**Course: Advanced Bioinformatics**

**Module title: Bioinformatics Tasks**

**Module no. : 211**

**Sequence Alignments**

You can select from a list of analysis methods to compare nucleotide or amino acid sequences using pairwise or multiple sequence alignment functions.

**Pairwise sequence alignment —** Efficient implementations of standard algorithms such as the Needleman-Wunsch (nwalign) and Smith-Waterman (swalign) algorithms for pairwise sequence alignment. The toolbox also includes standard scoring matrices such as the PAM and BLOSUM families of matrices (blosum, dayhoff, gonnet, nuc44, pam). Visualize sequence similarities with seqdotplot and sequence alignment results with showalignment.

**Multiple sequence alignment —** Functions for multiple sequence alignment (multialign, profalign) and functions that support multiple sequences (multialignread, fastaread, showalignment). There is also a graphical interface (seqalignviewer) for viewing the results of a multiple sequence alignment and manually making adjustment.

**Multiple sequence profiles —** Implementations for multiple alignment and profile hidden Markov model algorithms (gethmmprof, gethmmalignment, gethmmtree, pfamhmmread, hmmprofalign, hmmprofestimate, hmmprofgenerate, hmmprofmerge, hmmprofstruct, showhmmprof).

**Biological codes —** Look up the letters or numeric equivalents for commonly used biological codes (aminolookup, baselookup, geneticcode, revgeneticcode).

**Sequence Utilities and Statistics:** You can manipulate and analyze your sequences to gain a deeper understanding of the physical, chemical, and biological characteristics of your data. Use a graphical user interface (GUI) with many of the sequence functions in the toolbox (seqviewer).

**Sequence conversion and manipulation —** The toolbox provides routines for common operations, such as converting DNA or RNA sequences to amino acid sequences, that are basic to working with nucleic acid and protein sequences (aa2int, aa2nt, dna2rna, rna2dna, int2aa, int2nt, nt2aa, nt2int, seqcomplement, seqrcomplement, seqreverse).

You can manipulate your sequence by performing an in silico digestion with restriction endonucleases (restrict) and proteases (cleave).

**Sequence statistics —** Determine various statistics about a sequence (aacount, basecount, codoncount, dimercount, nmercount, ntdensity, codonbias, cpgisland, oligoprop), search for specific patterns within a sequence (seqshowwords, seqwordcount), or search for open reading frames (seqshoworfs). In addition, you can create random sequences for test cases (randseq).

**Sequence utilities —** Determine a consensus sequence from a set of multiply aligned amino acid, nucleotide sequences (seqconsensus, or a sequence profile (seqprofile). Format a sequence for display (seqdisp) or graphically show a sequence alignment with frequency data (seqlogo).

Additional MATLAB functions efficiently handle string operations with regular expressions (regexp, seq2regexp) to look for specific patterns in a sequence and search through a library for string matches (seqmatch).

Look for possible cleavage sites in a DNA/RNA sequence by searching for palindromes (palindromes).

**Protein Property Analysis**: You can use a collection of protein analysis methods to extract information from your data. You can determine protein characteristics and simulate enzyme cleavage reactions. The toolbox provides functions to calculate various properties of a protein sequence, such as the atomic composition (atomiccomp), molecular weight (molweight), and isoelectric point (isoelectric). You can cleave a protein with an enzyme (cleave, rebasecuts) and create distance and Ramachandran plots for PDB data (pdbdistplot, ramachandran). The toolbox contains a graphical user interface for protein analysis (proteinplot) and plotting 3-D protein and other molecular structures with information from molecule model files, such as PDB files (molviewer).

**Amino acid sequence utilities —** Calculate amino acid statistics for a sequence (aacount) and get information about character codes (aminolookup).