**Multiple Sequence Alignment (MSA)**

Multiple sequence alignment (MSA) is an extension of the similarity concepts to determine levels of homology (relatedness) between members of globally related sequences. In multiple sequence alignment homologous residues among a set of sequences are aligned together in columns. Homologus is meant in both the structural and evolutionary sense. Homologus are genes or proteins with the same function in different species. Ideally, a column aligned residues occupies similar 3-D structural positions and all diverge from a common ansectral residue.

Multiple sequence alignment is very important for finding similar domains in a set of sequences and for doing phylogenetic analysis.

Why do we need to conduct MSA? Aren’t the pair wise alignment sufficient? In the case of several genes, search can reveal a whole number of homologous sequences. However, pair-wise comparisons do not readily show positions that are conserved among a whole set of sequence conservations and tend to miss subtle similarities that become visible when observed simultaneously among many sequences.

Like pair wise alignments, multiple sequence alignments can also be done locally or globally. These are illustrated in figures 5.1(a) and 5.1 (b). Global alignments need to use gaps (representing insertions/deletions) while local alignments can avoid them, aligning regions between gaps.



Multiple sequence alignment can be looked at as an extension of the pair-wise alignment. The first step in multiple sequence alignment is pair-wise alignment of all the sequences. For example, you may have sequences-S1, S2, S3 and S4.  The alignment of these 4 sequences involves 6 pair-wise comparisons (S1 and S2, S1 and S3, S1 and S4, S2 and S3, S4 and S3 and S4). This is shown in figure 5.2. The result of this alignment can be represented as a tree or a dendogram.





Software- The most commonly used multiple alignment software is the Clustal package. This software is freely available by ftp from ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX, and is linked from the text World Wide Web (www) site.

Methods of Multiple sequence alignment

There are several approaches for conducting MSA. There are dynamic programming (DP) methods also for multiple sequence alignment, but they usually resource-intensive in nature.

1. Sum-of-Pairs (SP) Method.

Sum of pairs methodsis a dynamic programming methods. In this methods, instead of aligning two sequences at a time with dynamic programming, need to align two or more simultaneously.

In figure 5.5 the algorithm can be generalized to k sequences. An optimal alignment between k=3 sequences is given in figure 5.5 in three dimensions. The problem with sum-of-pairs is of large size and low speed.

The extension of this algorithm to large number of sequences can be done using the program MSA. MSA program is able to simply the many computational steps by making a projections of the multi dimensional alignment on to the alignment of each pair of sequences in the set.

MSA utilizes a variant of a multidimensional DP to produce an optimal global alignment between several sequences.



1. Progressive alignment- Most current programs use the method of **Progressive alignment,** which has the advantage of being fast. This involves making a preliminary assessment of how the sequences are related using pair-wise alignments, using this to form a **guide tree** and then guide tree to add sequences progressively to the alignment, beginning with the most closely related sequences and finishing with the most distant.

Progressive alignment is usually very effective, but it suffers from the problem that alignment error made early in the process can never be rectified.

