COMPUTATIONAL ANALYSIS OF UNKNOWN DNA SEQUENCES USING INTEGRATED BIOINFORMATICS TOOLS

DNA SEQUENCE ANALYSIS

Download GenBank Graphics

Homo sapiens cyclin dependent kinase inhibitor 1A (CDKN1A), transcript variant 1, mRNA

Sequence ID: NM_000389.5 Length: 2117 Number of Matches: 1

Range 1: 1 to 2117 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
3910 bits(2117)	0.0	2117/2117(100%)	0/2117(0%)	Plus/Plus

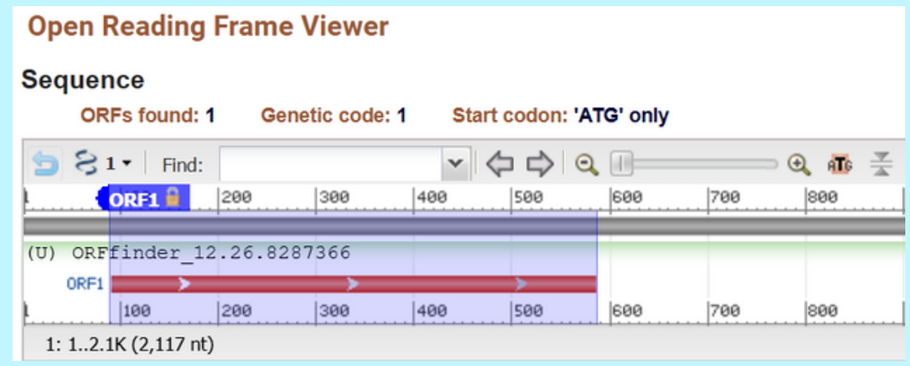
```
Query  1    GAGGTGTGAGCAGCTGCCGAAGTCAGTTCCTTGTGGAGCCGGAGCTGGGCGCGGATTCGC  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    GAGGTGTGAGCAGCTGCCGAAGTCAGTTCCTTGTGGAGCCGGAGCTGGGCGCGGATTCGC  60

Query  61   CGAGGCACCGAGGCACTCAGAGGAGGCGCCATGTCAGAACCGGCTGGGGATGTCCGTCAG  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   CGAGGCACCGAGGCACTCAGAGGAGGCGCCATGTCAGAACCGGCTGGGGATGTCCGTCAG  120

Query  121  AACCCATGCGGCAGCAAGGCCTGCCGCCGCCTCTTCGGCCCAGTGGACAGCGAGCAGCTG  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACCCATGCGGCAGCAAGGCCTGCCGCCGCCTCTTCGGCCCAGTGGACAGCGAGCAGCTG  180

Query  181  AGCCGCGACTGTGATGCGCTAATGGCGGGCTGCATCCAGGAGGCCCGTGAGCGATGGAAC  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  AGCCGCGACTGTGATGCGCTAATGGCGGGCTGCATCCAGGAGGCCCGTGAGCGATGGAAC  240
```

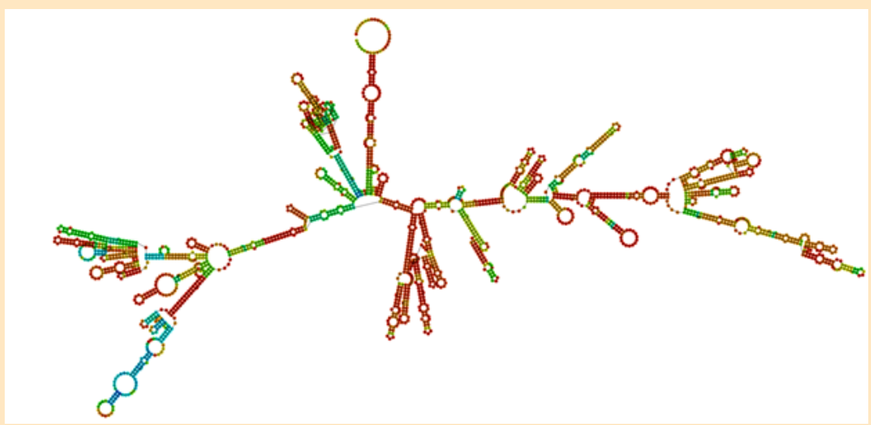
The BLASTn analysis shows a high-identity match of the query sequence with Homo sapiens **CDKN1A (p21) mRNA**, confirming the sequence corresponds to the CDKN1A gene.



ORF analysis identified a continuous open reading frame of expected length, indicating that nucleotide sequence has protein-coding potential.

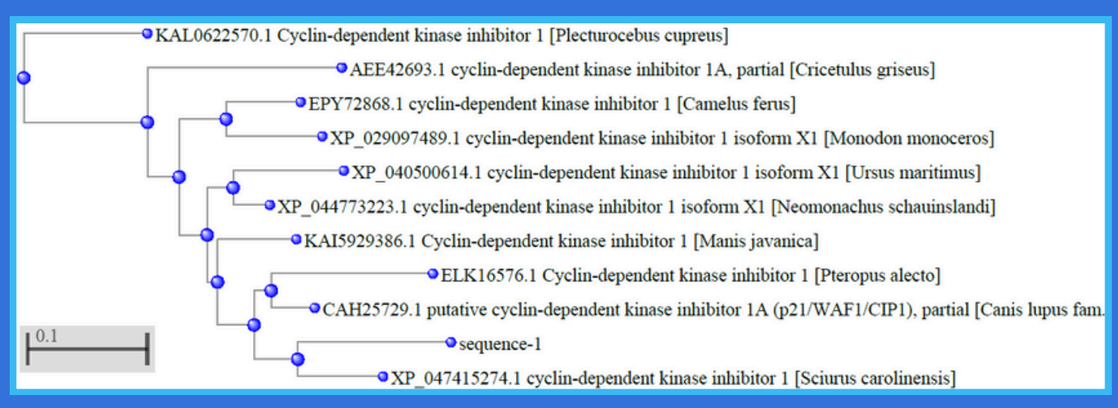
ExPASy translation revealed a valid amino acid sequence without premature stop codons, supporting the correct translation of the CDKN1A coding region

RNA ANALYSIS



RNA secondary structure predicted using **RNAfold**. The minimum free energy (MFE) structure has a free energy of −847.20 kcal/mol. The thermodynamic ensemble free energy is −876.35 kcal/mol, with the MFE structure occurring at 0.00% frequency, indicating high structural flexibility and a diverse folding ensemble (ensemble diversity = 390.52).

PHYLOGENETIC ANALYSIS



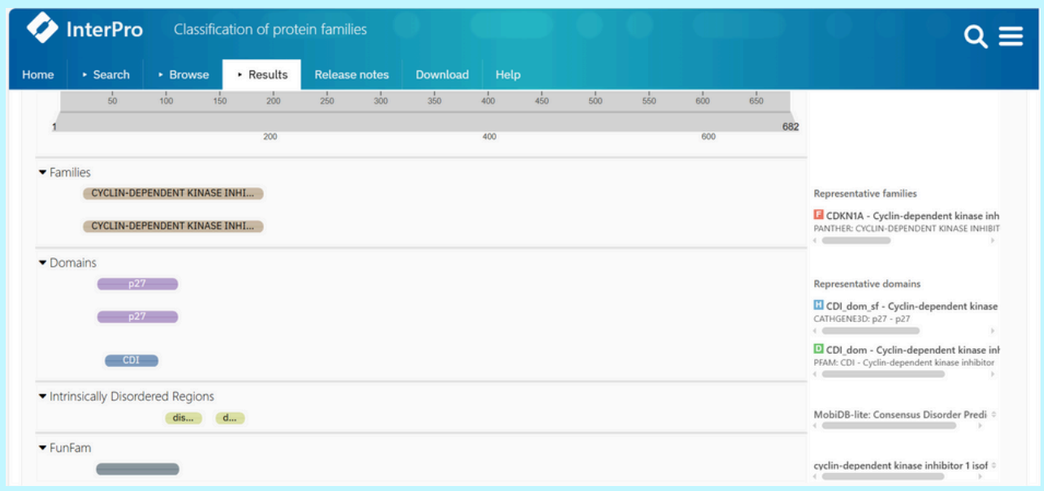
Phylogenetic tree of cyclin-dependent kinase inhibitor 1 (CDKN1A/p21) protein sequences from selected mammalian species constructed using NCBI COBALT. The tree illustrates evolutionary relationships based on multiple sequence alignment. Branch lengths represent evolutionary distance, with the scale bar indicating substitutions per site (0.1). The query sequence (sequence-1) clusters closely with mammalian CDKN1A homologs, confirming its evolutionary conservation.

PROTEIN SEQUENCE ANALYSIS

PROPERTIES

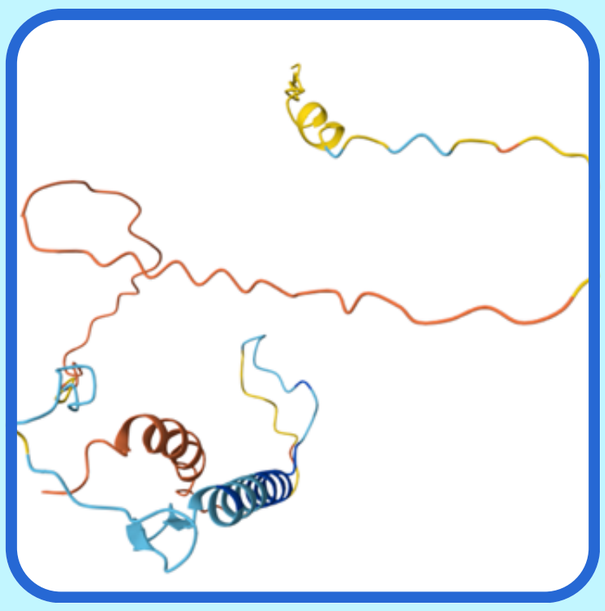
Number of amino acids	682
Theoretical pI	9.29
Molecular weight	74609.3
Instability index	73.89
Aliphatic index	69.79
GRAVY	0.293

protparam



InterPro domain analysis confirms that the translated sequence contains CDKN1A-specific conserved regions, validating its identity as a cyclin-dependent kinase inhibitor (p21) and supporting its known regulatory function in cell cycle control.

PROTEIN STRUCTURE ALPHA FOLD



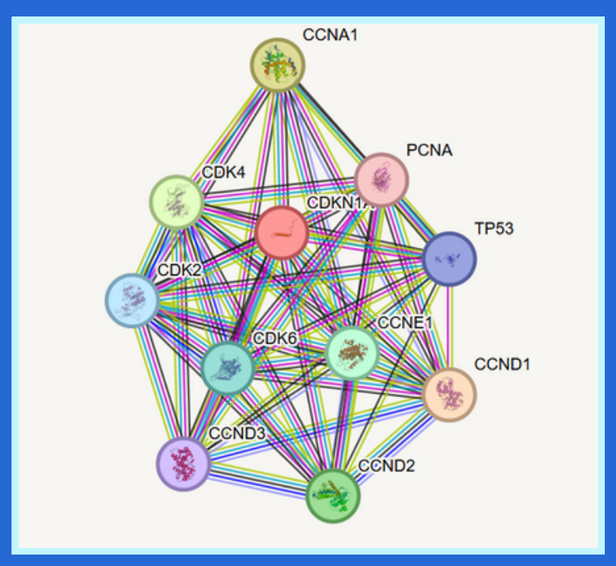
SUBCELLULAR LOCALIZATION

Sequence
Predicted localizations: Nucleus
Predicted signals: Nuclear localization signal

Localization	Cytoplasm	Nucleus	Extracellular	Cell membrane	Mitochondrion	Plastid	Endoplasmic reticulum	Lysosome/Vacuole	Golgi apparatus	Peroxisome
Probability	0.3635	0.9316	0.0375	0.0735	0.1148	0.0060	0.0809	0.0520	0.0770	0.0763

Both **DeepLoc and PSORT** analyses consistently predict nuclear localization of the protein, with high nuclear probability scores and minimal likelihood of alternative subcellular compartments

PROTEIN-PROTEIN INTERACTION



The STRING network identifies CDKN1A (p21) as a central hub linking cyclins, CDKs, PCNA, and TP53, highlighting its critical role in cell cycle inhibition, DNA damage response, and maintenance of genomic stability.