Applied Biology and Bioengineering (BT520)

EXPERIMENT 3

BIOINFORMATICS TOOLS FOR SEQUENCE ANALYSIS, DESIGN OF PRIMERS AND IDENTIFICATION OF RESTRICTION SITES

Objectives:

1. Retrieving DNA/protein sequences from databases.
2. Computing amino-acid composition, molecular weight, isoelectric point, and other parameters like hydrophobicity or hydrophilicity of a protein.
3. Predicting elements of secondary structure.
4. Predicting the domain organization of proteins.
5. Design of primers and identification of restriction sites within the given sequence/gene.

Introduction

The term homology implies a common evolutionary relationship between two sequences related by divergence from a common ancestor. BLAST (Basic Local Alignment Search Tool) uses an algorithm developed by NCBI (http://www.ncbi.nlm.nih.gov/LocusLink/) that seeks out local alignment (the alignment of some portion of two sequences) as opposed to global alignment (the alignment of two sequences over their entire length: multiple sequence alignment (MSA). The results of BLAST are converted to FastA format and MSA is done using T-COFFEE or clustalx to find out the conserved regions among the hits. The physiochemical properties like amino acid composition, pI, Mol.Wt., etc. can be computed from tools on ExPASy server (ProtParam, ScanSite, pI/Mw, ProtScale and SAPS). Secondary structure of protein (showing helices, sheets and coils) is predicted using PSIPRED VIEW, PROTEUS, Jpred, PredictProtein etc., servers. The domain organization (pattern and profile searches) can be computed using ScanProsite, PPSEARCH and InterPro Scan available from ExPASy network. Post-translational modification prediction can also be done on tools available on ExPASy network. Primers (forward and reverse) are designed for amplification of desired fragment or a particular gene from a genome. It can be done using tools like Primer 3. Restriction sites, or restriction recognition sites, are specific sequences of nucleotides that are recognized by restriction enzymes. The sites are generally palindromic, (because restriction enzymes usually bind as homodimers) and a particular restriction enzyme may cut the sequence between two nucleotides within its recognition site, or somewhere nearby. They can be pedicted using NEB cutter V2.0 software.

Requirements:
Comp. Lab , PC, internet connection, etc.

Procedure:

Retrieve a DNA sequence from NCBI (PubMed) and convert it to amino acid sequence using Transeq from ExPASy network or get a protein sequence from SwissProt/TrEMBL sites and convert the sequence in FastA format. Do homology search using PSI-BLAST and identify the good hits. Select the sequences (in FastA format) which have E (expectation value) ≤0.005 or percentage similarity ≥ 25%. Perform MSA using T-COFFEE or clustalx and edit using CINEMA or other relevant tools available on ExPASy network and build a evolutionary phylogenetic tree. The secondary structure prediction can be done from PSIPRED VIEW and PROTEUS server available freely on web. Pattern and profile of a protein from sequence can be done using tolls like ScanProsite, PPSEARCH and InterPro Scan available on ExPASy network. Primers are designed using Primer3 (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) or PRIDE (http://www.dkfz-heidelberg.de/tbi_old/services/Pride/search_primer). Restriction recognition sites can be identified from a sequence by making use of NEBcutter V2.0 software.

Observation:

Results and discussion:

List of instruments:

PC with internet connectivity.

Reference:

1. Bioinformatics for dummies, 2nd Ed. By by Jean-Michel Claverie,PhD and Cedric Notredame,PhD. (Wiley publication Inc.).

2. Bioinformatics: Sequence and Genome Analysis, 2nd Ed. by David W. Mount. Cold Spring Harbor Laboratory Press.