



Computational Analysis Of Unknown DNA Sequence Using Integrated Bioinformatics Tools



INTRODUCTION

The CFTR gene encodes an ATP-regulated chloride channel that maintains ion and fluid balance in epithelial tissues. It has specialized domains which enable controlled ion transport, essential for normal lung and pancreatic function.

DNA ANALYSIS

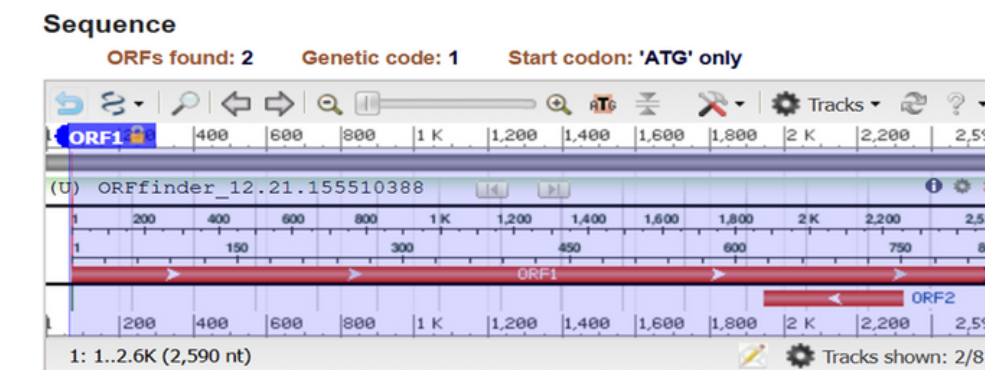
Descriptions	Graphic Summary	Alignments	Taxonomy
Sequences producing significant alignments			
	Description	Scientific Name	Max Score
✓	Homo sapiens CF transmembrane conductance regulator (CFTR), mRNA	Homo s...	4783
✓	Human cystic fibrosis mRNA, encoding a pre...	Homo s...	4767
✓	Pan troglodytes CF transmembrane conductance regulator (CFTR), mRNA	Pan tro...	4745

Homo sapiens CF transmembrane conductance regulator (CFTR), mRNA			
Sequence ID: NM_000492.4 Length: 6070 Number of Matches: 1			
Range 1: 1 to 2590			
Score	Expect	Identities	Gaps
4783 bits(2590)	0.0	2590/2590(100%)	0/2590(0%)
Query	1	GTAGTAGGTCTTTGGCATTAGGAGCTTGAAGCCAGAGCGCCCTAGCAGGACCCAGCCAGCSC	60
Sbjct	1	GTAGTAGGTCTTTGGCATTAGGAGCTTGAAGCCAGAGCGCCCTAGCAGGACCCAGCCAGCSC	60
Query	61	CCGAGAGACCATCGAGAGTCCGCTCTGGAAGAGCCAGCGCTTGTCTCCAACCTTTT	120
Sbjct	61	CCGAGAGACCATCGAGAGTCCGCTCTGGAAGAGCCAGCGCTTGTCTCCAACCTTTT	120
Query	121	CAGCTGACAGACCAATTTTGAAGAAAGATACAGACAGCCCTGGAATTGTCCAGACAT	180
Sbjct	121	CAGCTGACAGACCAATTTTGAAGAAAGATACAGACAGCCCTGGAATTGTCCAGACAT	180

blastn

The first hit represents fully annotated human CFTR mRNA sequence with 100% identity and coverage, making it the most accurate and biologically relevant match for the query

ORF FINDER



ORF 1 was selected because it is the longest continuous reading frame, containing a clear start and stop codon, making it the most likely to encode a functional protein

EXPASY

5'3' Frame 1
MQRSPLEKASVSKLFFSWTRPILRKGYRQRELSDIYQIPSVDSADNLSEKLEREWRELASKK
NPKLINALRRCFFWRMFYGFILYLGVEYTKAVQPLLGRITIASYPDNKEERSIAYLGLCLL
FIVRTLHLHPAIFGLHHIGQMRIAMFSLIYKTKLSSRVLDKISIGQLVSLSSNNLNKFDEGL
ALAHFVWIAPLQVALIMGLIWELLQASAFCGLGFLIVLALFQAGLGRMMKYRQDRAGKISERLV
ITSEMIENIQSVKAYCWEAEKMIENLRQTEKLTTRKAAVYKFNSSAFFSGFFVFLVSLPY
ALIKGILRKIFTTISFCIVLRMAVTRQFFWAVQTYWYDSLGAINKIQDFLQKQEKYKLTLENTTT
EUVVMENVTAFEWEEGFGELFEKAKQNNNRKTSNGDLSLFFSNFSLILGTPVLKDNFKIERGQLLA
VAGSTGAGKTSILMVMIGELEPSEGGIKHSGRISFCQSQFWSWIMPGTIKENIFGVSYDEYRYSV
IKACQLEEDISKFAEKDNIVLGEGGITLSGGQARISLARAVYKDADLYLLDSPFGYLDVLTKE
IFESCVCKLANKTRILVTSKMEHLKKADKILLHGGSSYFYGTSELQNLQPDFSSKLMGCDSE
DQFSAERRNSILTEHLRFSLLEGDAPVSWTETKKQSFQKTGEFGEKRKNSILNPINSIRKFSIVQ
KPTLQMNIGIEDSDEPLERRLSLVPDSEQGEAILPRISVISTGTPTLQARRQSVNLMTSHVNOG
QNIHRKTTASTRKVSLAPQANLTEDIYSRRLSQETGLEISEEINEEDLKECFDDMESI

LOCALIZATION

Results of the k-NN Prediction

PSORT
k = 9/23
55.6 %: endoplasmic reticulum
22.2 %: vacuolar
11.1 %: nuclear
11.1 %: Golgi
>> prediction for 176664918422686 is end (k=9)

The protein is primarily localized to the ER, as indicated by its signal peptide, transmembrane domain, and ER-related predictions, suggesting it undergoes trafficking through the secretory pathway

DeepLoc

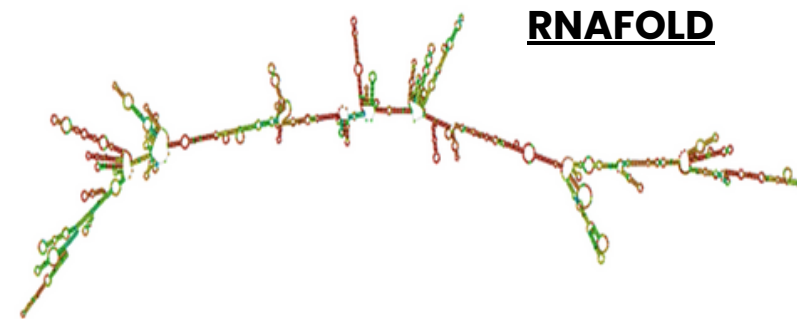
Sequence

Predicted localizations: Cell membrane, Lysosome/Vacuole

Predicted signals: Signal peptide, Transmembrane domain

Localization	Cytoplasm	Nucleus	Extracellular	Cell membrane	Mitochondrion	Plastid	Endoplasmic reticulum	Lysosome/Vacuole	Golgi apparatus	Peroxisome
Probability	0.0677	0.1588	0.0357	0.6748	0.1385	0.0196	0.4558	0.6736	0.4227	0.2516

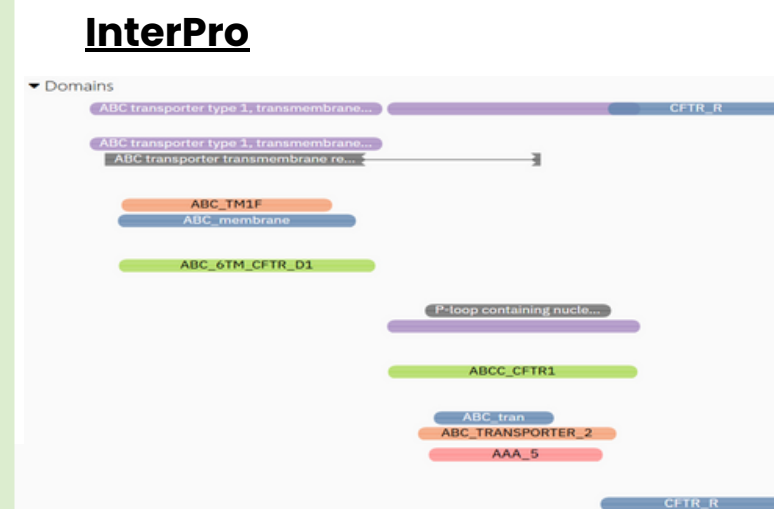
RNA ANALYSIS



RNAFOLD

The RNA is thermodynamically stable due to its very low minimum free energy of -673.70 kcal/mol.

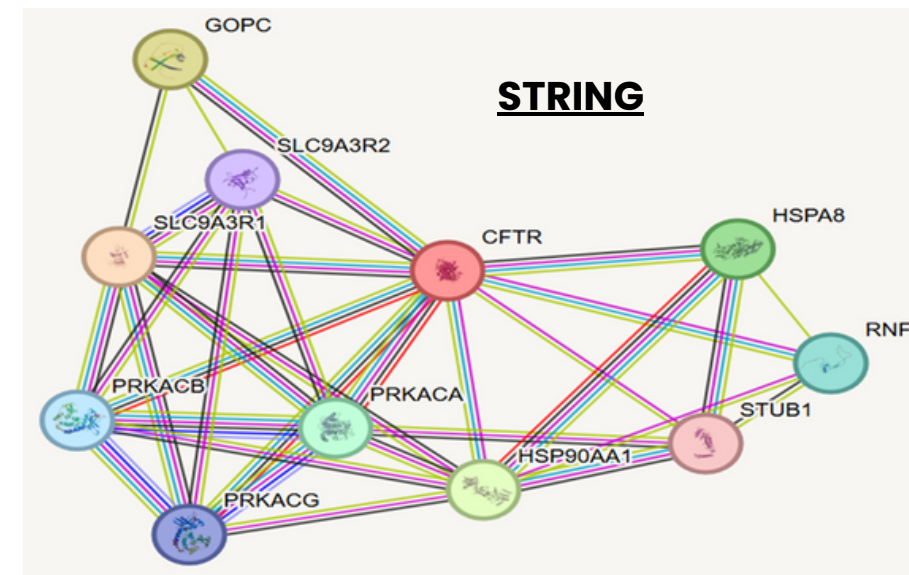
PROTEIN ANALYSIS



ProtParam

No of AA	840
pI	8.67
MW	95883.04
II	43.06
AI	96.94

INTERACTIONS



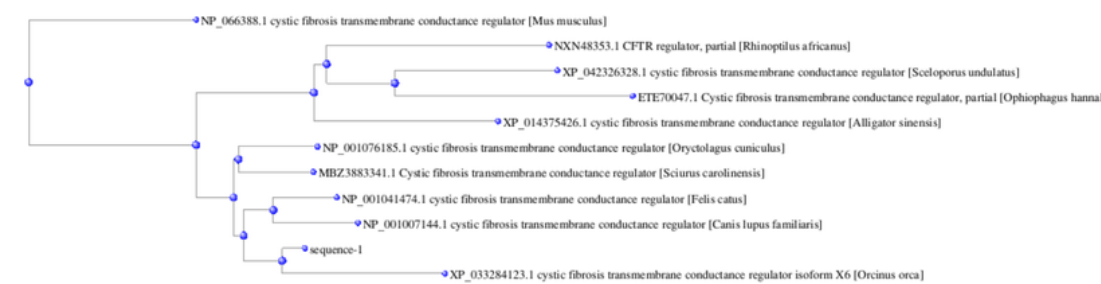
STRING

PHYLOGENY

MSA

Query_10001 1 MQRSPLEKASVSKLFFSWTRPILRKGYRQRELSDIYQIPSVDSADNLSEKLEREWRELASKK 68
Query_10002 1 ----ETQFFFIIPNTPILRKGYRQRELSDIYQIPSVDSADNLSEKLEREWRELASKK 62
Query_10003 1 MQRSPLEKASVSKLFFSWTRPILRKGYRQRELSDIYQIPSVDSADNLSEKLEREWRELASKK 68
Query_10004 1 MQRSPLEKASVSKLFFSWTRPILRKGYRQRELSDIYQIPSVDSADNLSEKLEREWRELASKK 68
Query_10005 1 MQRSPLEKASVSKLFFSWTRPILRKGYRQRELSDIYQIPSVDSADNLSEKLEREWRELASKK 68
Query_10006 1 ----HSELVAVQSLQGGHTTILRKGYRQRELSDIYQIPSVDSADNLSEKLEREWRELASKK 66
Query_10007 1 MQRSPLEKASVSKLFFSWTRPILRKGYRQRELSDIYQIPSVDSADNLSEKLEREWRELASKK 68
Query_10008 1 MQRSPLEKASVSKLFFSWTRPILRKGYRQRELSDIYQIPSVDSADNLSEKLEREWRELASKK 68
Query_10009 1 [19]ARRS[12]TTTQGGVENVHPR----HTPILRKGYRQRELSDIYQIPSVDSADNLSEKLEREWRELASKK 99
Query_10010 1 [12]HSS[15]PTREASVAG----HTPILRKGYRQRELSDIYQIPSVDSADNLSEKLEREWRELASKK 92
Query_10011 1 MQRSPLEKASVSKLFFSWTRPILRKGYRQRELSDIYQIPSVDSADNLSEKLEREWRELASKK 68
Query_10001 69 LINALRCFFAFRFPYGLLYLVEYTKAVQPLLGRITIASYPDNKEERSIAYLGLCLLPIVRLTHLHPAIFGLH 148
Query_10002 63 LINALRCFFAFRFPYGLLYLVEYTKAVQPLLGRITIASYPDNKEERSIAYLGLCLLPIVRLTHLHPAIFGLH 142
Query_10003 69 LINALRCFFAFRFPYGLLYLVEYTKAVQPLLGRITIASYPDNKEERSIAYLGLCLLPIVRLTHLHPAIFGLH 148
Query_10004 69 LINALRCFFAFRFPYGLLYLVEYTKAVQPLLGRITIASYPDNKEERSIAYLGLCLLPIVRLTHLHPAIFGLH 148
Query_10005 69 LINALRCFFAFRFPYGLLYLVEYTKAVQPLLGRITIASYPDNKEERSIAYLGLCLLPIVRLTHLHPAIFGLH 148
Query_10006 67 LINALRCFFAFRFPYGLLYLVEYTKAVQPLLGRITIASYPDNKEERSIAYLGLCLLPIVRLTHLHPAIFGLH 146
Query_10007 69 LINALRCFFAFRFPYGLLYLVEYTKAVQPLLGRITIASYPDNKEERSIAYLGLCLLPIVRLTHLHPAIFGLH 148
Query_10008 69 LINALRCFFAFRFPYGLLYLVEYTKAVQPLLGRITIASYPDNKEERSIAYLGLCLLPIVRLTHLHPAIFGLH 148
Query_10009 100 LINALRCFFAFRFPYGLLYLVEYTKAVQPLLGRITIASYPDNKEERSIAYLGLCLLPIVRLTHLHPAIFGLH 179
Query_10010 93 LINALRCFFAFRFPYGLLYLVEYTKAVQPLLGRITIASYPDNKEERSIAYLGLCLLPIVRLTHLHPAIFGLH 172
Query_10011 69 LINALRCFFAFRFPYGLLYLVEYTKAVQPLLGRITIASYPDNKEERSIAYLGLCLLPIVRLTHLHPAIFGLH 148

COBALT TREE



CONCLUSION

Bioinformatic analysis confirms that CFTR is a highly conserved and functionally important protein with well-defined structural and regulatory domains. Its stable mRNA and characteristic ATP-binding and membrane-spanning regions further support its essential role in epithelial ion transport and cellular homeostasis

REFERENCES

- NCBI
- ORF FINDER
- EXPASY
- RNAFOLD
- PSORT
- DeepLoc
- InterPro
- ProtParam
- COBALT
- STRING