Translation and Open Reading Frame Search

Regions of DNA that encode proteins are first transcribed into messenger RNA and then translated into protein. By examining the DNA sequence alone we can determine the sequence of amino acids that will appear in the final protein. In translation codons of three nucleotides determine which amino acid will be added next in the growing protein chain. It is important then to decide which nucleotide to start translation, and when to stop, this is called an open reading frame.

Once a gene has been sequenced it is important to determine the correct open reading frame (ORF). Every region of DNA has six possible reading frames, three in each direction. The reading frame that is used determines which amino acids will be encoded by a gene. Typically only one reading frame is used in translating a gene (in eukaryotes), and this is often the longest open reading frame. Once the open reading frame is known the DNA sequence can be translated into its corresponding amino acid sequence. An open reading frame starts with an atg (Met) in most species and ends with a stop codon (taa, tag or tga).

For example, the following sequence of DNA can be read in six reading frames. There are three ORFs in the forward direction and three ORFs in the reverse direction. The three reading frames in the forward direction are shown with the translated amino acids below each DNA seqeunce. Frame 1 starts with the "a", Frame 2 with the "t" and Frame 3 with the "g". Stop codons are indicated by an "*" in the protein sequence. The longest ORF is in Frame 1.

```
   5'                                                  3'
   atgcccaagctgaatagcgtagaggggttttcatcatttgaggacgatgtataa

 1 atg ccc aag ctg aat agc gta gag ggg ttt tca tca ttt gag gac gat gta taa
    M   P   K   L   N   S   V   E   G   F   S   S   F   E   D   D   V   *
 2  tgc cca agc tga ata gcg tag agg ggt ttt cat cat ttg agg acg atg tat
     C   P   S   *   I   A   *   R   G   F   H   H   L   R   T   M   Y
 3   gcc caa gct gaa tag cgt aga ggg gtt ttc atc att tga gga cga tgt ata
      A   Q   A   E   *   R   R   G   V   F   I   I   *   G   R   C   I
```