**Course: Advanced Bioinformatics**

**Module title: MSA Introduction**

**Module no. : 37**

**Multiple Sequence Alignment (MSA)**

**Definition:** A MSA is an alignment of n > 2 sequences obtained by inserting gaps ("-") into sequences such that the resulting sequences have all length L and can be arranged in a matrix of N rows and L columns where each column represents a homologous position

MSA Procedure

Given:

* Set of sequences
* Similarity score matrix
* Gap penalties

Find:

Alignment of sequences such that optimal score is achieved.

Result:

A collection of three or more protein or nucleic acid sequences that are partially or completely aligned, such that homologous residues are aligned in columns across the length of the sequences.

**Why MSA? Objectives:**

* Useful way to summarize the sequences of related proteins.
* Find important functional amino acids by assessing conservation over many sequences.
* Establish evolutionary relationships between sequences.
* What was sequence of events leading to current species?
* More precisely understand how to model 3D structures.
* Help prediction of the secondary and tertiary structures of new sequences;
* What other amino acids are acceptable in this structure?
* Preliminary step in molecular evolution analysis using Phylogenetic methods for constructing phylogenetic trees.

MSA Gold Standard: If sequences can be aligned, the alignment should reflect structural similarities. Thus, the alignment should lead to "match" of common structural and functional elements.

Multiple Sequence Alignment Example?

VTVLTALGAILKKKG

GATVLKKLGELLKAKG

KKVADALTNAVAHVD

KKVGDALTLAVGHLD

KKVMGGIALAVSKID

What are common characters across all the five sequences?