Introductory For Bioinformatics

DATA DATA DATA DATA DATA DATA DATA DATA

Dr. Mohammad Najafi

Iran University of Medical Sciences

David Roos: “We are swimming in rapidly rising of data …..how do we help from drowning?”

”..We must hook our individual computers into the worldwide network that gives us access to daily changes in the databases and also makes immediate our communications with each other……. The programs that display and analyze the material for us must be improved - and we must learn to use them more effectively. Like the purchased kits, they will make our life easier, but also like the kits, we must understand enough of how they work to use them effectively…”

Walter Gilbert (1991)
“Towards a paradigm shift in biology” *Nature News and Views* 349:99

BIOINFORMATICS

The field of science is in which Biology ,Computer science and Informational technology merge to form a single discipline

Bioinformatics

The creation and advancement of algorithms, computational and statistical techniques for management and analysis of biological data.

Clinical Bioinformatics

The hypothesis driven of a medical subject using computers and carried out with experiment.

Biochemical Pathways

To computerize the current knowledge of molecular interactions

Metabolic pathways

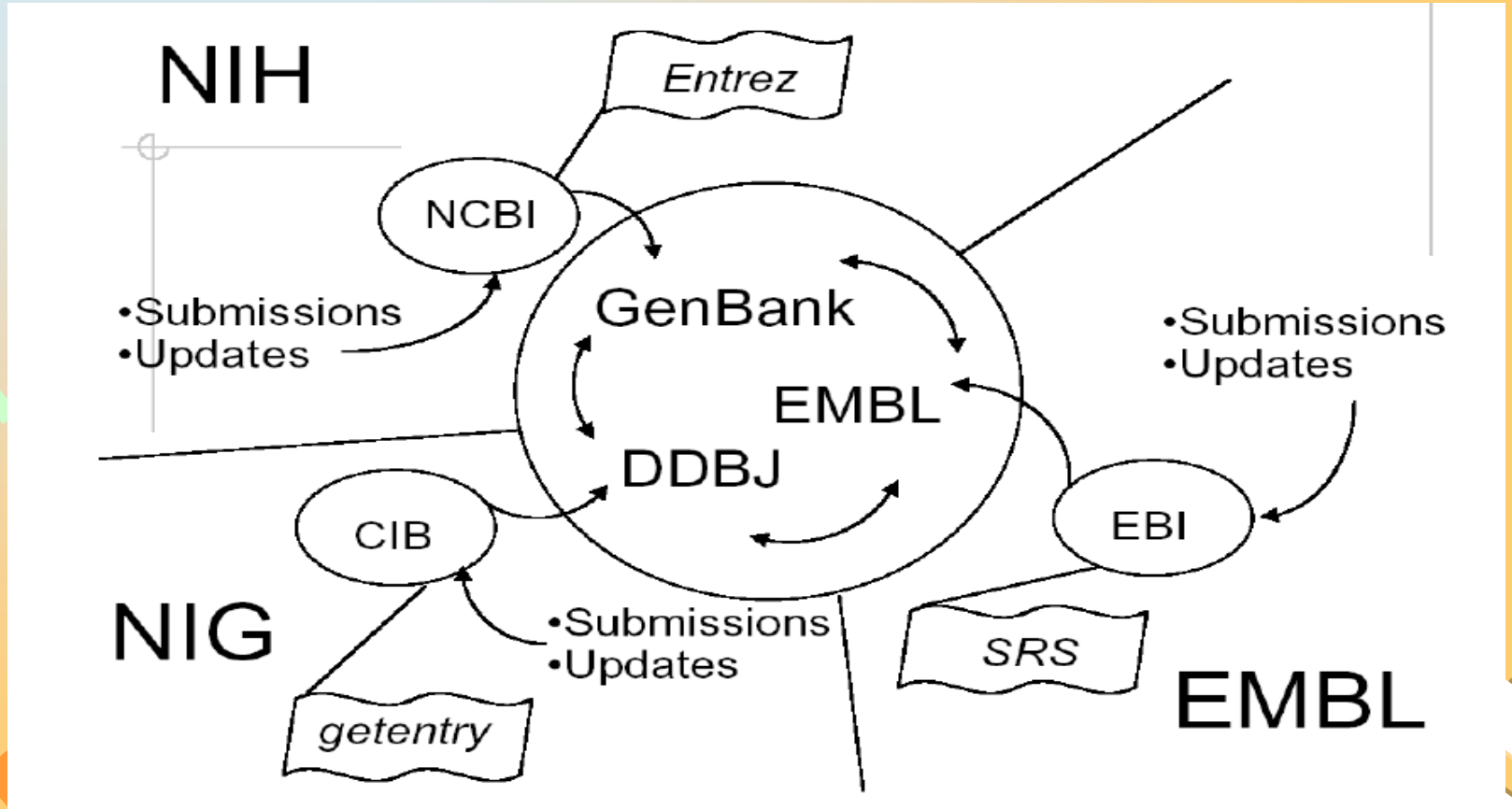
Regulatory pathways

Molecular assemblies

Current servers

- BioCarta
- KEGG
- MetaCyc

Sectors and Subsectors



Gene Databases

Protein Database

PIR (**P**rotein **I**nformation **R**esource)

PDB (**P**rotein **D**ata **B**ank)

HPRD (**H**uman **P**rotein **R**eference **D**atabase)

ALIGNMENT

An *alignment* is a mutual arrangement of two sequences, which exhibits where the two sequences are similar, and where they differ.

An *optimal alignment* is one that exhibits the most correspondences and the least differences. It is the alignment with the highest score. May or may not be biologically meaningful.

Similarity between two sequences

Global

```
A:   T C A G A C G A G
B:   T C G G A G C T G
```

Local

```
A:   T C A G A C G A G T G
B:   T C G G A G C T G
```

what is the best alignment between the two sequences?

```
A:    T C A G A C G A G T G
B:    T C G G A G C T G

I.    T C A G A C G A G T G
      T C G G A - - G C T G

II.   T C A G A C G A G T G
      T C G G A - G C - T G

III.  T C A G A C G A G T G
      T C G G A - G - C T G
```

How should alignments be scored?

How should gaps be scored?

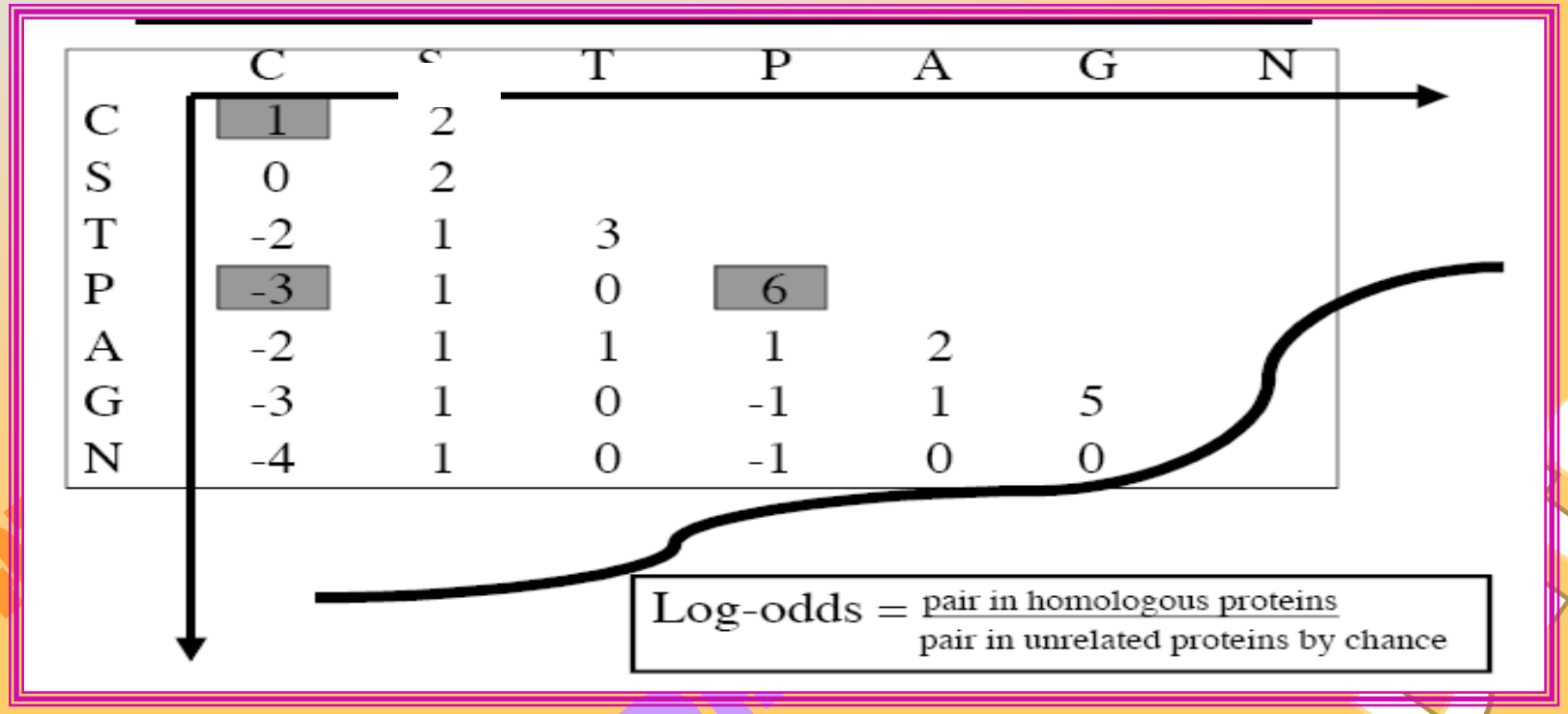
Scoring system for nucleic acids

- Match = +1 (ex; A-A, T-T, C-C, G-G)
- Mismatch = -1(ex; A-T, A-C, etc.)

T	C	A	G	A
T	C	G	G	A
+1	+1	−1	+1	+1

PAM-Point Accepted Mutation [Evolutionary Model]

- **Margaret Dayhoff, 1978**
- **related proteins**



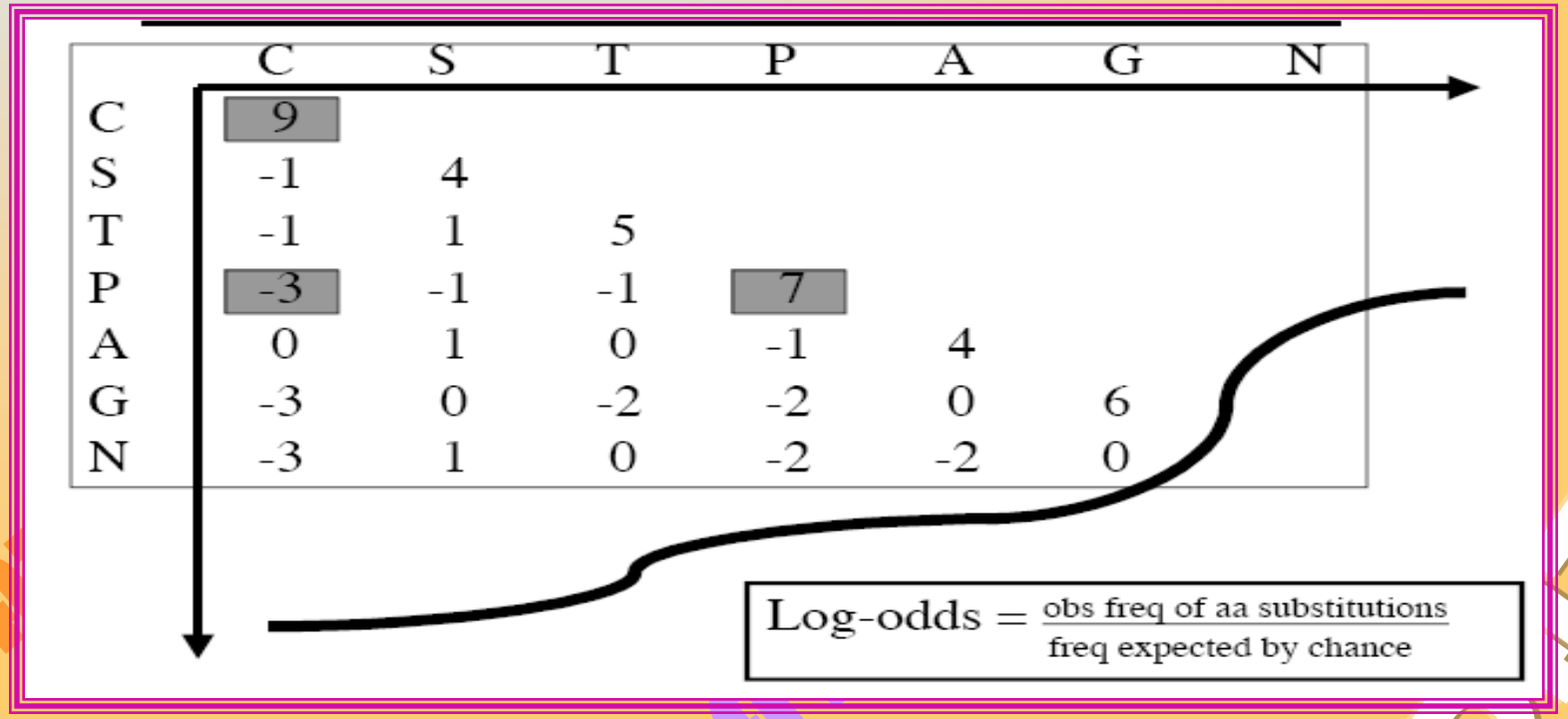
	C	S	T	P	A	G	N
C	1	2					
S	0	2					
T	-2	1	3				
P	-3	1	0	6			
A	-2	1	1	1	2		
G	-3	1	0	-1	1	5	
N	-4	1	0	-1	0	0	

$$\text{Log-odds} = \frac{\text{pair in homologous proteins}}{\text{pair in unrelated proteins by chance}}$$

BLOSUM-Block Substitutions [Conserved Sequence]

Henikoff and Henikoff, 1992

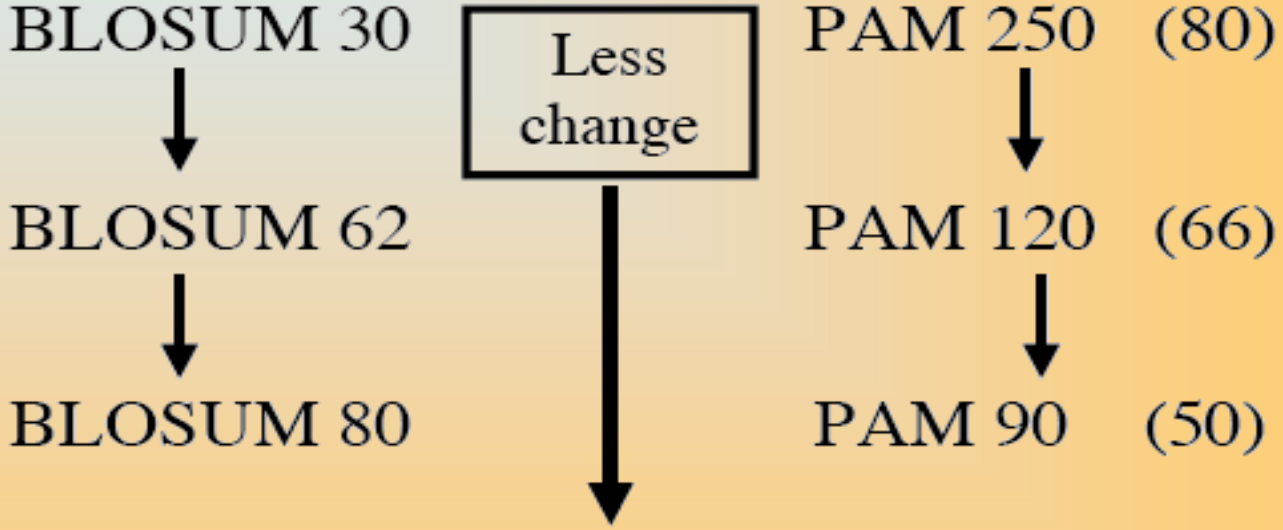
conserved, ungapped regions of a protein family



	C	S	T	P	A	G	N
C	9						
S	-1	4					
T	-1	1	5				
P	-3	-1	-1	7			
A	0	1	0	-1	4		
G	-3	0	-2	-2	0	6	
N	-3	1	0	-2	-2	0	

$$\text{Log-odds} = \frac{\text{obs freq of aa substitutions}}{\text{freq expected by chance}}$$

Substitution Matrices



Dynamic Programming

- Provides very best or optimal alignment
- Compares every pair of characters (e.g. bases or amino acids) in the two sequences
- Puts in gaps and mismatches

Dynamic Programming

Match = +2

Mismatch = -1

Gap = -3

	Gap	T	A	T	A	A	T
Gap	0	-3	-6	-9	-12	-15	-18
G	-3	[redacted]	[redacted]				
T	-6		G / T = -1	T- / - G = -6			
T	-9						
A	-12		- T / G - = -6	-1			
C	-15						
G	-18						
T	-21						
A	-24						
A	-27					[redacted]	

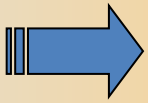
Alignment Tools

Pairwise alignment

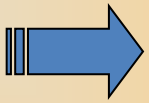
BLAST

FASTA

Dot Plat



Statistical Value



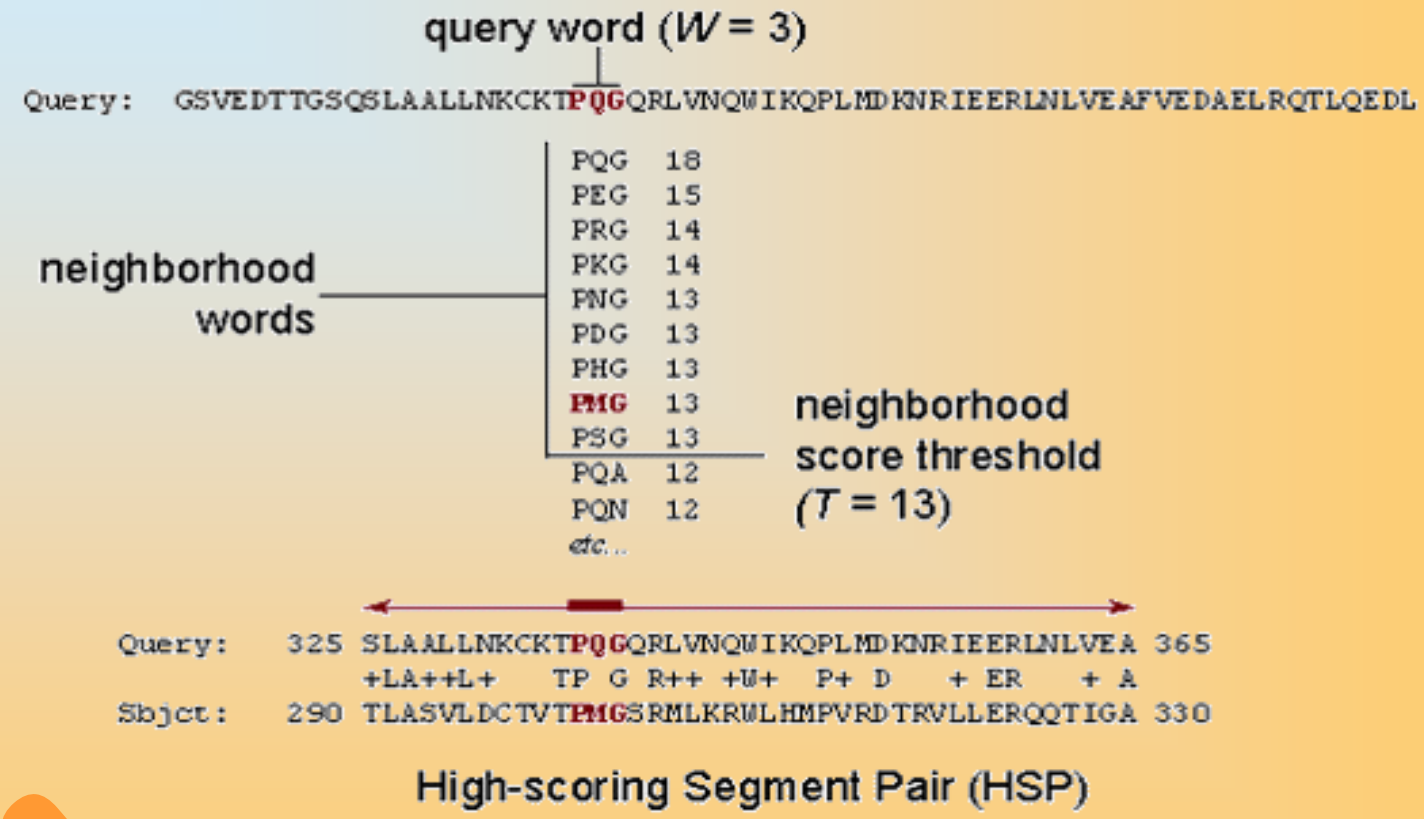
Graph

Multiple sequence alignment (MSA)

Induced pairwise method

Progressive method

The BLAST Search Algorithm



Multiple Sequence Alignment

MSA has many uses:

- Detect the overall similarity of a set of sequences.
- Find similar regions in sequences.
- As the starting point of a phylogenetic analysis to determine evolutionary relatedness.
- Find overlapping DNA fragments as part of genome sequencing efforts.

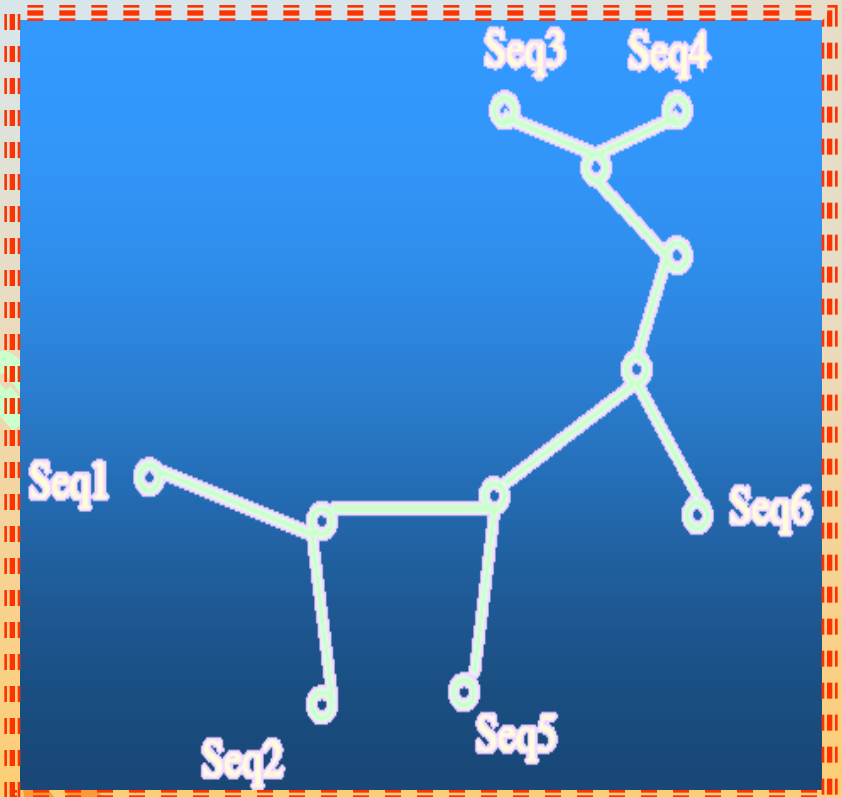
Progressive Method (MSA)

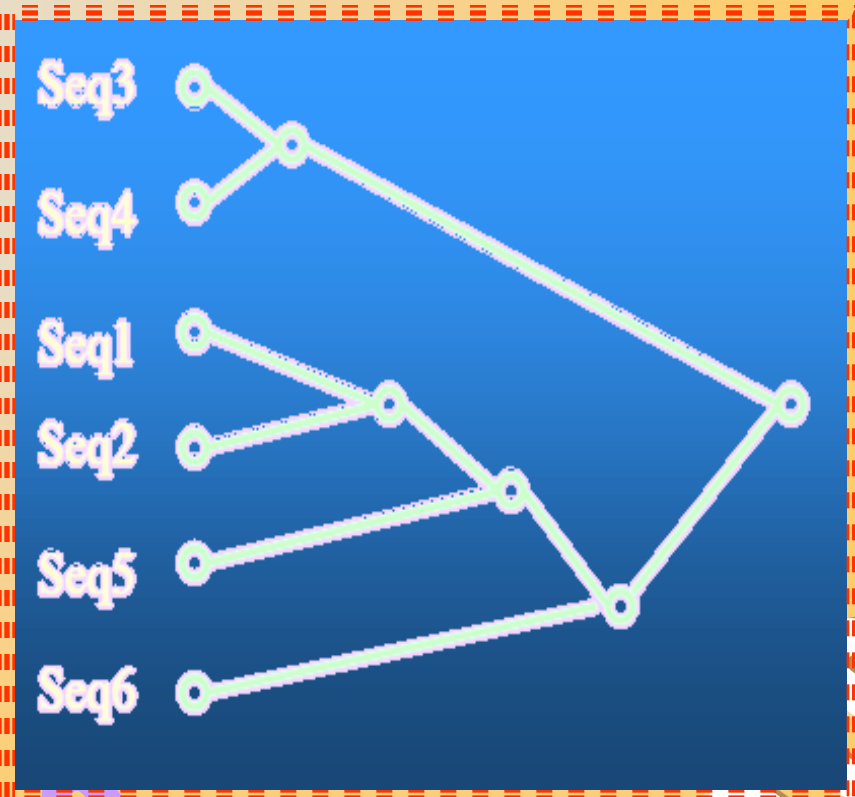
The relatedness of the sequences is done by creating a *tree*.

The tree shows how closely related the sequences to be aligned are.

The tree can be created by doing pairwise comparisons of all the sequences.

Progressive Method (Tree)





Secondary Protein Database

The Fruits of analyses of the sequences that exit within the primary sources

Secondary Protein Database

Prosite	PS
blocks	BL
Prodom	PD
prints	PR
Interpro	IPR
pfam	PF

Gene Regulation Databases

Related to *compilation* of

DNA Elements

&

Protein Regulatory Factors

Gene Regulation Databases

Promoter region

The element of bind with polymerase II

- EPD

The sequence beyond TATA box

- TRRD

-TRANCFAC

Genomics

Gene Prediction

- ab initio
- Homology
- Consensus

Proteomics

Mapping ***protein expression*** to estimate

what's actually happing in tissue following

intervention

Proteome analysis

1. A procedure for ***purification***

   2-DE

2. A method to determine ***structural information***

   Sequencing
   Ms

3. Use of ***databases***

   genebank

4. Prediction of ***functional information***

   using annotations

References

Najafi M,Jangravi **Z.** Human PON promoters**:** from similarity to prediction of polymorphic positions within transcription factor elements**.** Mini Rev Med Chem**.** 2010 Sep**;**10(10):938-45.

Najafi M, Roustazadeh A. Scoring of Matrix Gla protein (MGP) promoter variants within predicted transcription elements. IWBBIO 2013, 393-395.

Najafi M, Alipoor B, Shabani M, Amirfarhangi A, Ghasemi H.Association between rs4673 (C/T) and rs13306294 (A/G) haplotypes of NAD(P)H oxidase p22phox gene and severity of stenosis in coronary arteries. Gene. 2012 10;499(1):213-7.

Najafi M, Roustazadeh A**,** Alipoor B**.** Ox-LDL Particles**:** Modified components**,** Cellular uptake**,** Biological roles and Clinical assessments**.** Cardiovasc Hematol Disord Drug Targets**.** 2011 Oct 26.

Najafi M, Bioinformatics Tools as Power Hypothetical Predictors.
Biochem Physiol 2012 1:e113.

OMICS INTERNATIONAL

www.omicsonline.org

OMICS International (and its subsidiaries), is an Open Access publisher and international conference Organizer, which owns and operates peer-reviewed Clinical, Medical, Life Sciences, and Engineering & Technology journals and hosts scholarly conferences per year in the fields of clinical, medical, pharmaceutical, life sciences, business, engineering, and technology. Our journals have more than 3 million readers and our conferences bring together internationally renowned speakers and scientists to create exciting and memorable events, filled with lively interactive sessions and world-class exhibitions and poster presentations. Join us!

OMICS International is always open to constructive feedback. We pride ourselves on our commitment to serving the Open Access community and are always hard at work to become better at what we do. We invite your concerns, questions, even complaints. Contact us at contact.omics@omicsonline.org. We will get back to you in 24-48 hours. You may also call 1-800-216-6499 (USA Toll Free) or at +1-650-268-9744 and we will return your call in the same timeframe.

Biochemistry & Physiology: Open Access Related Journals

- Biochemistry & Pharmacology: Open Access
- Biochemistry & Analytical Biochemistry
- Journal of Plant Biochemistry & Physiology

Biochemistry & Physiology: Open Access

Related Conferences

http://www.omicsonline.org/international-scientific-conferences/

OMICS publishing Group Open Access Membership enables academic and research institutions, funders and corporations to actively encourage open access in scholarly communication and the dissemination of research published by their authors.
For more details and benefits, click on the link below:
http://omicsonline.org/membership.php