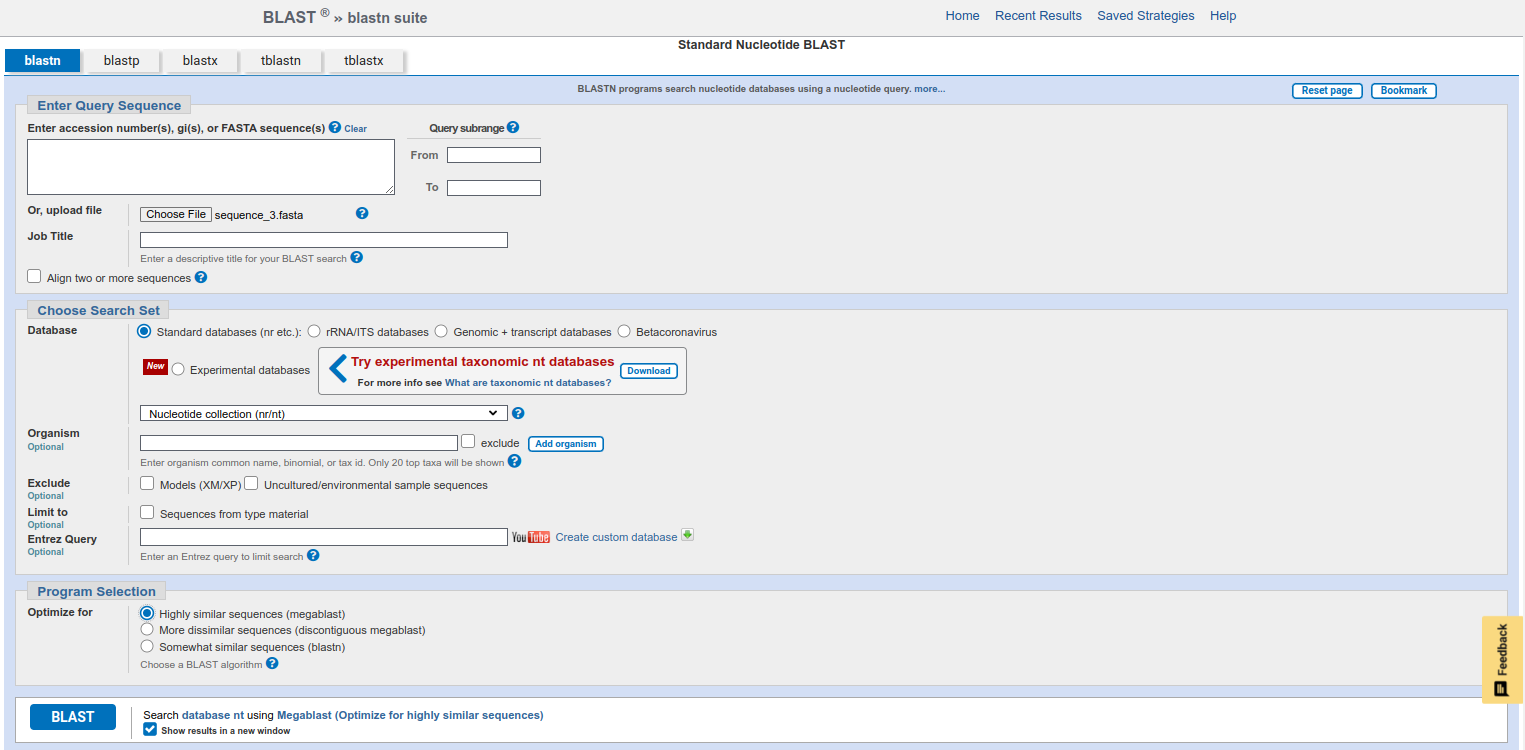
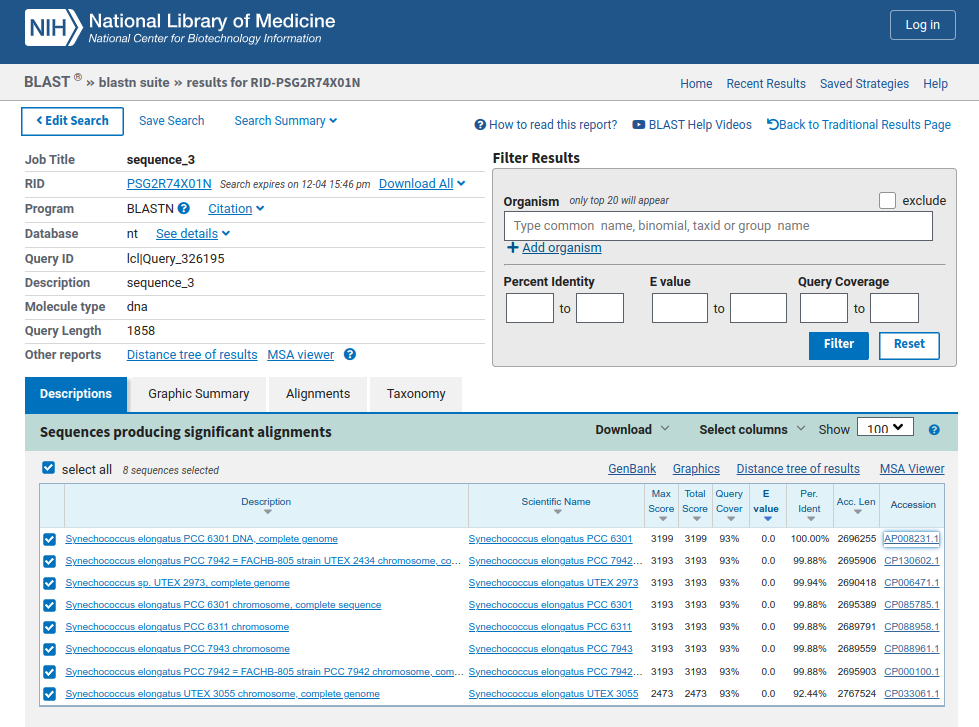
The task is to analyse a specific nucelotide sequence using blast. phylogenetic tree analysis using MEGA, multiple sequence alignment, swiss model repository and open reading frame, and also homologs. we will need to explain and explain the results obtained from the tests above, especially the phylogenetic tree relationships.

**BLAST Analysis**

BLAST url: https://blast.ncbi.nlm.nih.gov/Blast.cgi

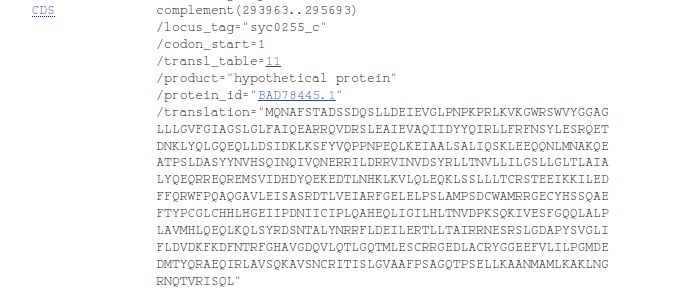
As the sequence is DNA, we will use Nucleotide BLAST (blastn).

Figure 1: NCBI Nucleotide Blast Search Page

Figure 2: NCBI Nucleotide Blast Result Page

Here we get sequence\_3 has 100% identity with AP008231.1 (Synechococcus elongatus PCC 6301).

Sequence\_3 matches position 293963-295693 in AP008231.1 sequence which is a coding sequence having protein\_id BAD78445.1. We can download that protein sequence as fasta file to use in SWISS-MODEL Repository.

Figure 3: Position of Sequence 3 in AP008231.1

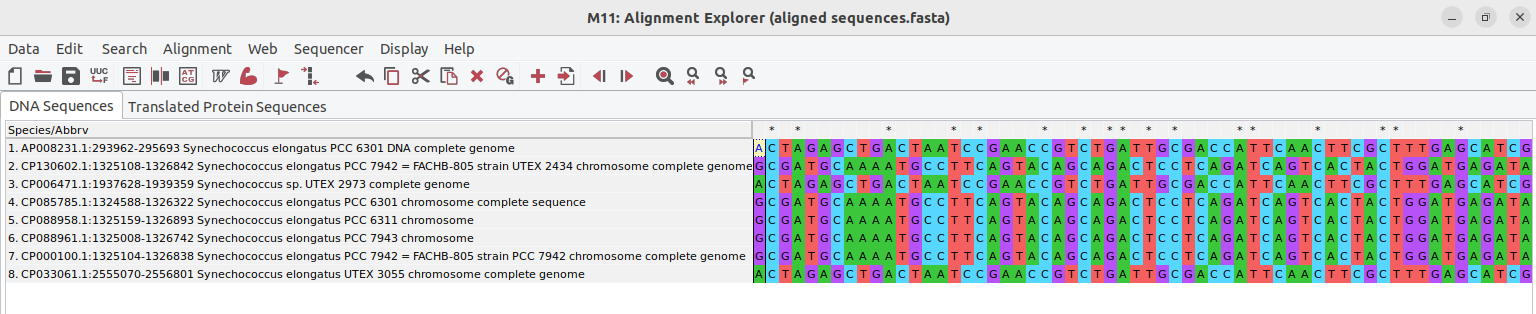
**Homolog sequences are:**

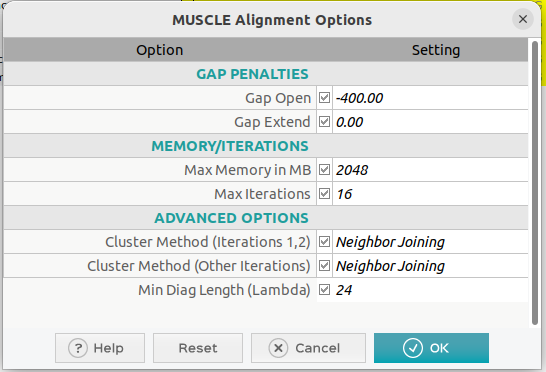
1. AP008231.1
2. CP130602.1
3. CP006471.1
4. CP085785.1
5. CP088958.1
6. CP088961.1
7. CP000100.1
8. CP033061.1

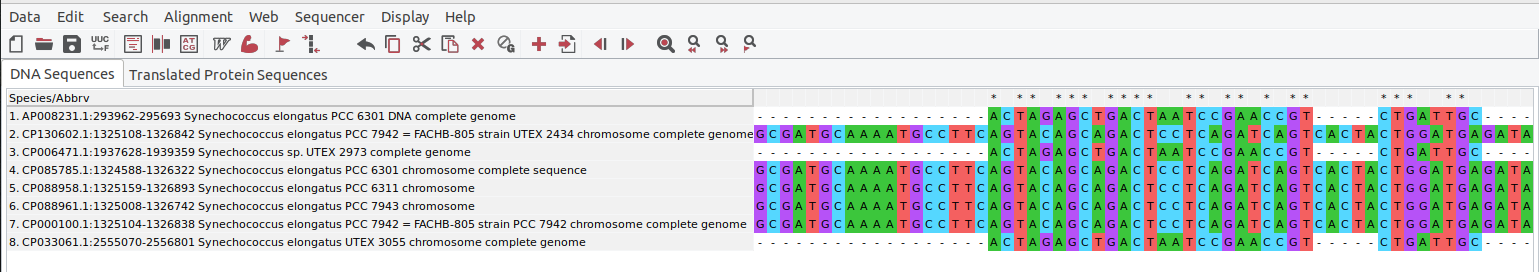
**Multiple Sequence Alignment (MEGA)**

Open MEGA and load fasta file having homolog sequences:

ALIGN → Edit/Build Alignment → Retrieve a sequence from a file → Open ( homolog sequences.fasta)

Figure 4: Loaded Homolog Fasta File in MEGA

Figure 5: Parameters of Muscle Alignment for MSA

Figure 6: First Few Bases of Aligned sequences using MUSCLE

**Phylogenetic Tree Analysis (MEGA)**

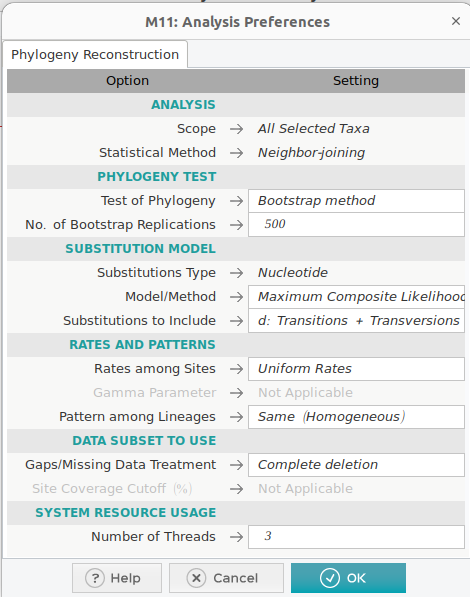
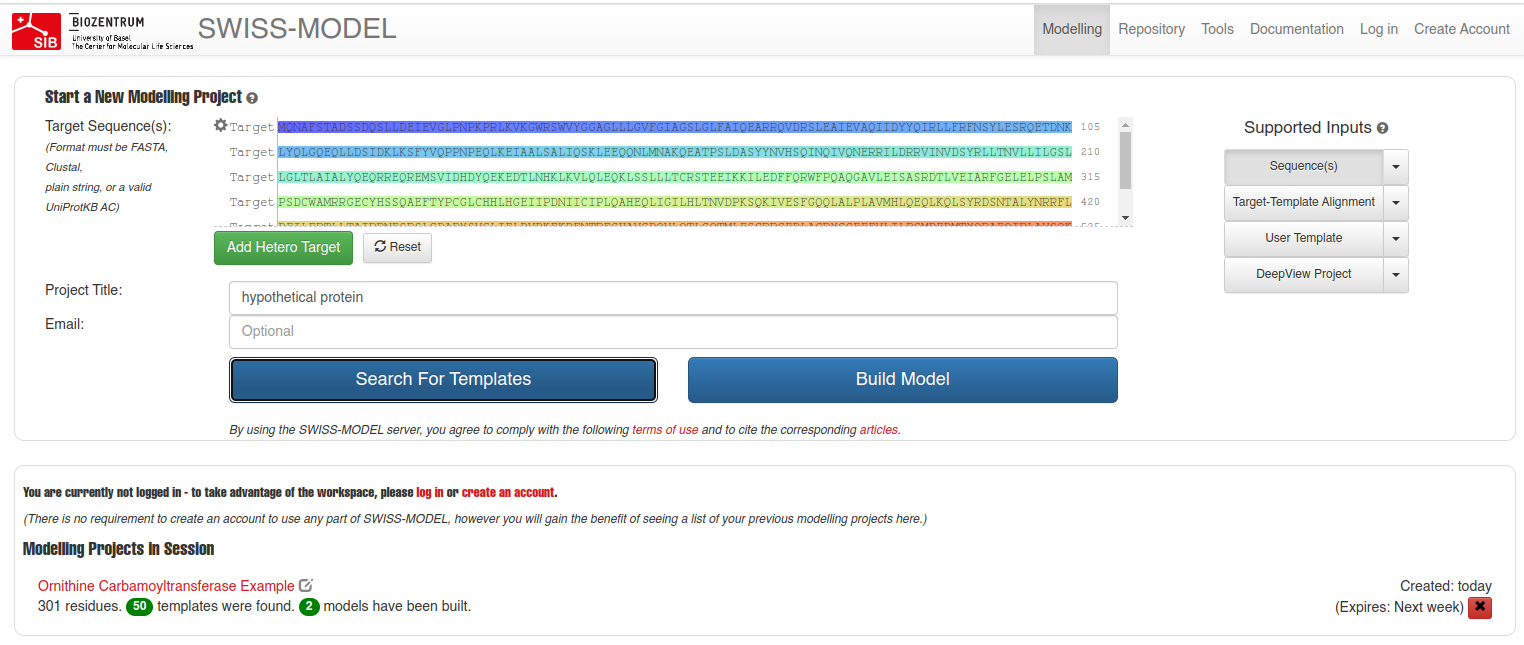
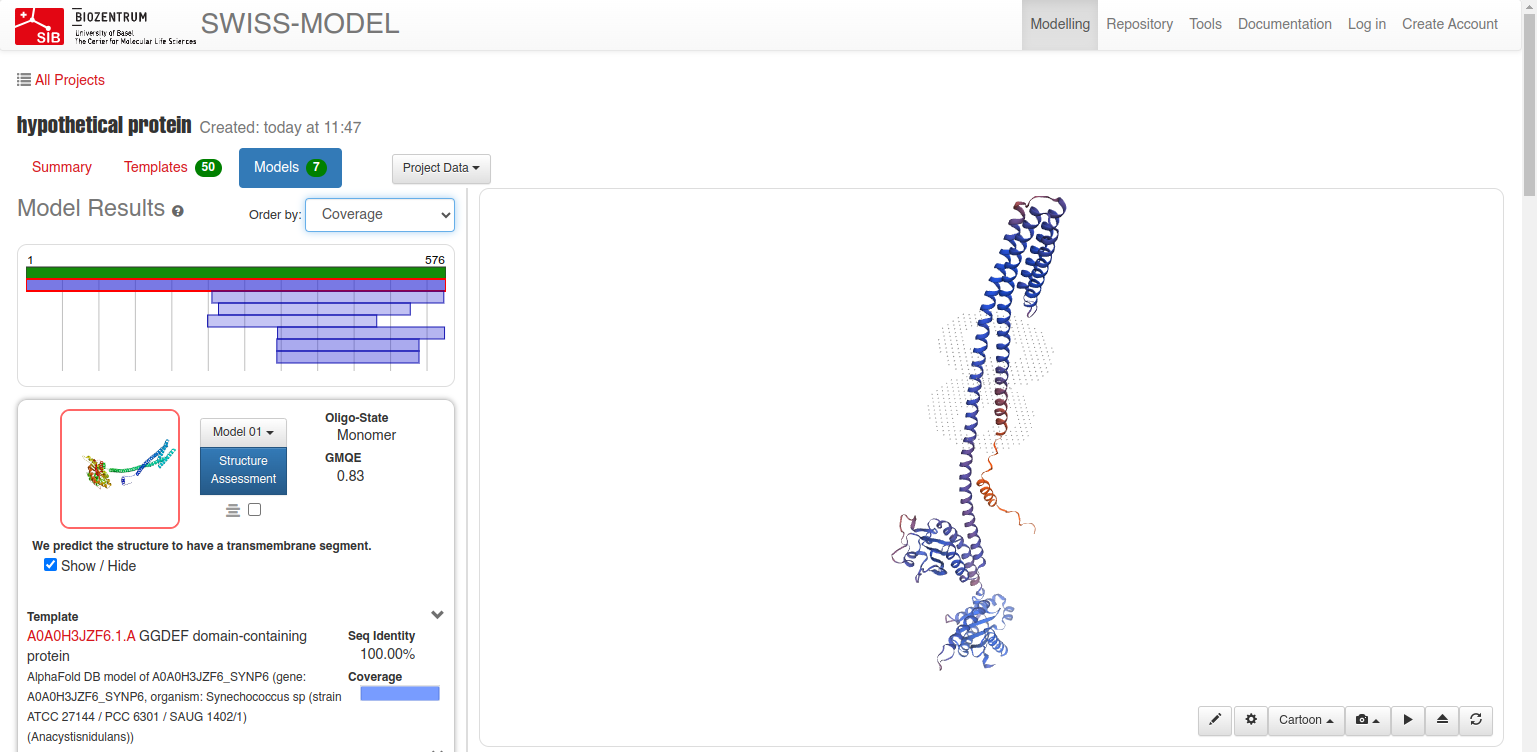
Figure 7: Parameters for Phylogenetic Analysis in MEGA

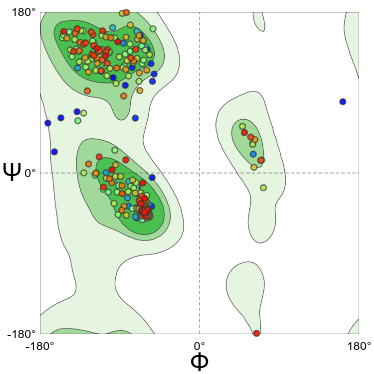
Figure 8: Drawn Phylogenetic Tree with Bootstrap Confidence

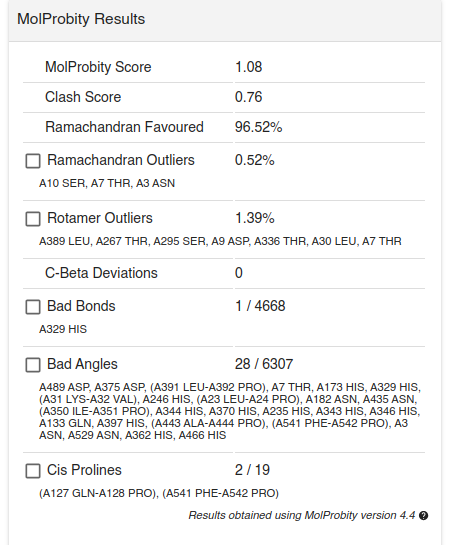
SWISS-MODEL Repository

SWISS-MODEL url: <https://swissmodel.expasy.org/interactive>

Figure 9: Building New Model Page in SWISS-MODEL

Figure 10: 3D Structure for Hypothetical Protein

Figure 11: Ramachandran Plots for Assessment of Protein Structure

Figure 12: MolProbity Results as Assessment of Protein Structure

Ramachandran Favoured score is 96.52%.

6 Domains and 1 Alpha Fold.

Domains are:

* GGDEF\_dom - IPR000160
* GGDEF - TIGR00254
* GGDEF - SM00267
* GGDEF - PS50887
* GGDEF - PF00990
* GGDEF - cd01949

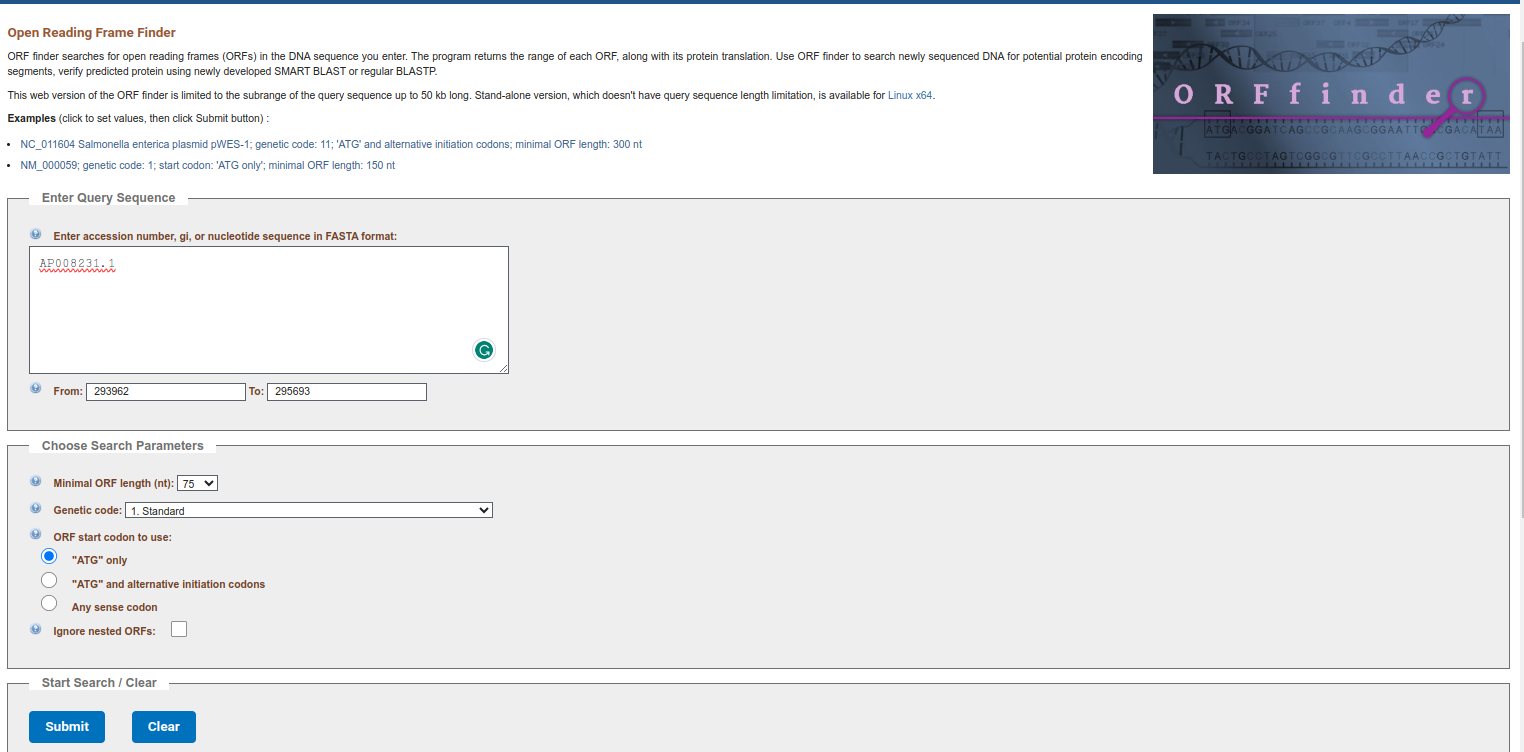
Additional Info:

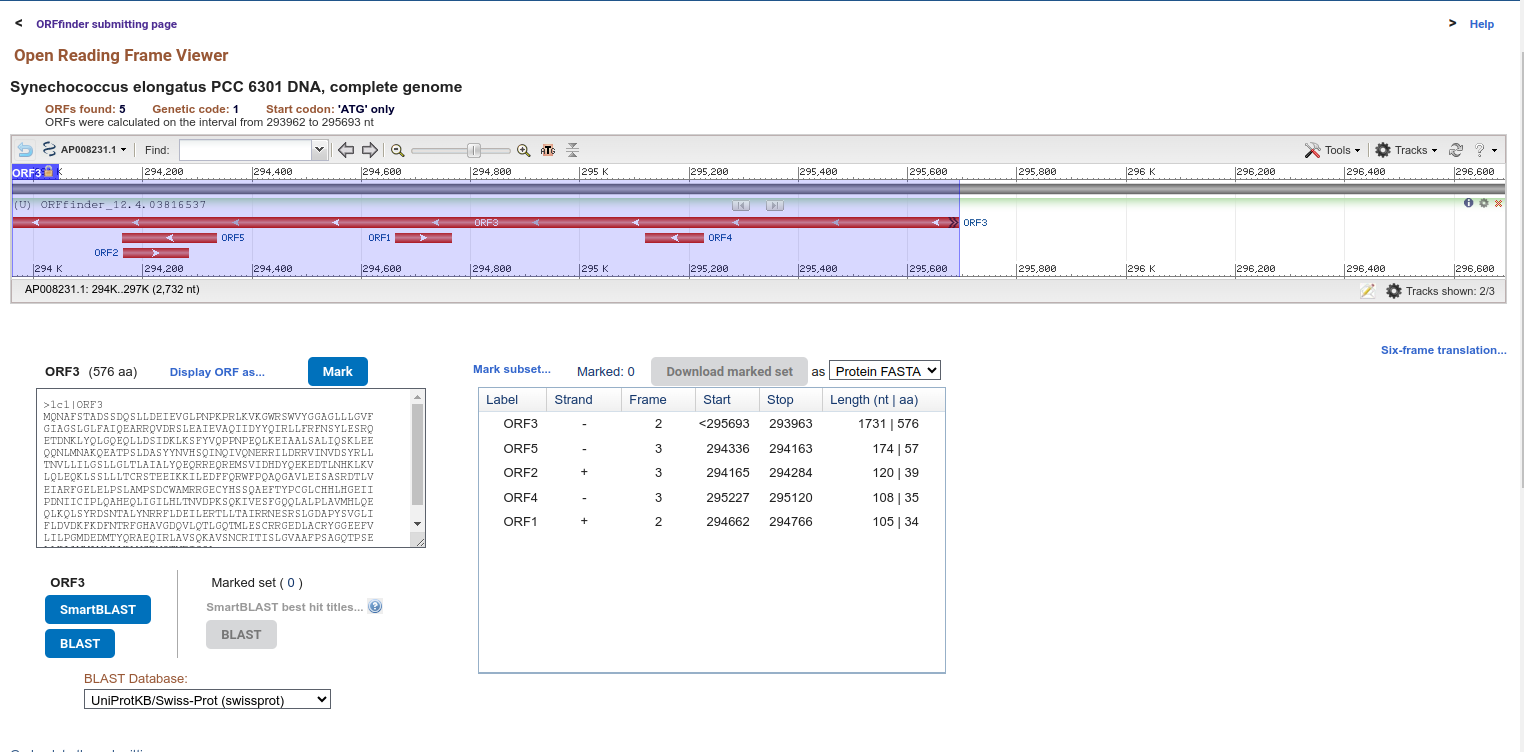
* [https://www.uniprot.org/uniprotkb/A0A0H3JZF6/entry#subcellular\_location](https://www.uniprot.org/uniprotkb/A0A0H3JZF6/entry" \l "subcellular_location)
* <https://www.ebi.ac.uk/interpro/protein/UniProt/A0A0H3JZF6/>

**Open Reading Frame Analysis:**  
  
NCBI orffinder url: https://www.ncbi.nlm.nih.gov/orffinder/

Start Codon: ATG

Stop Codons: TAA, TAG, TGA

Figure 13: NCBI OrfFinder Search Page

Figure 14: NCBI OrfFinder Result Page