Computational Analysis of Unknown DNA Sequences Using Integrated Bioinformatics Tools

Official name: CD4 molecule
Gene type: protein coding
O**rganism:** Homo sapien
Also known as: OKT4; IMD79
CD4mut

This gene encodes the CD4 membrane glycoprotein of T lymphocytes. The CD4 antigen acts as a coreceptor with the T-cell receptor on the T lymphocyte to recognize antigens displayed by an antigen presenting cell in the context of class II MHC molecules

1 DNA ANALYSIS

1.1 BLAST

Nucleotide BLAST

Score	Expect	Identities	Gaps	Strand
5631 bits(3049)	0.0	3049/3049(100%)	0/3049(0%)	Plus/Plus

```
Query  1    CTCTCTTCATTTAAGCACGACTCTGCAGAAGGAACAAAGCACCCTCCCCACTGGGCTCCT  60
Sbjct  1    CTCTCTTCATTTAAGCACGACTCTGCAGAAGGAACAAAGCACCCTCCCCACTGGGCTCCT  60
Query  61   GGTTGCAGAGCTCCAAGTCCTCACACAGATACGCCTGTTTGAGAAGCAGCGGGCAAGAAA  120
Sbjct  61   GGTTGCAGAGCTCCAAGTCCTCACACAGATACGCCTGTTTGAGAAGCAGCGGGCAAGAAA  120
Query  121  GACGCAAGCCCAGAGGCCCTGCCATTTCTGTGGGCTCAGGTCCCTACTGGCTCAGGCCCC  180
Sbjct  121  GACGCAAGCCCAGAGGCCCTGCCATTTCTGTGGGCTCAGGTCCCTACTGGCTCAGGCCCC  180
```

	Name	Score	Score	Cover	value	Ident
	Homo sapiens	5631	5631	100%	0.0	100.00%
..	Homo sapiens	5489	5489	98%	0.0	99.70%
	Pan paniscus	5430	5430	100%	0.0	98.79%

1.2 ORF FINDER

ORFs found: 74 Genetic code: 1 Start codon: stop-to-stop

ORF14 WAS SELECTED BECAUSE IT IS AMONG THE LONGEST PREDICTED ORFS WITH A CLEAR START AND STOP CODON, INDICATING HIGH CODING POTENTIAL. ITS MINIMAL OVERLAP WITH OTHER ORFS MAKES IT A RELIABLE CANDIDATE FOR PROTEIN-CODING ANALYSIS.

1.3 DNA TO PROTEIN TRANSLATION

Expasy

5'3' Frame 1

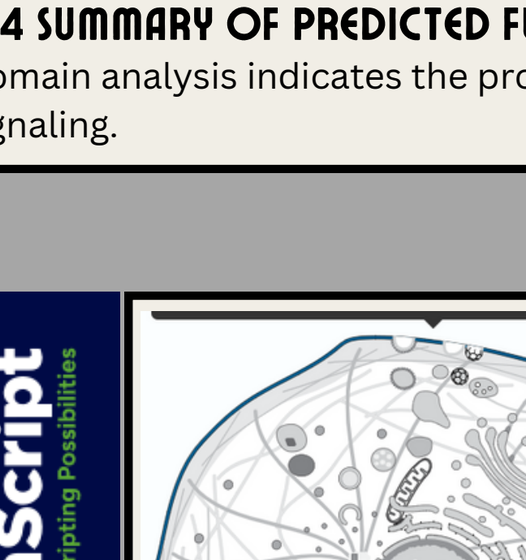
```
LSSFKHDSAEGTKHPPHWAPGCRAPSPHTDTPV-RAAGKKDASPEALPFLWAQVPTGSGPCLPRQGHNEPGSPF-ALASGAATGAPPSSHSGKESGAGQKRGYSGTDLYSFP
EEEHTIPLEKLQPDKDSGKSGLLLN-RSIQAE-SR-LKKKPLGPRKLSPDHQES-DRRLRYLHL-SGGPEGGGAIASVRIDCQL-HPPASGAEPDPDLGEPPW--PLSAN-E
SKG-KHTGGEDPLRVSAGAPG-WHLDMHCLAEPEEGGVQNRHRGASFPEGLQHSL-ERGGTGGVLLPTRLYS-KADGQWRAVVAGGEGLLQVLDHL-PEEQGSVCKTGYPG
P-APDGQEAPAPPHPAPGLASVCWLMKPHPGP-SENRKVASGSEPGGDESHSAPEKFDL-GVGTHLP-ADAEFETGEQGGKGLEAGEGGVGAEP-GGDVAVSAE-LGTGPAG
IQNQGSAHMVHPGAANGPDCAGGRRRPFAFHWARHLLLCQVFAPKAFSRADVSDQETPQ-EEDLFVSSPVSEDM-PHLRHEARQIPLAASPGVCPAFPACGPDECSRSPASG
LLFASSTICHCFSWVRPRLHWLSVAL-FPEA-SHRPPRHFLFLQA-PFSHYFSLTLSPLLIWIPGECSGPALAGMEGEAGCLEAWSMGLFFYKTGPWDHRGQELAQNHTAKP
VKDGCRSRGFWQPVPPAPCCPLLTLCGWDHRLTS-PCTNSPSGHSPMYTASRDVSHPLSI-DLEKSYKAGSDRTKMIISSL-TRTRAQRG-MIDYQVPD-QVLESGLTQVPC
PKFHCCLLNAGTNATRLSPVASGGYSMCTFGFTEAQHFWEGSISENLRAGMKASLSKIPPSSPAGGRICYQRTKPLALLIRAQAGSTGTAGENAQ-PVTDPVQNLLEASFAG
RGGS-PEREPSKGPQR-LCQALRLPHTLPYPPPDHSGHREIRVTNLLDPLLSGSPLFLPFLTTSLSPNSFSLFPSPPVFKACLFQEDPPIAAGAPHLLTLHLCPLSTPAPLS
-NKNTINL
```

5'5' Frame 1

```
SLLYFYFSSGEQGWRVGTNAK-ANGEPQQQ-GGLPGRGRL-RQEEKEIGKRSWD-GKW-GRVGREGILREVDQEDL-P-FPCVLNGLEEGKGGCGAGAEPGTITFEVP
SQASYPLSQQSSLPGGSAQGQ-LVTGHSLLQCLCSQLAL-LEEPKALSSGNRFCHQRGMKEGE-TGRPSSLLVNSLRWTLFMGAVLL-TQKYTLSTHH-PLVRAVWHL
RGSSGTWDKGPGLVPTPALASPALGNQSSSLSELWFWSINWR-SS-FCHCQPCRIFLSLK-TEDVRHPLRPCTWGCVQRGCLCKVRM-VCGPTHIG-EAGSMGQEVLA
FICIHP-LAWLCDFVQVPALCGPRVLSCKRTVPCSMLPDTQPHPPCQPGLALNTPLGSK-AVGRGSERNNERRARLEGKGNGVEDGVIKPLETREQHSTSEAGA-PRR
L-RRRTGGQRLGICYIHLVRRQETRGRHLGRLQVGSAWFRASNGATCLLKPVRTLAGLLLTEESLDLRHPLCLAPSVPAPDTEEDA-PNEKQEAGDAPQHNQGHWLHR
QNLDVGFQQDLSRVTQQTLPHPRLRVQHPHRLLPLRDLCLLVLQFQTQHQLRGGGSPHLTGQIFLELSGSHHHQVHFLMQLSCFRFKGQGEVSRASILRQGLGQGEV
AHLELRVLGNPFYRHFLVLQVKGDPRLGGGSPLRLPPQLATARQLFNCKGEWEGELHLFPLFLIDYAGGLLES-HHDVYFELHLLLVLQDSACPGATILELQLRHGE
LYVFTPWTPTLH-GATTRGALQGQGQALPLKQVGVRVGSQSEH-QLHLLLLVLHFTDVSI-VFYLKILDDQGKVSLVPKASS-VSAIIQLGWTFS-EGALISQNLYLV
/ELYALLLGSCTGQFHCIPFFAQHHFLSLSGCWEERQLQHQKQVPKRDSPVHCGLAEGGRGLSQ-GPEPTEMAGPLGLRLSCPLLLKQAYLCEDLELCNQEPSGEGAL
RA-MKR
```

2. RNA ANALYSIS

RNAfold WebServer



RNAfold analysis indicates a highly stable RNA structure (MFE = −1151.19 kcal/mol).

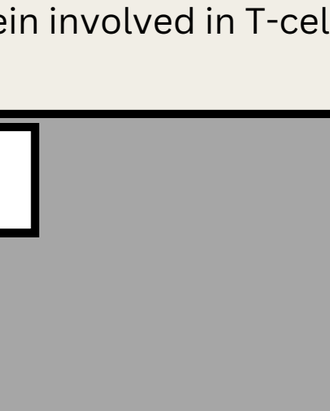
3 PHYLOGENETIC ANALYSIS

3.1 MSA

```
CLUSTAL O(1.2.4) multiple sequence alignment


.
------------------------------------------------  0
GTTGGGGGCCGGCCCAGCCCTAGGGTGTGGAGGAGCCTCGCCATCGGGCTTCCTGTCTCTTCATTTAAGC
ACGACTCTGCAGAAGGAACAAAGCACCCTCCCCACTGGGCTCCTGGTTGCAAAGCTCCAA    60
ATCAGAAGAGCTGTGGGCATTCTCTTGCCAGAGGTGCCATGTCTCCAAGGTACCTCAGCCTCTCCCATAT
------------------------------------------------  0
CTCTCTTCATTTAAGCACGACTCTGCAGAAGGAACAAAGCACCCTCCCCACTGGGCTCCTGGTTGCAAAG
------------------------------------------------  0
```

3.2 PHYLOGENETIC TREE

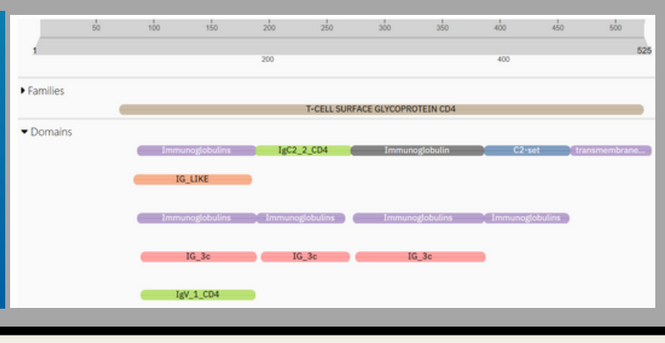


The phylogenetic tree indicates that human CD4 is evolutionarily conserved and closely related to mammalian CD4 homologs, highlighting its critical immune function.

4 PROTEIN ANALYSIS(STRUCTURE & PROPERTIES)

4.1 DOMAIN AND MOTIFS

InterPro



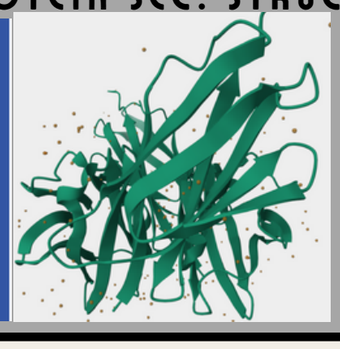
4.2 TABLE OF PROTEIN PROPERTIES

Expasy

Parameter	Value
Number of amino acids	525
Molecular weight (Da)	58,477.16
Theoretical pI	9.71
Instability index (II)	42.82 (Unstable)

4.3PROTEIN SEC. STRUCTURE

PSIPRED



4.4 SUMMARY OF PREDICTED FUNCTION:

Domain analysis indicates the protein is CD4, an immunoglobulin-like transmembrane glycoprotein involved in T-cell activation and immune signaling.

5 SUBCELLULAR LOCALIZATION

GenScript Scripting Possibilities

CELL MEMBRANE
THE CELL MEMBRANE IS THE SELECTIVELY PERMEABLE MEMBRANE WHICH SEPARATES THE CYTOPLASM FROM ITS SURROUNDINGS. KNOWN AS THE CELL INNER MEMBRANE IN PROKARYOTES WITH 2 MEMBRANES.

6 PROTEIN-PROTEIN INTERACTION

6.1 BIOLOGICAL SIGNIFICANCE

CD4 PROTEIN INTERACTIONS ARE CRUCIAL FOR THE PROPER FUNCTIONING OF THE ADAPTIVE IMMUNE SYSTEM, PRIMARILY BY FACILITATING T-HELPER (CD4\({}^{+}\) T) CELL ACTIVATION, DIFFERENTIATION, AND COMMUNICATION WITH OTHER IMMUNE CELLS. THESE INTERACTIONS ARE ALSO NOTABLY EXPLOITED BY THE HUMAN IMMUNODEFICIENCY VIRUS (HIV) FOR CELLULAR ENTRY.

STRING