

Computational Analysis of Unknown DNA Sequence Using Integrated Bioinformatics Tools



AGCAGGCAGCGGGCGGGCGGGAGCAGCATGGAGCCGCGGGCGGGGAGCAGCATGGAGCCTTCGGCTGACTGGCTGGCCACGGCCGCGCCCGGGGTCGGG
TAAGGAGGTGCGGGCGCTGCTGGAGCGGGGCGCTGCCAACGCACCGAATAGTTACGTCGAGGCCGATCCAGGTCATGATGATGGGCAGCGCCCGAGT
GGCGGAGCTGCTGCTGCCAGCGCGGAGCCAACTGCGCCGACCCCGCCACTCTACCCGACCCGTGCACGACGTCGCCGGGAGGGCTTCTGGACACG
CTGGTGGTCTGCACCGGGCGGGGCGGGCTGGACGTGCGCGATGCCTGGGCGCTGTCCTGGACCTGGCTGAGGAGCTGGGCCATCGCGATGTCGCACG
GTACCTGCGCGCGCTGCGGGGGGACCAGAGGCAGTAACCATGCCGCATAGATGCCGCGAAGGTCCCTCAGACATCCCGATTGAAAGAACCAGAGAGGC
TCTGAGAAACCTCGGAAACTTAGATCATCAGTCACCGAAGGTCTACAGGGCCACAACCTGCCCGCCACAACCCACCCGCTTTCGATGTTTTCATTAGAAAA
TAGAGCTTTTAAAAATGCTCTGCCTTTTAAACGTAGATATATGCTTCCCCACTACCGTAAATGTCCATTATATATTTTATATATTCTTATAAAATGTAAAAAGA
AAAACACCGCTTCTGCCTTTTCACTGTGTTGGAGTTTCTGGAGTGAGCACTACGCCCTAAGCGCACATTCATGTGGGCATTCTTGCAGGCCTCGCAGCCTCCG
GAAGTGTGCACTTCATGACAAGCATTTTGTGAAGTAGGGAAGCTCAGGGGGGTACTGGCTTCTTGTAGTCACACTGCTAGCAATGGCAGAACCAAGCTCA
AATAAAAAATAAATAATTTTCATTCACTCACT

GENE: CDKN2A
ORGANISM: HOMO SAPIENS
LOCATION: CHROMOSOME 9
PROTEIN: P14
BP: 138394717
AMINO ACIDS: 153
ACCESSION NO: NC_000009.12

SUBMITTED BY: SYEDA
AYESHA MAHMOOD
BS 1TH (M)



DNA SEQUENCE ANALYSIS

NCBI BLASTN

Range 1: 4 to 977 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

	Score	Expect	Identities	Gaps
	1799 bits(974)	0.0	974/974(100%)	0/974(0%)

Query 1 AGCAGGCAGCGGGCGGGCGGGAGCAGCATGGAGCCGCGGGCG
Sbjct 4 AGCAGGCAGCGGGCGGGCGGGAGCAGCATGGAGCCGCGGGCG
Query 61 TCGGCTGACTGGCTGGCCACGGCCGCGGGCCGGGTCGGGT
Sbjct 64 TCGGCTGACTGGCTGGCCACGGCCGCGGGCCGGGTCGGGT
Query 121 CTGGAGGCGGGGCGCTGCLCAACGCACCGAAAGTTACGGT
Sbjct 124 CTGGAGGCGGGGCGCTGCLCAACGCACCGAAAGTTACGGT
Query 181 ATGATGATGGGCAGCGCCCGAGTGGCGGAGCTGCTGCTCT
Sbjct 184 ATGATGATGGGCAGCGCCCGAGTGGCGGAGCTGCTGCTCT

ORF FINDER

NIH National Library of Medicine
National Center for Biotechnology Information

ORF Finder submitting page
Open Reading Frame Viewer

Homo sapiens cyclin dependent kinase inhibitor 2A (CDKN2A), transcript variant 1, mRNA

ORFs found: 7 Genetic code: 1 Start codon: ATG only

ORF Finder: 12-22:19174751

ORF1 (156 aa) Display ORF as: Mark

Marked set (0)

Label	Strand	Frame	Start	Stop	Length (nt)
ORF1	+	1	31	501	471 (156)
ORF4	-	1	456	51	456 (152)
ORF5	-	2	491	159	353 (117)
ORF3	+	2	467	407	141 (47)
ORF2	+	2	353	466	114 (38)
ORF6	-	3	865	785	81 (27)
ORF7	-	3	691	611	81 (27)

EXPASY TRANSLATE

RNA sequence:

CCGGGGUCGGGUAGAGGAGGUGCGGGCGCUGCGUAGGCGGGGGCGCUGCCCAACGCACCGAAU
AGUUAACGGUCGGAGGCCGAUCCAGGUCUAUGAUGAUGGGCAGCGCCCGAGUGGCGGAGCUGCUGC
UGCUCACCGCGCGGAGCCCAACUGCGCCGACCCCGCCACUCUCACCCGACCCGUGCACGACGCG
UGCCCGGGAGGGCUUCUGGACACGUGGUGGUGCUGCACCAGGGCGGGGCGCGGUGGACGUG
CGCGAUGCCUGGGGGCGCUGGCCGUGGACUGGUGAGGAGCUGGGCGCAUCGCGAUGUCGCAC
GGUACCUUGCGCGCGCGCUGCGGGGGGACACAGAGGCAGUAACCAUGCCCGCAUAGAUGCCGCGGA
AGGUCCUCAGACAUCCCCGAUUGA

RNA ANALYSIS

RNA FOLD

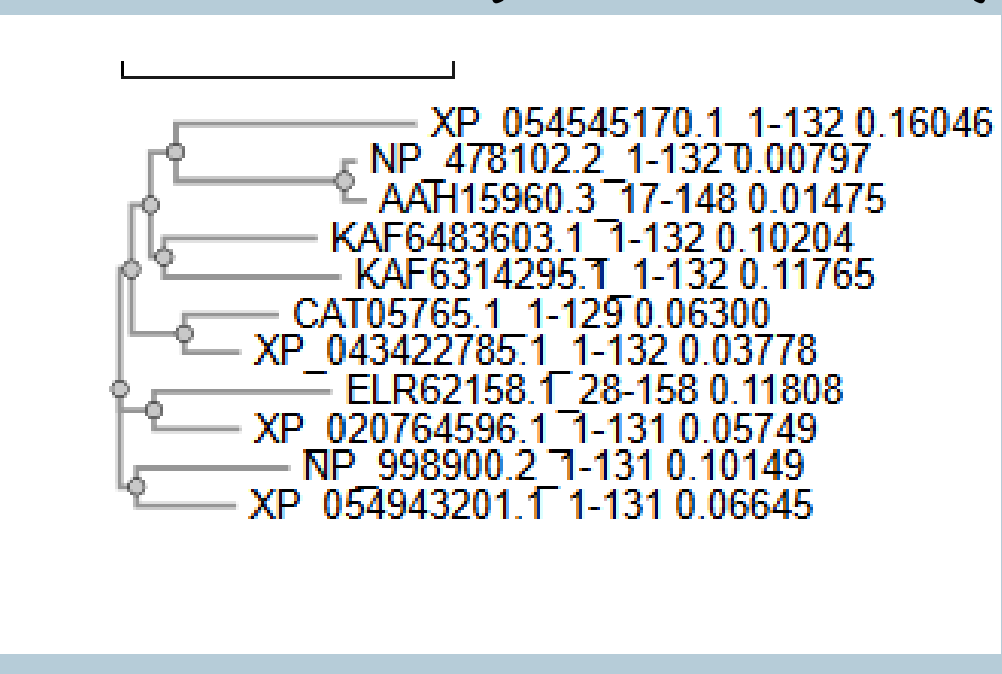
The highly negative Minimum Free Energy (MFE) value indicates that the mRNA transcript forms a stable, low-energy secondary structure that is thermodynamically favorable

PHYLOGENNETIC ANALYSIS

CLUSTAL MSA

```
(
  XP_054545170.1_1-132:0.16046,
  NP_478102.2_1-132:0.08797,
  AAH15960.3_17-148:0.01475)
:0.11227)
:0.01650,
(
  KAF6483603.1_1-132:0.10204,
  KAF6314295.1_1-132:0.11765)
:0.08907)
:0.01274,
(
  CAT05765.1_1-129:0.06300,
  XP_043422785.1_1-132:0.03778)
:0.03482)
:0.00788,
(
  ELR62158.1_28-158:0.11808,
  XP_020764596.1_1-131:0.05749)
:0.02296,
(
  NP_998900.2_1-131:0.10149,
  XP_054943201.1_1-131:0.06645)
```

CLUSTAL ALTREE(PHYLOGENETIC)

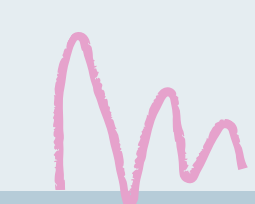


PROTEIN ANALYSIS STRUCTURE AND PROPERTIES

PROT PARAM

Number of amino acids: 153
Theoretical pI: 11.70
Molecular weight: 16989.79

PROPERTIES



INTER PRO

Entry matches to this protein⁰

AlphaFold Confidence

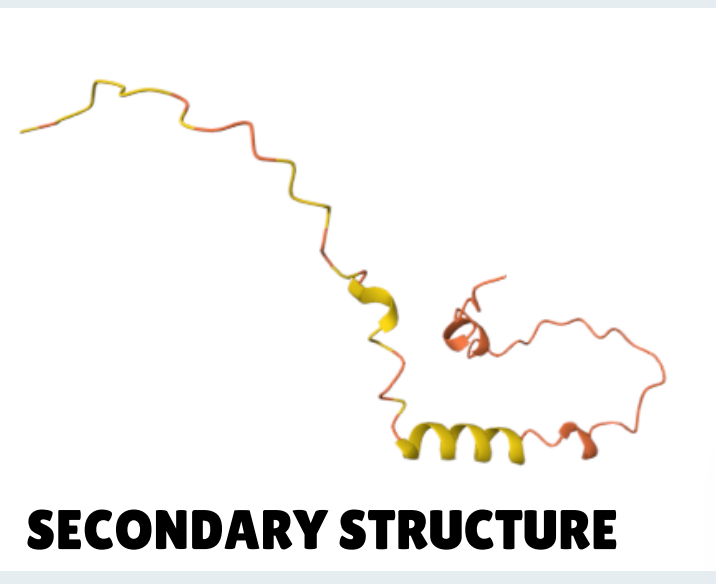
Families

Domains

Pathogenic And Likely Pathogenic Variants

cathepsinB G3DSA:1.25.40.20
Ankyrin repeat-containing domain
Integrated: IPR036770
1 - 134

ALPHA FOLD



SECONDARY STRUCTURE

SUBCELLULAR LOCALIZATION

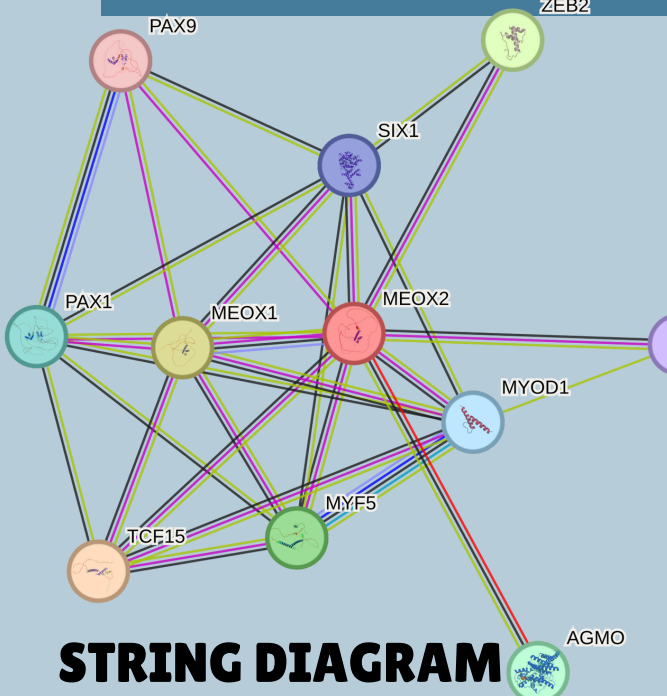
KAI006799.1
Predicted localizations: Cytoplasm, Extracellular
Predicted signals:

Localization	Cytoplasm	Nucleus	Extracellular	Cell membrane	Mitochondrion	Plastid	Endoplasmic reticulum	Lysosome/Vacuole	Golgi apparatus	Peroxisome
Probability	0.5231	0.4789	0.6884	0.0028	0.5563	0.0407	0.1800	0.0553	0.1532	0.0411

Sorting Signal Importance: Download: PNG / CSV
KAI006799.1
Predicted Signals:

Based on DeepLoc 2.0 analysis, the protein is primarily localized to the extracellular space (0.6884) and cytoplasm (0.5231). Moderate probabilities for the mitochondrion (0.5563) and nucleus (0.4789) suggest a multi-localization pattern. High-importance sorting signals between residues 40–50 and 120–130 likely facilitate its transition between the internal cellular environment and the external matrix.

PROTEIN PROTEIN INTERACTION



STRING DIAGRAM

STRING analysis identifies the protein as a key hub for somite development ($FDR = 4.74 \times 10^{-8}$) and embryonic segmentation ($FDR = 8.96 \times 10^{-6}$). High-confidence interactions with MEOX1, MEOX2, PAX1, and MYOD1 highlight its essential role in chordate development and RNA polymerase II transcription, confirming its significance in skeletal muscle patterning.