American Journal of Biology and Natural Sciences
https://biojournals.us/index.php/AJBNS
**ISSN: 2997-7185**

# Search For Dna Sequences in Orf Finder

**Nasriddinova P. M.**
Karshi state university, Kashkadarya, Uzbekistan
nasriddinova-parizod@qarshidu.uz

**Received:** 2025, 15, Jul
**Accepted:** 2025, 21, Aug
**Published:** 2025, 30, Sep

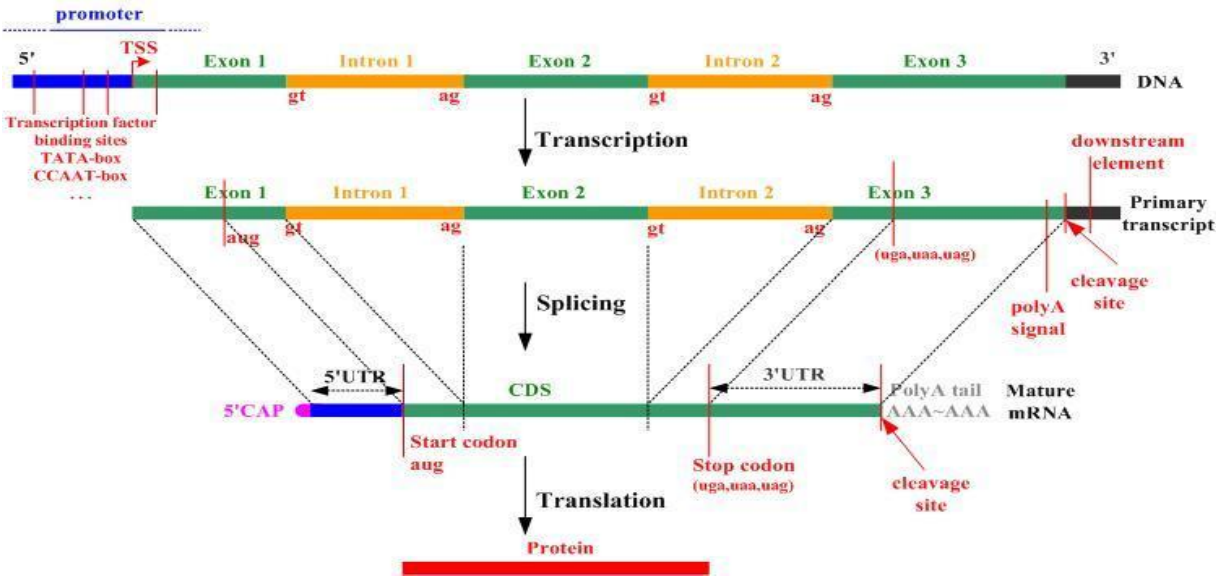
**Annotation:** In addition to DNA carrying the genetic information of a living organism or cell, it also plays a crucial role in the formation of amino acids. Before the DNA base sequence is encoded into amino acids, a search for open reading frames (ORFs) occurs. These ORFs form DNA regions potentially capable of being converted into amino acids. Demonstrating the full potential of ORFs in DNA sequences is important because it can aid in genetic analysis research.

**Keywords:** Open reading frame, start codon, stop codon, nucleotide sequence, protein synthesis, genetic code, transcription, software analysis, amino acid sequence, genome, BLAST.

**Introduction.** Genes are information carriers that affect the characteristics of an organism that are passed down from generation to generation. Genes consist of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). DNA is a polymer made up of four different types of monomers called nucleotides [1].

In addition to being a carrier of information about the nature of an organism or cell, DNA also plays an important role in the process of protein formation. Each protein molecule consists of a series of amino acids in a specific sequence, and the formation of each amino acid is encoded by the sequence of nitrogen bases in DNA.

Since prokaryotes lack introns, the ORF is the coding sequence of a gene that is directly transcribed into mRNA. Eukaryotic chromosomes contain introns as well as exons. In multi-exon eukaryotic genes, introns are removed and the exons are spliced together after transcription to form the final mRNA for protein translation.



Before the sequence of nitrogenous bases in DNA is encoded, there is a search process for open reading frames (ORFs), which produce parts of the DNA sequence that have the potential to be translated into amino acids. There are several websites that offer an ORF Finder page, but they do not display results for common ORFs in the encoded DNA sequence.

An open reading frame, or ORF, is the sequence of nucleotides between a start codon and a stop codon. An open reading frame (ORF) is a reading frame that is capable of being transcribed into RNA and translated into protein. Such an ORF may contain a start codon (usually AUG in RNA) and cannot extend beyond a stop codon (usually UAA, UAG, UGA). This start codon (which does not have to be the first) indicates where synthesis can begin, while the stop codon indicates where to stop. If transcription stops before the stop codon, translation will produce an incomplete protein. An ORF is a DNA sequence that begins with the start codon “ATG” and ends with any of three stop codons (TAA, TAG, or TGA). An ORF can be the complete mRNA portion of a gene. Thus, an ORF is a portion of an organism’s genome that contains a sequence of bases that could potentially encode an individual protein. Once the ORF for a gene or its mRNA is known, the nucleotide sequence can be converted to the corresponding amino acid sequence.

**Materials and Methods.** *Study Area.* Nucleotides are the simple units that make up DNA, consisting of a sugar (deoxyribose) to which phosphate groups are attached, and nitrogenous bases in the form of adenine, guanine, cytosine, and thymine. Thus, nucleotides are called A, G, C, and T [2].

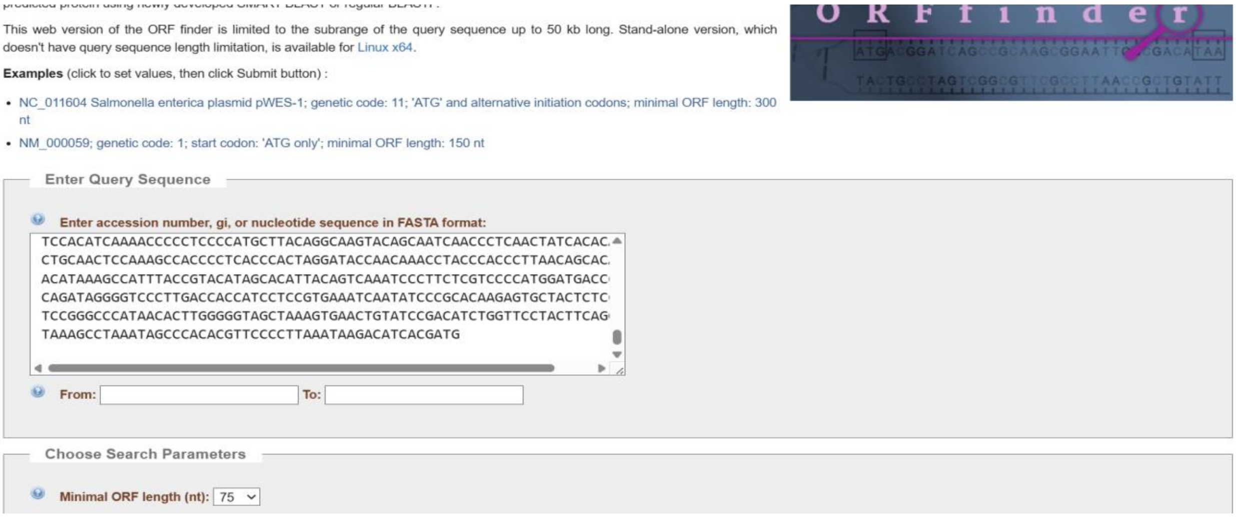
An open reading frame (ORF) is a triplet nucleotide sequence that is read as a codon that specifies an amino acid, with three possible reading frames on a single DNA strand. The length of an ORF can indicate the coding region of a protein in a DNA sequence [3]. The characteristics of an ORF include:

1. Start codon. A start codon in DNA initiates the translation of the first amino acid in a polypeptide chain. The first three bases of the mRNA coding sequence that is translated into proteins are called start codons. ATG is the most common start codon in DNA and codes for the amino acid methionine (Met) in eukaryotes and formyl methionine (fMet) in prokaryotes. During protein synthesis, ATG is converted to AUG in mRNA, which is then recognized as a start codon by tRNA with the help of several initiation factors and begins translation of the mRNA [4].
2. An ORF contains a sequence of nucleotide bases located between the start and stop codon positions.
3. Termination codon or Stop codon (TAA, TAG, TGA in DNA). A stop codon is essentially three nucleotide bases in mRNA that signal the end of the translation process. Most codons in

mRNA are associated with the addition of specific amino acids to the growing protein chain in a specific order. The formation of the amino acid chain stops at this point [5].

We applied this method (https://www.ncbi.nlm.nih.gov/nuccore/NC_001477) using the publicly available fasta NC_001477 data. The genetic sequence of the Homo sapiens apoptosis-inducing factor mitochondria-associated 1 (AIFM1) was analyzed. The DNA sequence was retrieved from the NCBI database and analyzed using the ORF finder program.

**Results and analysis.** This research phase begins with the process of entering DNA sequence data in FASTA format, followed by the adjustment of the minimum ORF length, start codon, and genetic code parameters.



This web version of the ORF finder is limited to the subrange of the query sequence up to 50 kb long. Stand-alone version, which doesn't have query sequence length limitation, is available for Linux x64.

**Examples** (click to set values, then click Submit button) :

- NC_011604 Salmonella enterica plasmid pWES-1; genetic code: 11; 'ATG' and alternative initiation codons; minimal ORF length: 300 nt
- NM_000059; genetic code: 1; start codon: 'ATG only'; minimal ORF length: 150 nt

Enter Query Sequence

Enter accession number, gi, or nucleotide sequence in FASTA format:

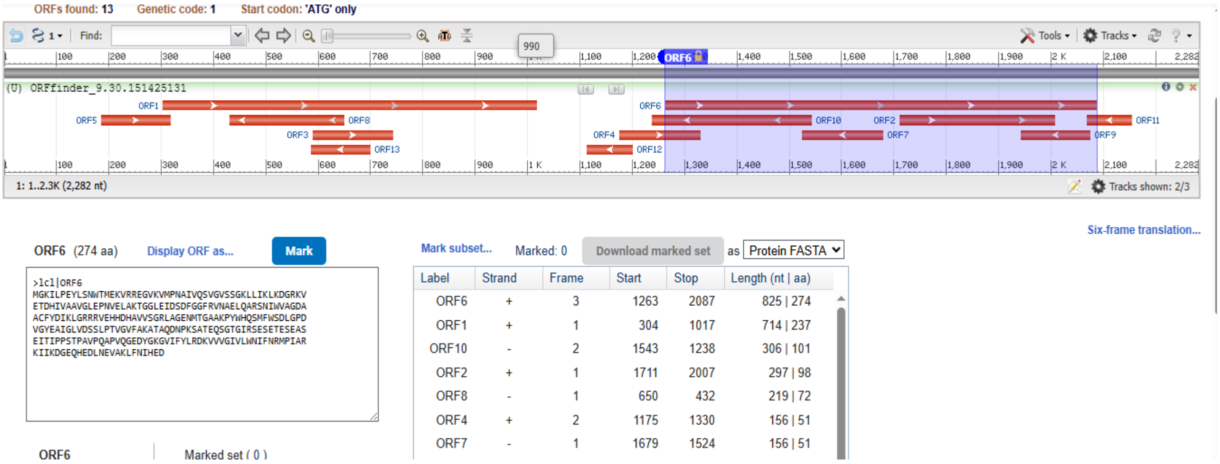
```
TCCACATCAAAACCCCCTCCCCATGCTTACAGGCAAGTACAGCAATCAACCCTCAACTATCACAC
CTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCACCCTTAACAGCAC
ACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACC
CAGATAGGGGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTC
TCCGGGCCCATAACACTTGGGGGTAGCTAAAGTGAACTGTATCCGACATCTGGTTCCTACTTCAG
TAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATG
```

From: To:

Choose Search Parameters

Minimal ORF length (nt): 75

The program presents the ORFs present in the input sequence and their amino acid sequences. The process of applying the dynamic programming method in this study is divided into several steps. The first step is to find the start codon and stop codon, the results of which are used to select each ORF, which allows it to be translated into an amino acid. In this step, each start and stop position in the nucleotide sequence is stored in a new list.



ORFs found: 13 Genetic code: 1 Start codon: 'ATG' only

1: 1..2.3K (2,282 nt)

ORF6 (274 aa)

```
>lcl|ORF6
MGKILPEYLSNWTMEKVRREGVKVMPNAIVQSVGVSSGKLLIKLKDGRKV
ETDHIVAAVGLEPNVELAKTGGLEIDSDFGGFRVNAELQARSNIWVAGDA
ACFYDIKLGRRRVEHHDHAVVSGRLAGENMTGAAKPYWHQSMFWSDLGPD
VGYEAIGLVDSSLPTVGVFAKATAQDNPKSATEQSGTGIRSESETESEAS
EITIPPSTPAVPQAPVQGEDYGKGVIFYLRDKVVVGIVLWNIFNRMPIAR
KIIKDGEQHEDLNEVAKLFNIHED
```

ORF6 Marked set ( 0 )

Marked: 0

| Label | Strand | Frame | Start | Stop | Length (nt \| aa) |
|---|---|---|---|---|---|
| ORF6 | + | 3 | 1263 | 2087 | 825 \| 274 |
| ORF1 | + | 1 | 304 | 1017 | 714 \| 237 |
| ORF10 | - | 2 | 1543 | 1238 | 306 \| 101 |
| ORF2 | + | 1 | 1711 | 2007 | 297 \| 98 |
| ORF8 | - | 1 | 650 | 432 | 219 \| 72 |
| ORF4 | + | 2 | 1175 | 1330 | 156 \| 51 |
| ORF7 | - | 1 | 1679 | 1524 | 156 \| 51 |

Thirteen open reading frames (ORFs) consisting of 1113 amino acid sequences were identified in the genetic sequence of Homo sapiens apoptosis-inducing factor 1 (AIFM1). This protein sequence was analyzed using SmartBLAST and revealed a 97.45% identity with the apoptosis-inducing factor 1, mitochondrial precursor protein of Rattus norvegicus. This suggests that this protein has important molecular functions in living organisms. Overall, the analyses demonstrated that bioinformatics methods can be used to increase the accuracy and efficiency of molecular biology. This approach may be widely used in the future in the areas of genetic disease detection, drug development, and genetic engineering.

## References:

1. Neil A. Campbell and Jane Reece, *Campbell Biology 11th Edition*. New York: Pearson Higher Education, 2003.
2. Bruce Albert and Alexander Johson, *Molecular Biology of The Cell Sixth Edition*. New York: Garland Science, 2014.
3. Nature.com. nature.com. [Online]. https://www.nature.com/subjects/open-reading-frames Accessed on November 16, 2018
4. Biology Wise Staff. (2018) Biology Wise. [Online]. https://biologywise.com/start-codon Accessed on September 29, 2018
5. BiologyWiseStaff. (2018) Biology Wise. [Online]. https://biologywise.com/stop-codon Accessed on September 29, 2018