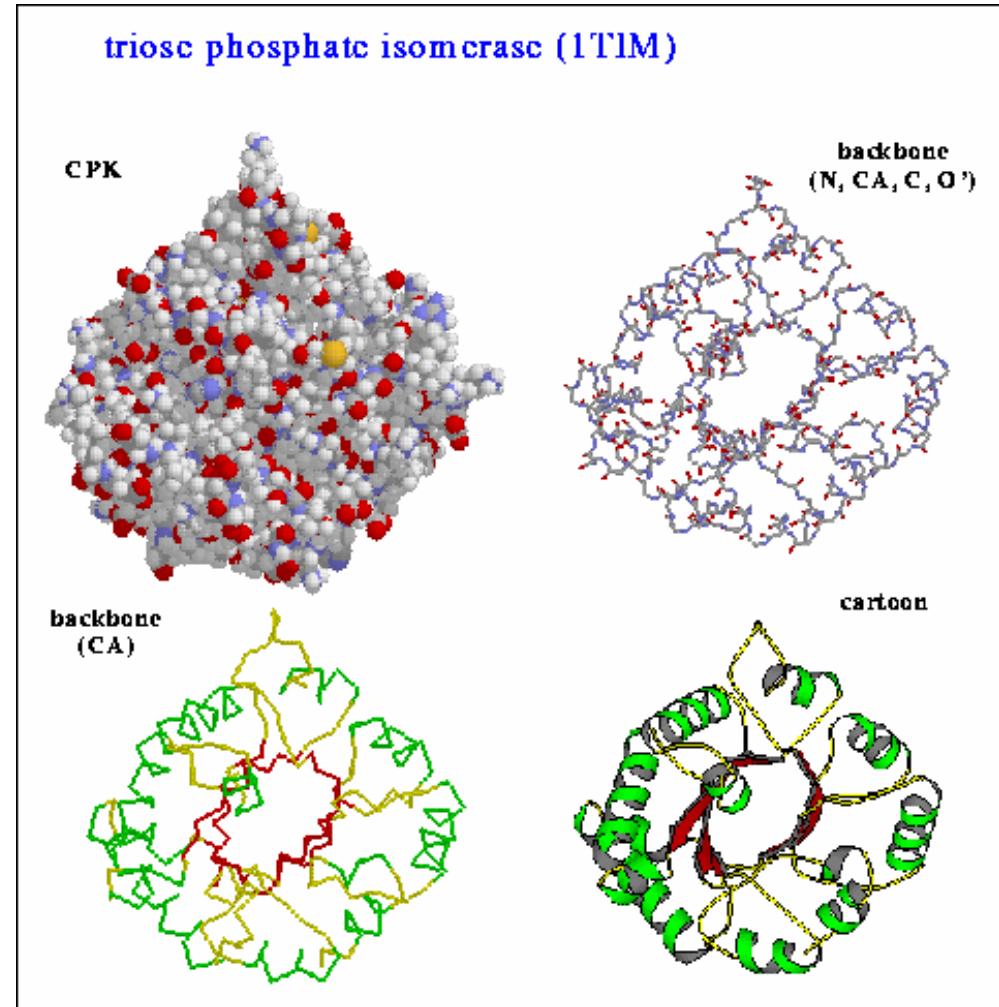
Homo sapiens

~ 30,000

Protein dizisinden Yapısına

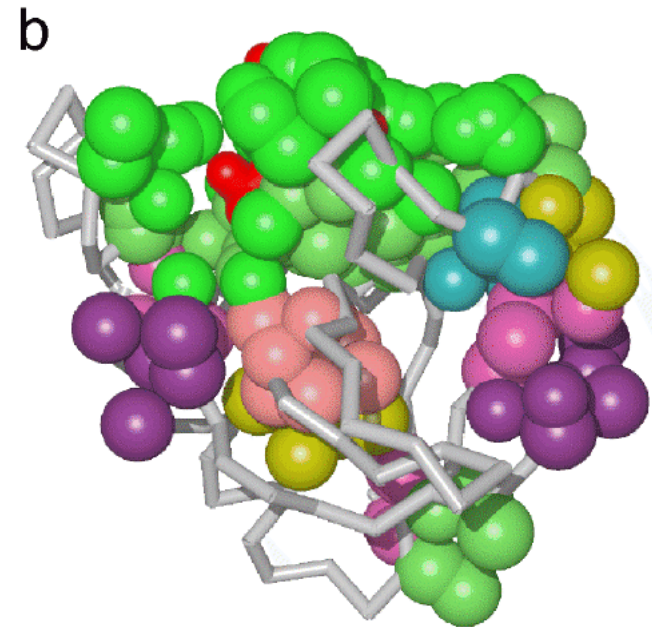
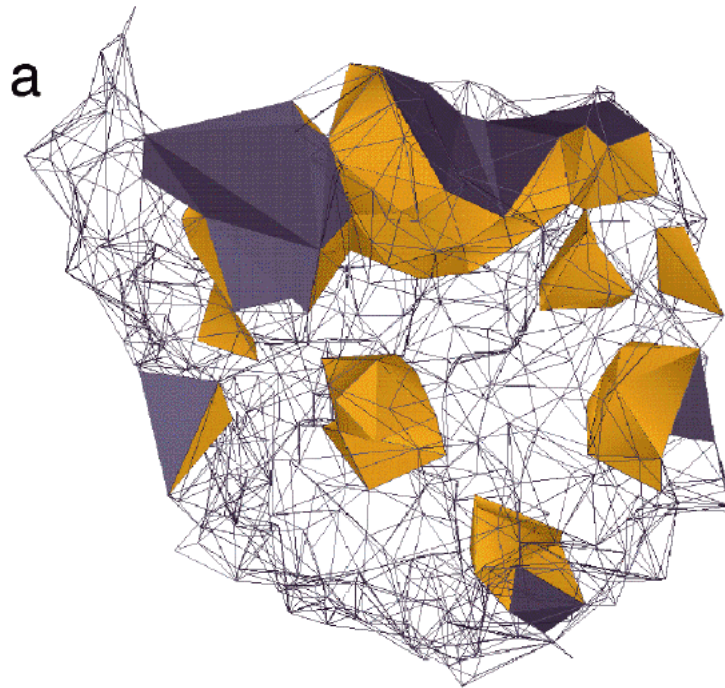
APRKFFVGGNWKMNGD
KKSLGELIHTLNGAKL
SADTEVVCGAPSIYLD
FARQKLDKIGVAAQN
CYKVPKGAFTGEISPA
MIKDIGAAWVILGHSE
RRHVFGESDELIGQKV
AHALAEGLGVIACIGE
KLDEREAGITEKVVFE
QTKAIADNVKDWSKVV
LAYEPVWAIGTGKTAT
PQQAQEVHEKLRGWLK
SHVSDAVAQSTRIIYG
GSVTGGNCKELASQHD
VDGFLVGGASLKPEFV
DIINAKH

=



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

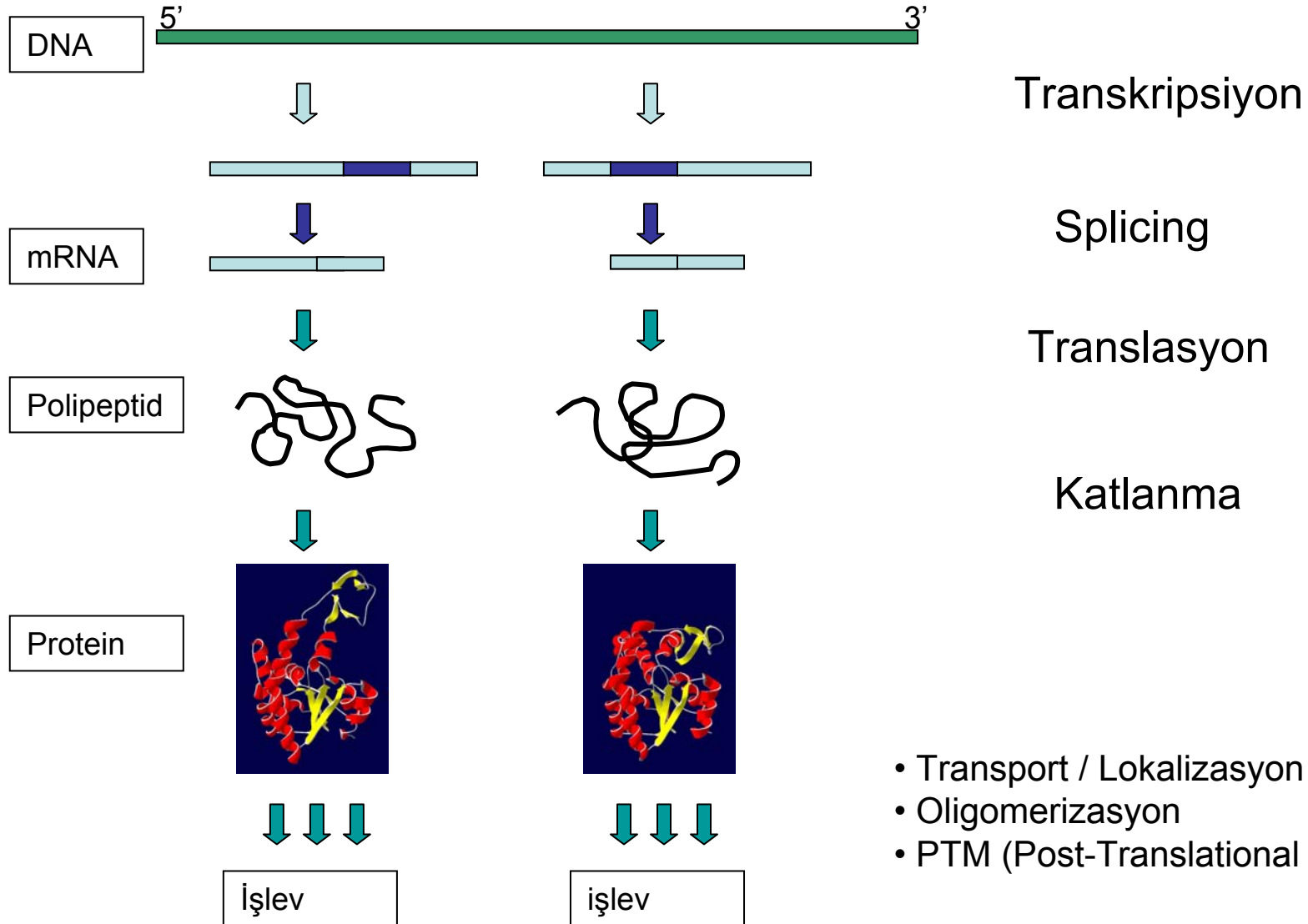
Pockets in Ribonuclease A



Biyoinformatik

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

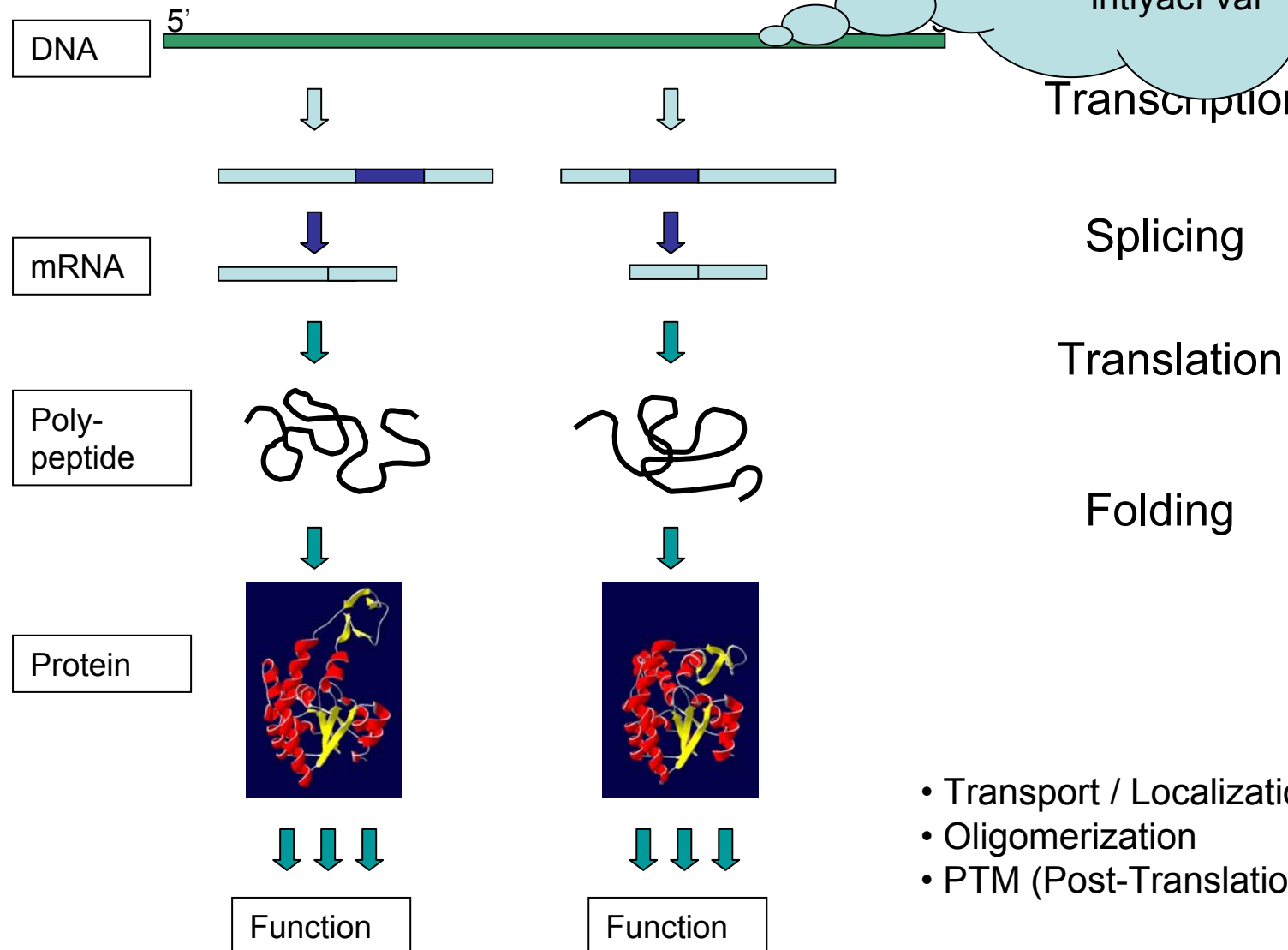
Biyoinformatiğe neden ihtiyacımız var?



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

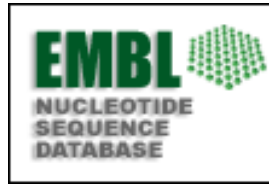
Biyoinformatiğe neden ihtiyacımız var?

Genom projelerinden
gelen DNA dizilerinin
saklanmaya ve
organize edilmeye
ihtiyacı var



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

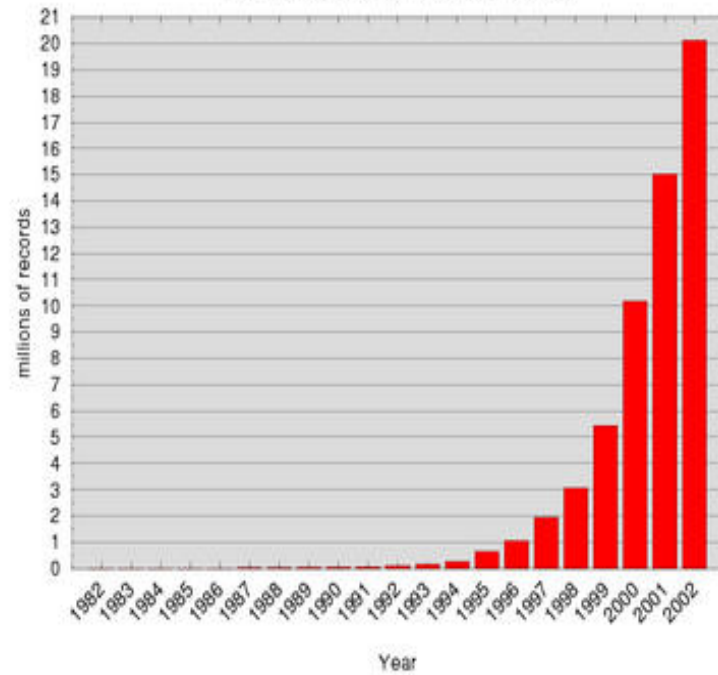
DNA veribankaları



- 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/>

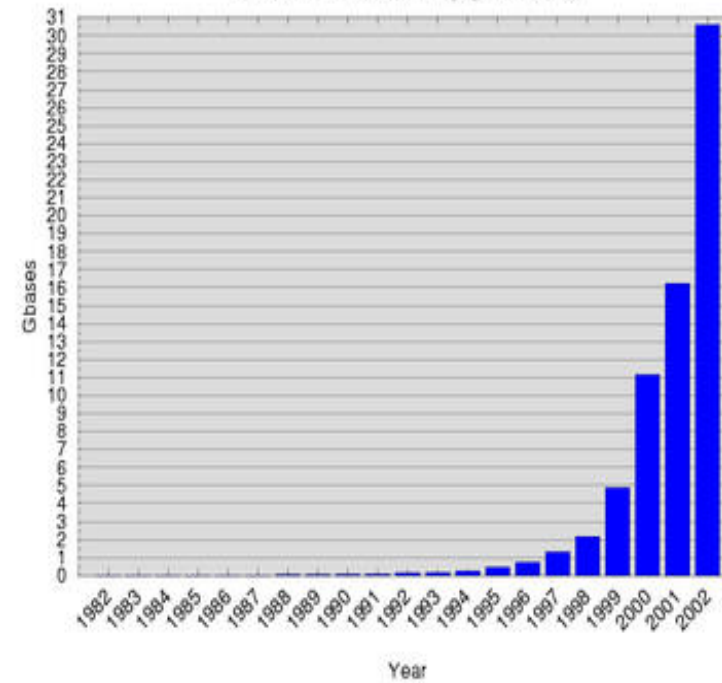
EMBL Database Growth

total record number (millions)

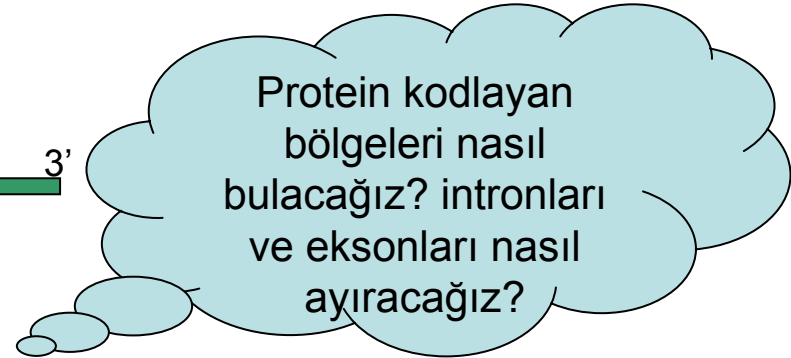
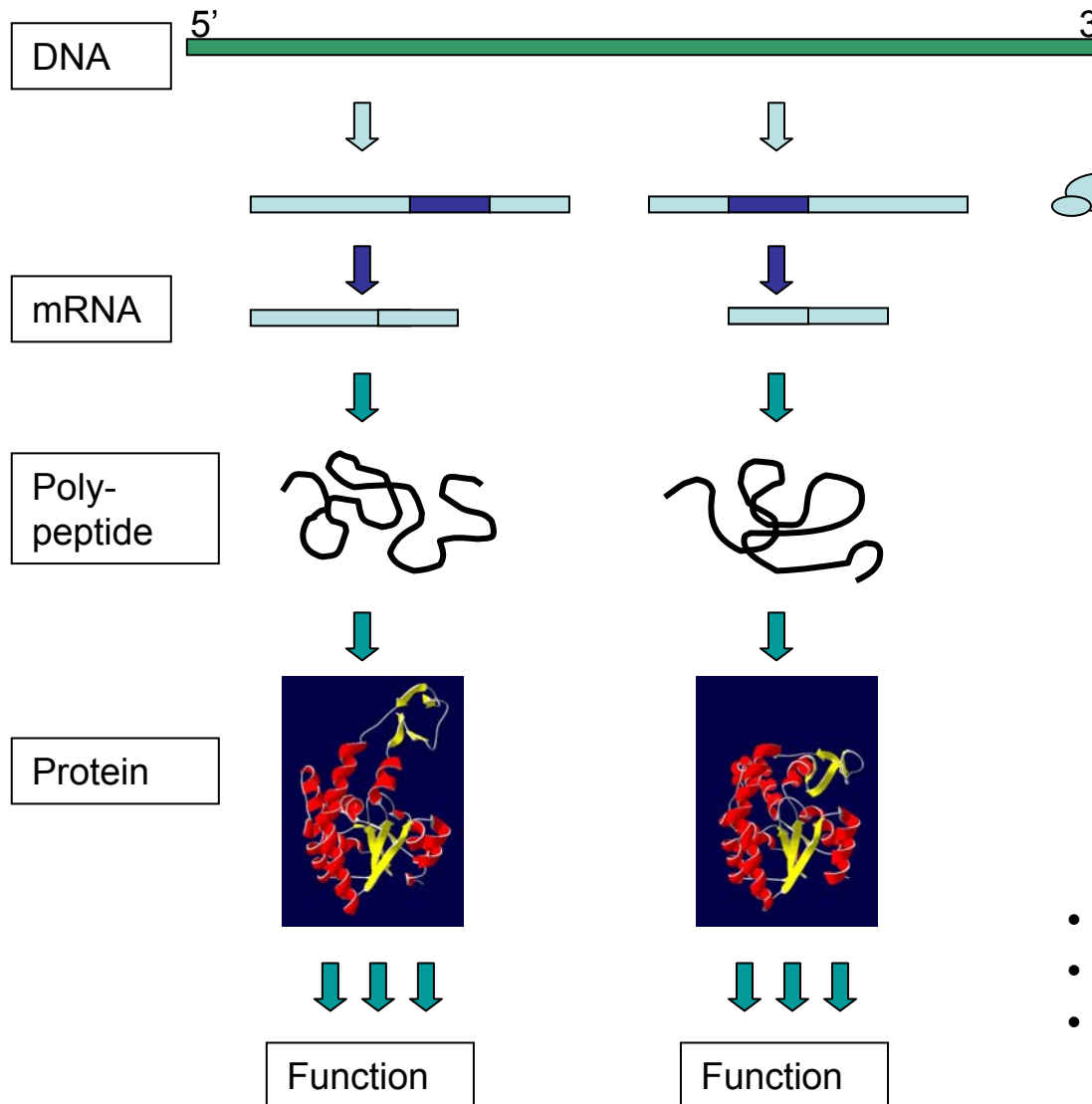


EMBL Database Growth

total nucleotides (gigabases)



Biyoinformatiğe neden ihtiyacımız var?



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:

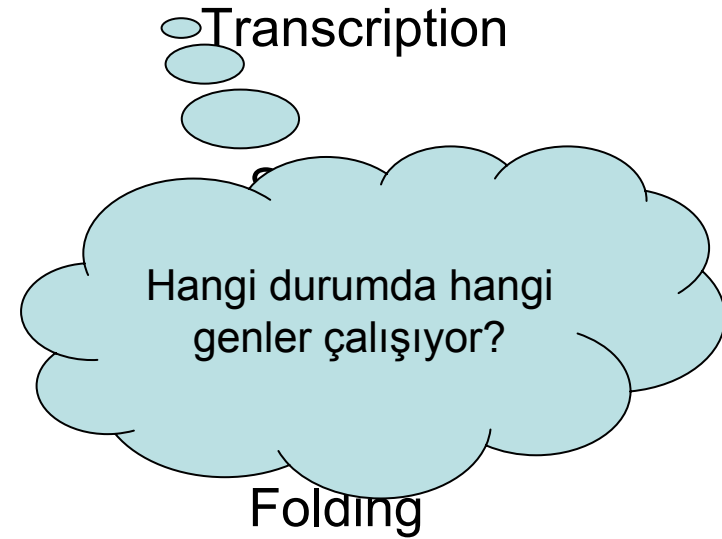
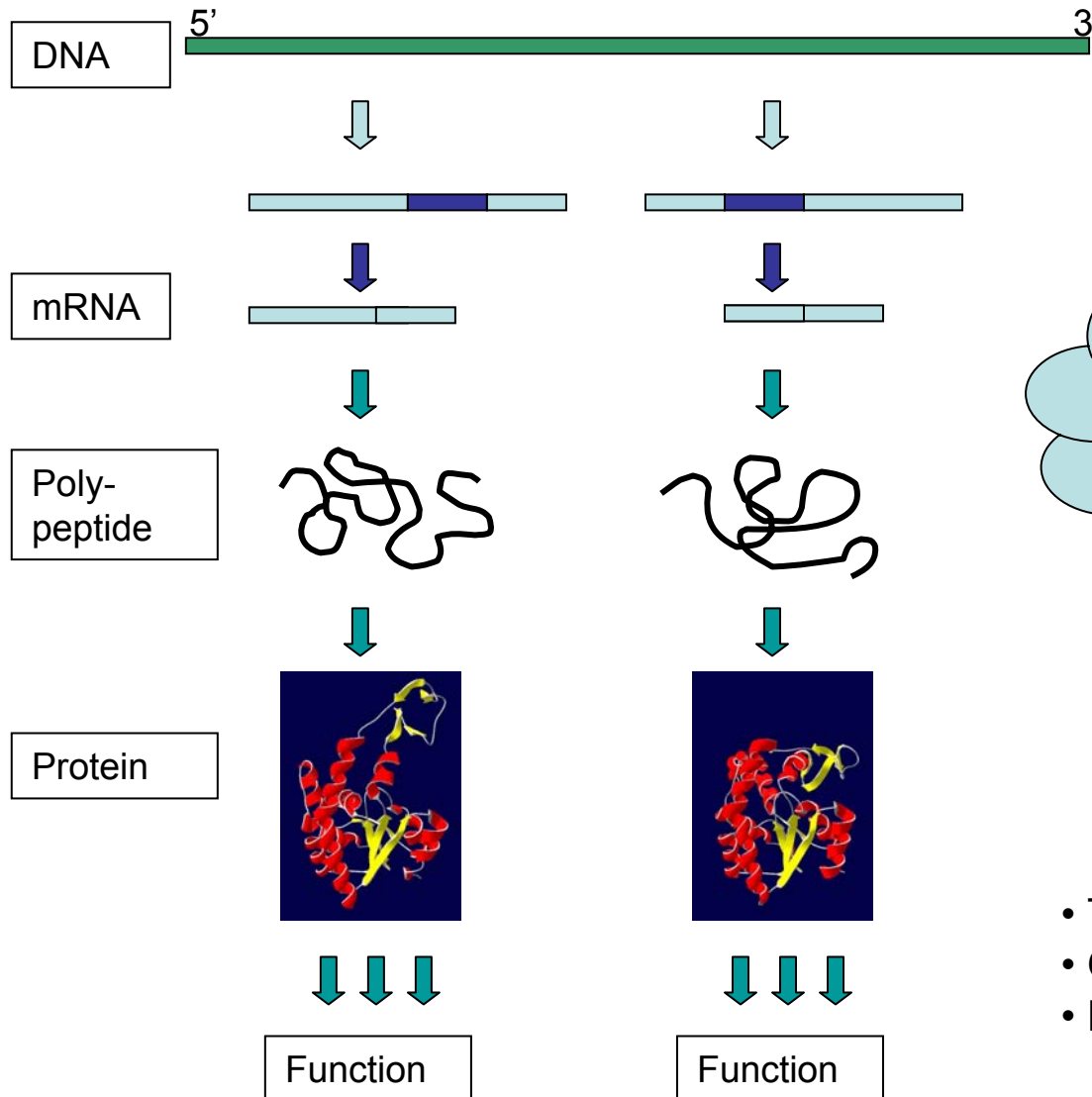
- Bilinen proteinlere benzerlik

Ökaryotik genomlarda genler nasıl bulunur?

Bazı problemler:

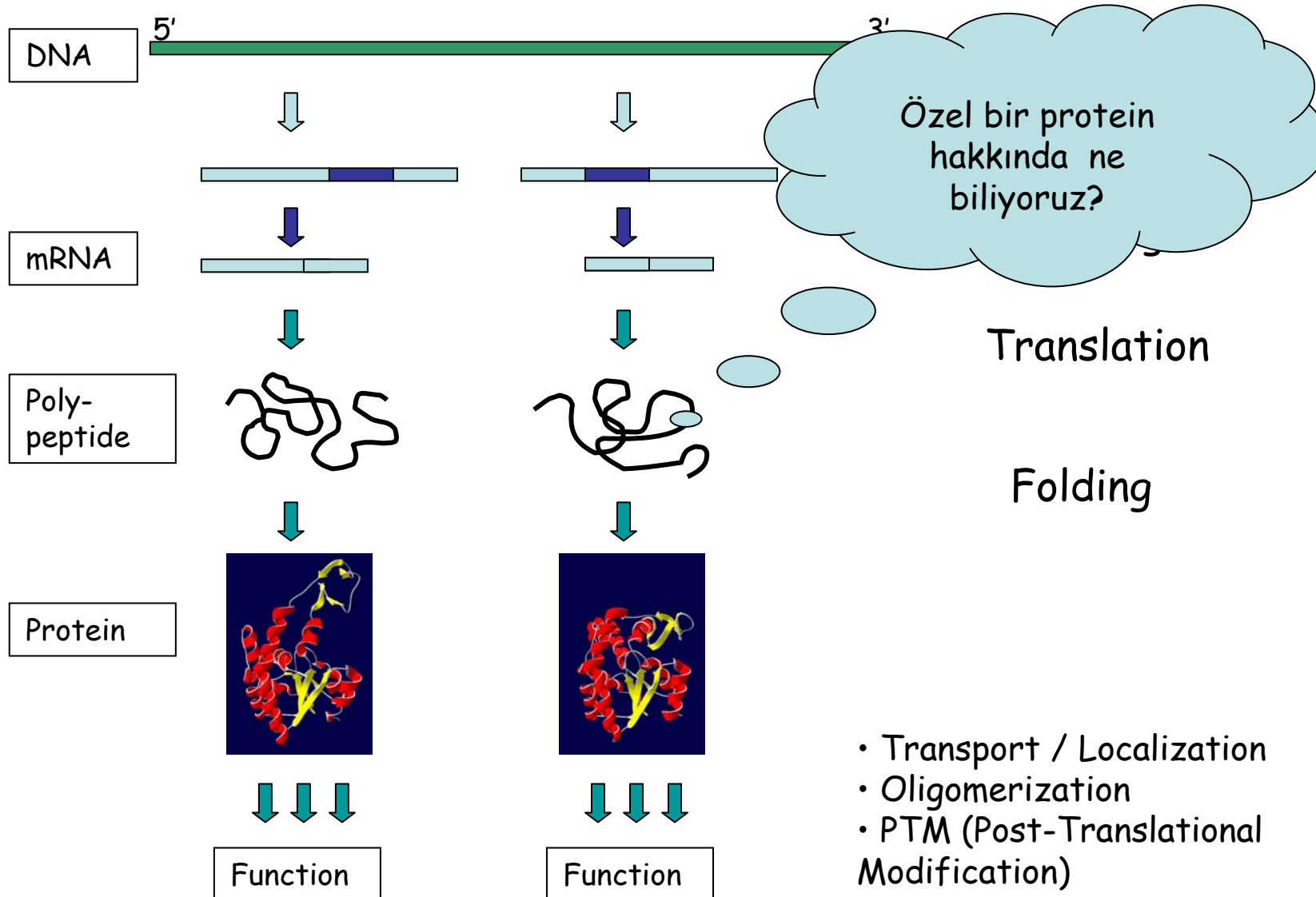
- Genomun kompleksliği organizmanın kompleksliği ile orantılıdır
- 1-2 % kodlama yapan bölge
- Intron – exon yapıları
- Alternatif kırılma (bütün genlerin ~30%)
- Pseudogenler

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
- Dokümantasyon

SWISS-PROT/TrEMBL



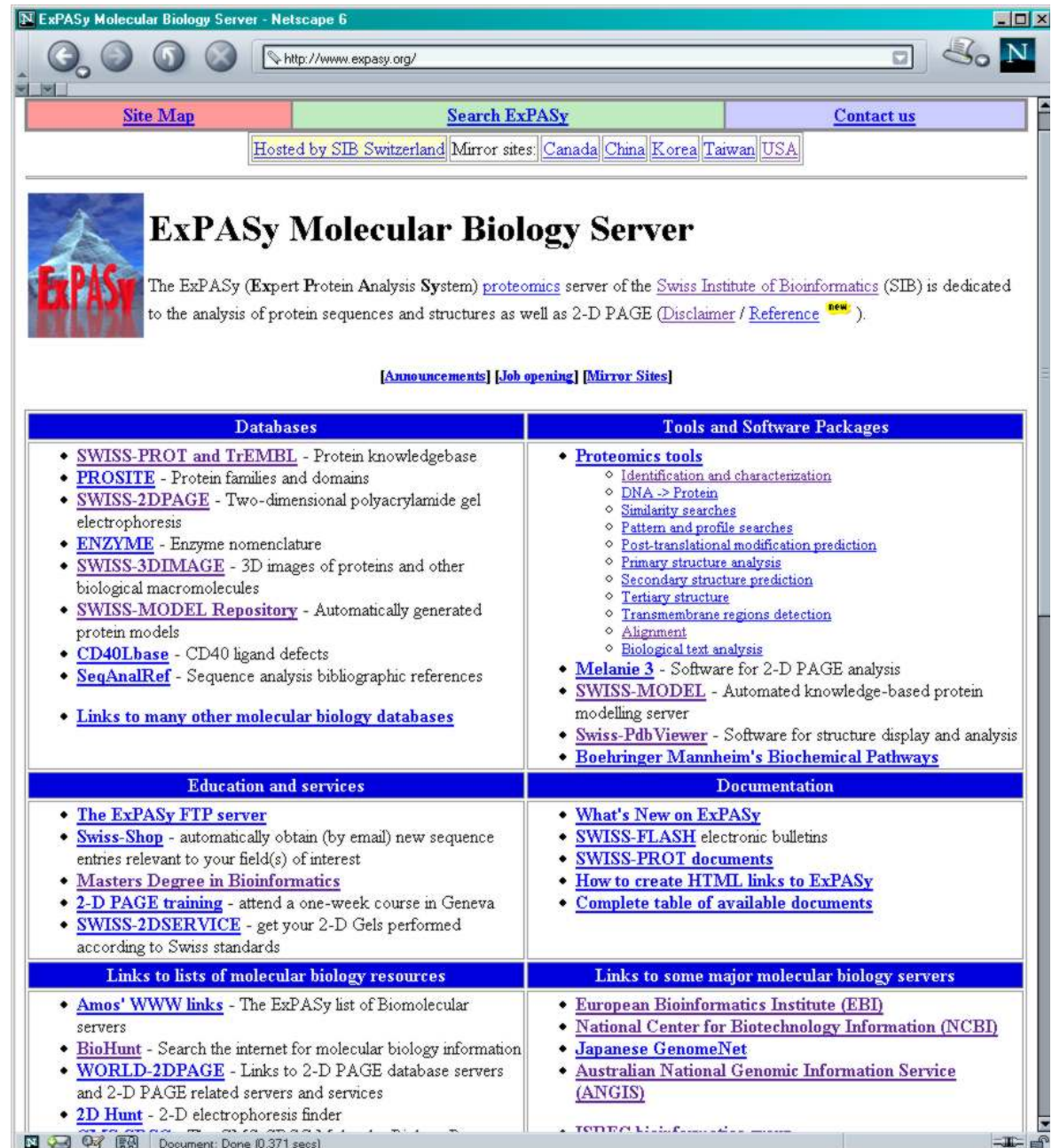
- the SIB (CH) ve EMBL/EBI (UK) arası ortaklık
- SWISS-PROT: protein dizisi veritabanı.
- TrEMBL

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

ExPASy Web Server

ExPASy =

Expert
Protein
Analysis
System



ExPASy Molecular Biology Server

The ExPASy (Expert Protein Analysis System) [proteomics](#) server of the [Swiss Institute of Bioinformatics](#) (SIB) is dedicated to the analysis of protein sequences and structures as well as 2-D PAGE ([Disclaimer](#) / [Reference](#) ^{new}).

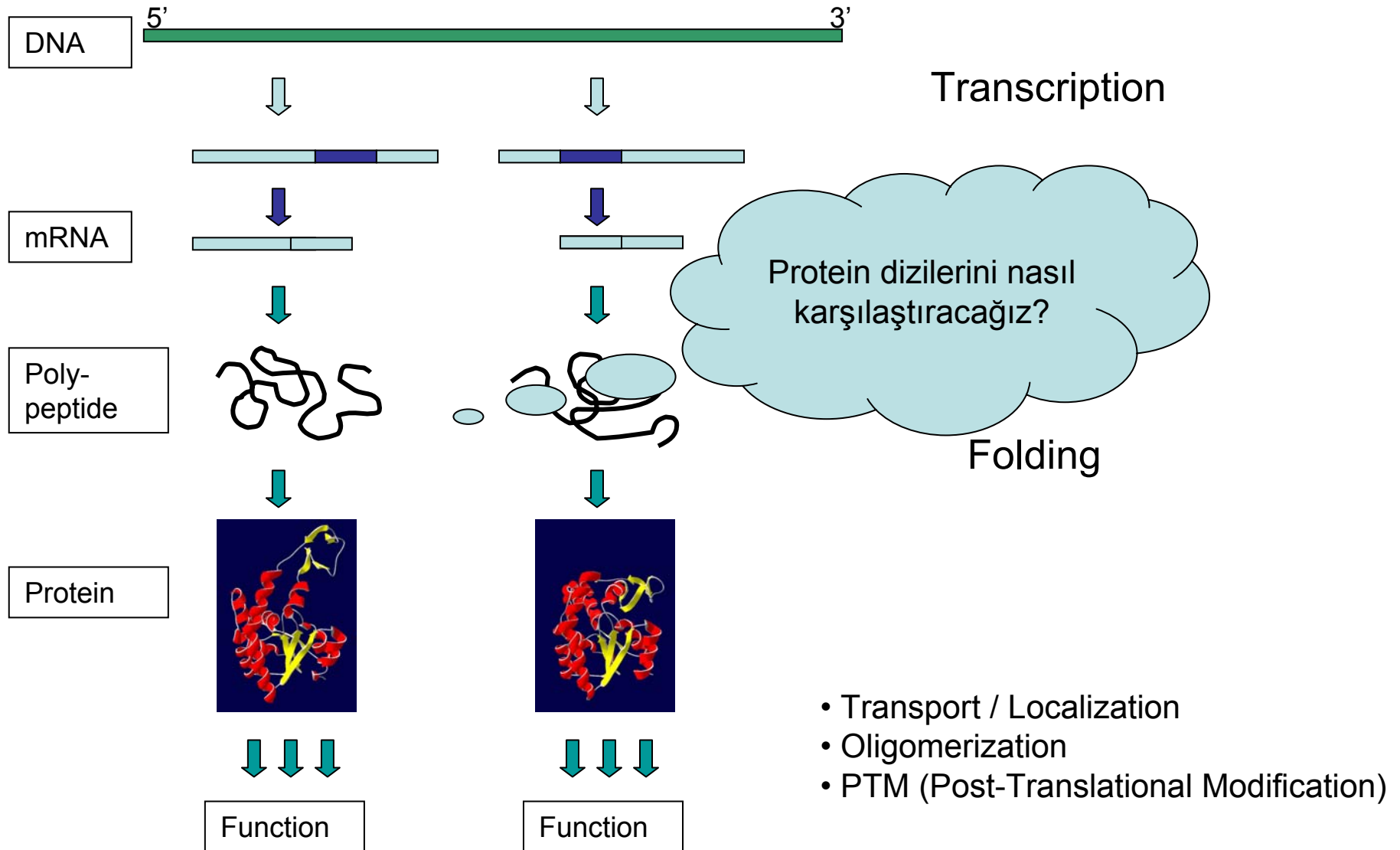
[\[Announcements\]](#) [\[Job opening\]](#) [\[Mirror Sites\]](#)

Databases	Tools and Software Packages
<ul style="list-style-type: none"> • SWISS-PROT and TrEMBL - Protein knowledgebase • PROSITE - Protein families and domains • SWISS-2DPAGE - Two-dimensional polyacrylamide gel electrophoresis • ENZYME - Enzyme nomenclature • SWISS-3DIMAGE - 3D images of proteins and other biological macromolecules • SWISS-MODEL Repository - Automatically generated protein models • CD40Lbase - CD40 ligand defects • SeqAnalRef - Sequence analysis bibliographic references • Links to many other molecular biology databases 	<ul style="list-style-type: none"> • Proteomics tools <ul style="list-style-type: none"> ◦ Identification and characterization ◦ DNA -> Protein ◦ Similarity searches ◦ Pattern and profile searches ◦ Post-translational modification prediction ◦ Primary structure analysis ◦ Secondary structure prediction ◦ Tertiary structure ◦ Transmembrane regions detection ◦ Alignment ◦ Biological text analysis • Melanie 3 - Software for 2-D PAGE analysis • SWISS-MODEL - Automated knowledge-based protein modelling server • Swiss-PdbViewer - Software for structure display and analysis • Boehringer Mannheim's Biochemical Pathways
Education and services	Documentation
<ul style="list-style-type: none"> • The ExPASy FTP server • Swiss-Shop - automatically obtain (by email) new sequence entries relevant to your field(s) of interest • Masters Degree in Bioinformatics • 2-D PAGE training - attend a one-week course in Geneva • SWISS-2DSERVICE - get your 2-D Gels performed according to Swiss standards 	<ul style="list-style-type: none"> • What's New on ExPASy • SWISS-FLASH electronic bulletins • SWISS-PROT documents • How to create HTML links to ExPASy • Complete table of available documents
Links to lists of molecular biology resources	Links to some major molecular biology servers
<ul style="list-style-type: none"> • Amos' WWW links - The ExPASy list of Biomolecular servers • BioHunt - Search the internet for molecular biology information • WORLD-2DPAGE - Links to 2-D PAGE database servers and 2-D PAGE related servers and services • 2D Hunt - 2-D electrophoresis finder 	<ul style="list-style-type: none"> • European Bioinformatics Institute (EBI) • National Center for Biotechnology Information (NCBI) • Japanese GenomeNet • Australian National Genomic Information Service (ANGIS) • ISDPS - Information System for Data Processing and Storage

■ 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, O-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?



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:

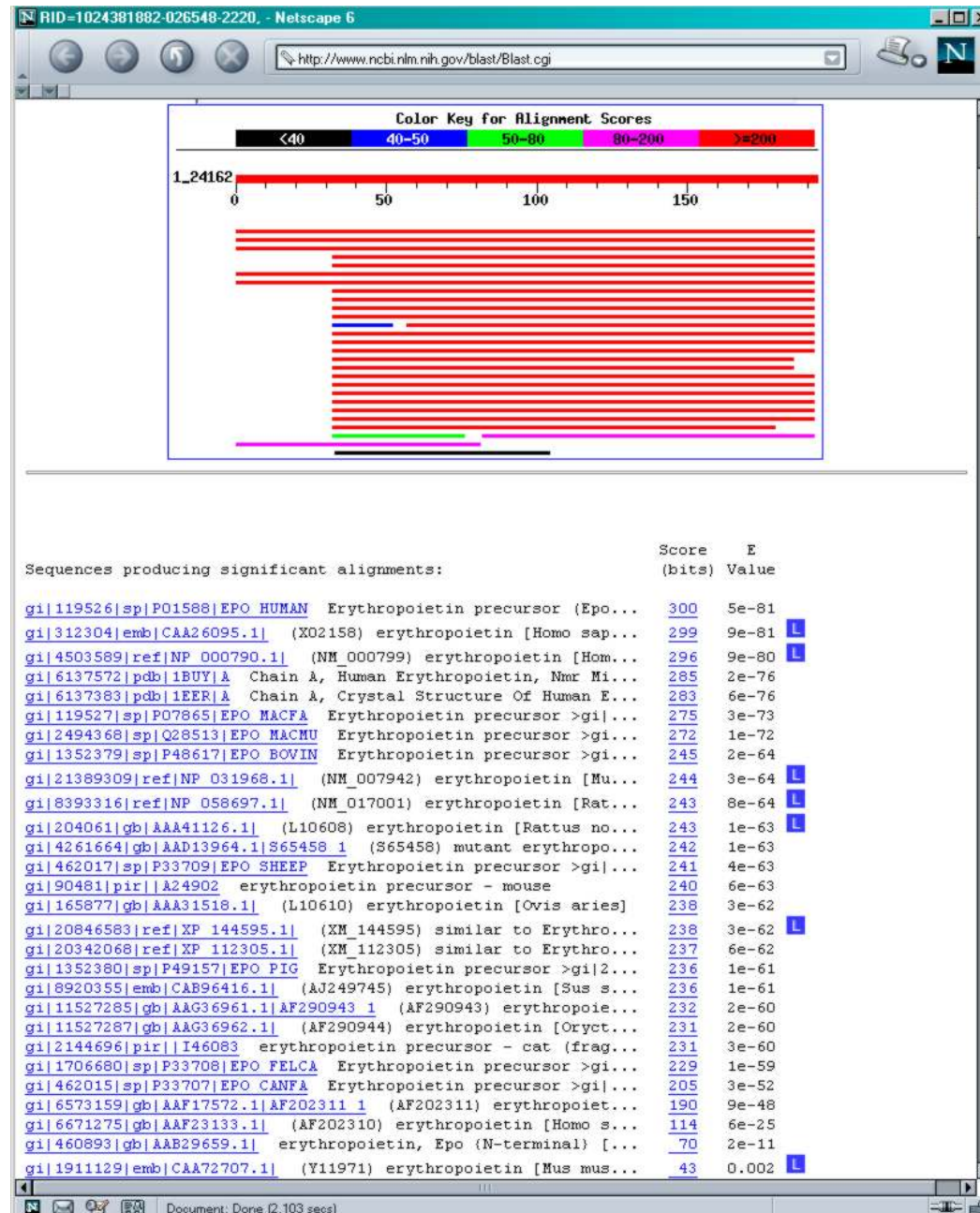
- CLUSTALW
- T_COFFEE
- MULTALIGN

HBA_CHICK	VL-SAADKNNVKGIFTKIAGHAE EYGAETLERMFTTYPPTKTYFPHF-DL	48
HBAD_CHICK	ML-TAEDKKLIQQAW EKAASHQE EFGAEALTRMFTTYPQTKTYFPHF-DL	48
HBPI_CHICK	AL-TQAEKAAVTTIWAKVATQIESIGLES LERLFASYPQTKTYFPHF-DV	48
HBB_CHICK	VHWTAE EQLITGLWGKV--NVAECGAEALARLLIVYPWTQRFFASFGNL	48
HBE_CHICK	VHWSAE EQLITSVWSKV--NVEECGAEALARLLIVYPWTQRFFASFGNL	48
HBRH_CHICK	VHWSAE EQLITSVWSKV--NVEECGAEALARLLIVYPWTQRFFDNFGNL	48
MYG_CHICK	GL-SDQEWQQVLTIWGKVEADIAGHGHEVLMRLFHDHPETLDRFDKFKGL	49
* . . . * * * * . . * * * * ..	
HBA_CHICK	SH-----GSAQIKGHGKKVVAALIEAANHIDDIAGTLSKLSDLHAHKLRV	93
HBAD_CHICK	SP-----GSDQVRGHGKKVLGALGNAVKNVDNLSQAMAELSNLHAYNLRV	93
HBPI_CHICK	SQ-----GSVQLRGHGSKVLNAIGEAVKNIDDIRGALAKLSELHAYILRV	93
HBB_CHICK	SSPTAILGNPMVRAHGK KVLTSFGDAVKNLDNIKNTFSQLSELHCDKLHV	98
HBE_CHICK	SSPTAIMGNPRVRAHGK KVLSSFGEAVKNLDNIKNTYAKLSELHCDKLHV	98
HBRH_CHICK	SSPTAIIGNPKVRAHGK KVLSSFGEAVKNLDNIKNTYAKLSELHCEKLHV	98
MYG_CHICK	KTPDQMKGSEDLKKHGATVLTQLGKILKQKGNHESELKPLAQTHATKHKI	99
	. * . . . ** . * * . . * . .	
HBA_CHICK	DPVNFKLLGQCFLVVVAIHHPAALTPEVHASLDKFLCAVGTVLTAKYR--	141
HBAD_CHICK	DPVNFKLLSQCIQVVLAVHMGKDYTPEVHAAFDKFLSAVSAVLAEKYR--	141
HBPI_CHICK	DPVNFKLLSHCILCSVAARYPSDFTPEVHAEWDKFLSSISSVLTEKYR--	141
HBB_CHICK	DPENFRLLGDILIIIVLAAHFSKDFTPECQAAWQKLVRVVAHALARKYH--	146
HBE_CHICK	DPENFRLLGDILIIIVLASHFARDFTPACQFAWQKLNVVVAHALARKYH--	146
HBRH_CHICK	DPENFRLLGNILIIIVLAAHFTKDFTPTCQAVWQKLVSVAHALAYKYH--	146
MYG_CHICK	PVKYLEFISEV I I K VIAEKHAADFGADSQAAMKKALELFRNDMASKYKEF	149
 * * . . . ** .	
HBA_CHICK	---- 141	
HBAD_CHICK	---- 141	
HBPI_CHICK	---- 141	
HBB_CHICK	---- 146	
HBE_CHICK	---- 146	
HBRH_CHICK	---- 146	
MYG_CHICK	GFQG 153	

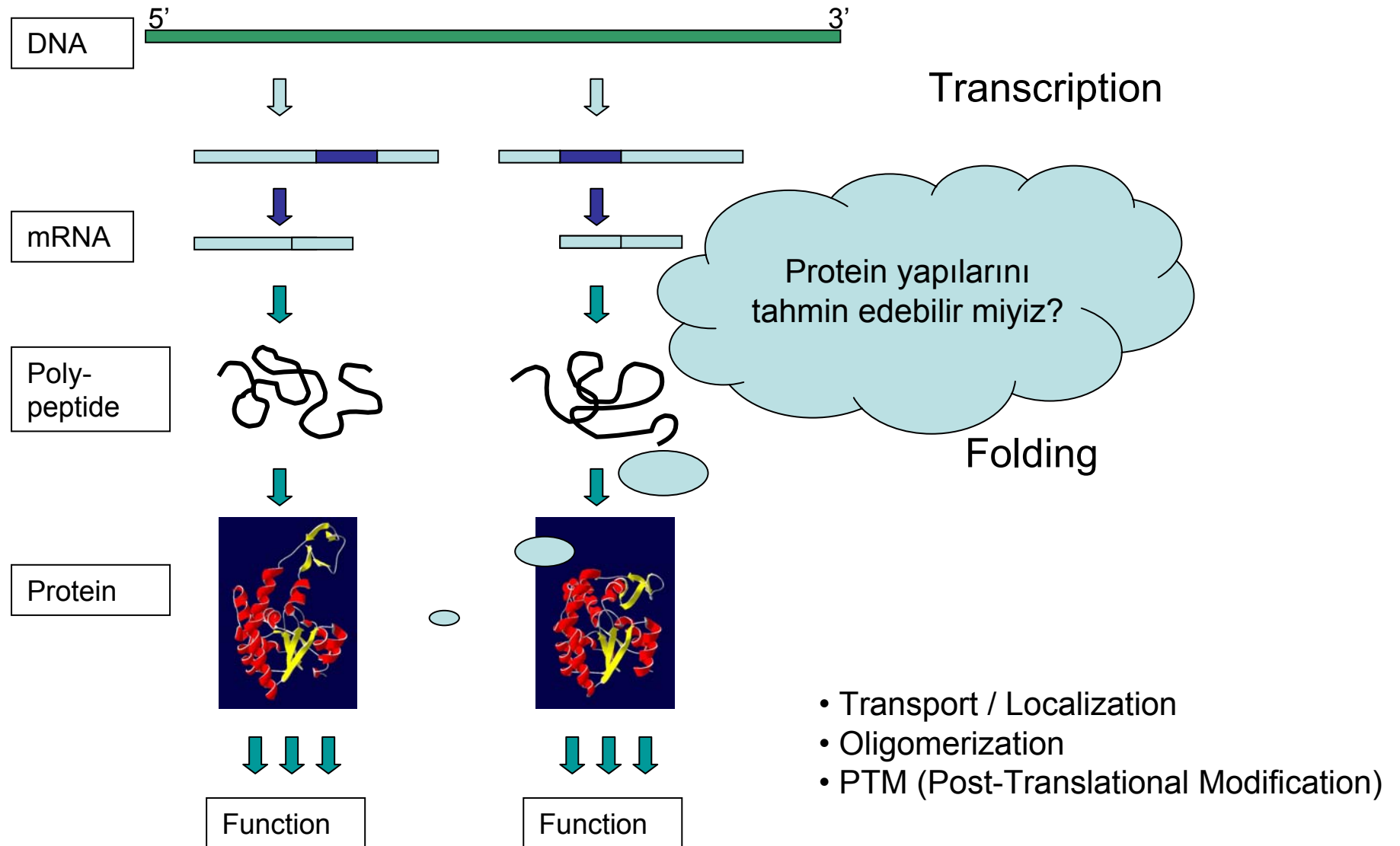
Consensus length: 154; Identity : 19 (12.3%); Similarity: 51 (33.1%)
 Character to show that a position in the alignment is perfectly conserved: '*'
 Character to show that a position is well conserved: '.'

BLAST:

Basic Local Alignment Search Tool



Biyoinformatiğe neden ihtiyacımız var?



Protein Yapısı Modelleme

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

MNIFEMLRID	EGLRLKIYKD	TEGYTIGIG
HLLTKSPSLN	AAKSELDKAI	GRNCNGVITK
DEAEKLFNQD	VDAAVRGILR	NAKLKPVYDS
LDAVRRCALI	NMVFQMGETG	VAGFTNSLRM
LQQKRWDEAA	VNLAWSRWYN	QTPNRAKRVI
TTFRTGTWDA	YKNL	

