

Nucleic acid sequence analysis
Reading frames, Transcriptional & Translational signals
prediction

- In molecular genetics, an open reading frame (ORF) is the part of a reading frame that contains no stop codons.
- The transcription termination pause site is located after the ORF, beyond the translation stop codon, because if transcription were to cease before the stop codon, an incomplete protein would be made during translation.
- Normally, inserts which interrupt the reading frame of a subsequent region after the start codon cause frame shift mutation of the sequence and dislocate the sequences for stop codons.

- One common use of open reading frames is as one piece of evidence to assist in gene prediction.
- Long ORFs are often used, along with other evidence, to initially identify candidate protein coding regions in a DNA sequence.
- The presence of an ORF does not necessarily mean that the region is ever translated.
- For example in a randomly generated DNA sequence with an equal percentage of each nucleotide, a stop-codon would be expected once every 21 codons.

- A simple gene prediction algorithm for prokaryotes might look for a start codon followed by an open reading frame that is long enough to encode a typical protein, where the codon usage of that region matches the frequency characteristic for the given organism's coding regions.
- By itself even a long open reading frame is not conclusive evidence for the presence of a gene.

Open Reading Finding Tool:

- **ORF Finder:** The ORF Finder (Open Reading Frame Finder) is a graphical analysis tool which finds all open reading frames of a selectable minimum size in a user's sequence or in a sequence already in the database.
- This tool identifies all open reading frames using the standard or alternative genetic codes.
- The deduced amino acid sequence can be saved in various formats and searched against the sequence database using the WWW BLAST server.
- The ORF Finder should be helpful in preparing complete and accurate sequence submissions.
- It is also packaged with the Sequin sequence submission software.

ORF Investigator:

- ORF Investigator is a program which not only gives information about the coding and non-coding sequences but also can perform pairwise global alignment of different gene/DNA regions sequences.
- The tool efficiently finds the ORFs for corresponding amino acid sequences and converts them into their single letter amino acid code, and provides their locations in the sequence.
- The pairwise global alignment between the sequences makes it convenient to detect the different mutations, including single nucleotide polymorphism.
- Needleman and Wunsch algorithms are used for the gene alignment.
- The ORF Investigator is written in the portable Perl programming language, and is therefore available to users of all common operating systems.

ORFPredictor:

- OrfPredictor is a web server designed for identifying protein-coding regions in expressed sequence tag (EST)-derived sequences.
- For query sequences with a hit in BLASTX, the program predicts the coding regions based on the translation reading frames identified in BLASTX alignments, otherwise, it predicts the most probable coding region based on the intrinsic signals of the query sequences. The output is the predicted peptide sequences in the FASTA format, and a definition line that includes the query ID, the translation reading frame and the nucleotide positions where the coding region begins and ends.
- OrfPredictor facilitates the annotation of EST-derived sequences, particularly, for large-scale EST projects.