

Task 3: Subcellular Localization Prediction

Instruction By scrutinizing the protein's amino acid composition and sequence motifs, forecast its intracellular localization in eukaryotic cells.

Input

MEDEAVLDRGASFLKHVCDEEEVEGHHTIYIGVHVPKSYRRRRRHKRKTGHREKKEKERISENYSDKSDVENADESSSSILKPLISPAAERIRFILGEEDDSPAPPQLFTELDELLAVDQGEMEWKETARWIKFEEKVEQGGERWSKPHVATLSLHSLFELRTCMEKGSIMLDREASSLPQLVEMIVDHQIETGLLKPDLDKDKVITYTLRKHRRHQTKKSNLRSLADIGKTVSSASRMFTNPDNGSPAMTHRNLTSSSLNDISDKPEKDQLKNKFMKKLPRDAEASNVLVGEVDFLDSPFI AFVRLQQA VMLGALTEVPVPT RFLFILLGPKGKA KSYHEIGRAIATLMSDEVFHDIA YKAKDRQDLIAGIDEFLDEVIVLPPGEWDPAIRIEPPKSLPSSDKRKNMYSGGENVQMNGDTPPDGGHGGGGGHADCEELQRTGRFCGGLIKDIKRKAPFFASDFYDALNIQALSAILFIYLATVTN AITFGGLLGDATDNMQGVLESFLGTAVSGAIFCLFAGQPLTILSSTGPVLVFERLLFNFSKDHNF DYLEFRLWIGLWSAFLCLILVATDASFLVQYFTRFTEEGFSSLSIFIFYDAFKKMIKLADYYPINSNFKVG YNTQFSCVCMPPDPVNISVSNDTTLAPEDLPTISSSNMYHNATFDWAFLT TKECLKYGGKLVGN NCGFVPDITLMSFILFLGTYTSSMALKKFKTSPYFPTTARKLISDFAILPILIFCVIDALVGVDTPKLIVPSEFKPTSPNRGW FVAPFGGNPWVWYLA AAPALLVTILIFMDQQITAVIVNRKEHKLKKGAGYHLDLFWVAILMVVCSFMALPWYVAATVISIAHIDSLKMETETSAPGEQPKFLGVREQ RVTGTLVFILTGLSVFMAPILKFIMPVLYGVFLYMGVASLNGVQFMDRLKLLMPLKHQP DFIYLRHVPLRRVHLFTFLQVLC LALLWILKSTVAAIIFPV MILALVAVRKGMDYLF SQHDL SFLDDVIPEKD KKKKKEDEK KKKKKKGSVDS DNDSDC PYSEK VPSIKIPMDIMEQQPFLSDSKPSDRERSPTFLERHTSC

Ground Truth membrane

OPI-tuned model membrane

To predict the intracellular localization of the given protein sequence in eukaryotic cells, we can analyze its amino acid composition, sequence motifs, and structural characteristics that are typically associated with specific cellular compartments. Here's a detailed analysis:

1. Signal Peptide for Secretion or Membrane Localization: The sequence contains several stretches of hydrophobic amino acids (e.g., LAVDQGEMEWKETARWIKFEEKVEQGGERWSKPHVATLSLHSLFELRTCMEKGSIMLDREASSLPQLVEMIVDHQIETGLLKPDLDKDKVITYTLRKHRRHQTKKSNLRSLADIGKTVSSASRMFTNPDNGSPAMTHRNLTSSSLNDISDKPEKDQLKNKFMKKLPRDAEASNVLVGEVDFLDSPFI AFVRLQQA VMLGALTEVPVPT RFLFILLGPKGKA KSYHEIGRAIATLMSDEVFHDIA YKAKDRQDLIAGIDEFLDEVIVLPPGEWDPAIRIEPPKSLPSSDKRKNMYSGGENVQMNGDTPPDGGHGGGGHADCEELQRTGRFCGGLIKDIKRKAPFFASDFYDALNIQALSAILFIYLATVTN AITFGGLLGDATDNMQGVLESFLGTAVSGAIFCLFAGQPLTILSSTGPVLVFERLLFNFSKDHNF DYLEFRLWIGLWSAFLCLILVATDASFLVQYFTRFTEEGFSSLSIFIFYDAFKKMIKLADYYPINSNFKVG YNTQFSCVCMPPDPVNISVSNDTTLAPEDLPTISSSNMYHNATFDWAFLT TKECLKYGGKLVGN NCGFVPDITLMSFILFLGTYTSSMALKKFKTSPYFPTTARKLISDFAILPILIFCVIDALVGVDTPKLIVPSEFKPTSPNRGW FVAPFGGNPWVWYLA AAPALLVTILIFMDQQITAVIVNRKEHKLKKGAGYHLDLFWVAILMVVCSFMALPWYVAATVISIAHIDSLKMETETSAPGEQPKFLGVREQ RVTGTLVFILTGLSVFMAPILKFIMPVLYGVFLYMGVASLNGVQFMDRLKLLMPLKHQP DFIYLRHVPLRRVHLFTFLQVLC LALLWILKSTVAAIIFPV MILALVAVRKGMDYLF SQHDL SFLDDVIPEKD KKKKKEDEK KKKKKKGSVDS DNDSDC PYSEK VPSIKIPMDIMEQQPFLSDSKPSDRERSPTFLERHTSC) indicative of potential transmembrane domains. This suggests that the protein could be localized to the plasma membrane or internal membranes like the endoplasmic reticulum (ER).

2. Nuclear Localization Signals (NLS): The sequence has several stretches of basic amino acids, such as RRRRRHKRKTGHREKKEKERISENY, which are characteristic of Nuclear Localization Signals (NLS). This motif suggests that the protein might localize to the nucleus.

3. Mitochondrial Targeting Sequences (MTS): There is no strong evidence of mitochondrial targeting sequences, which typically contain alternating positively charged and hydrophobic amino acids.

4. Endoplasmic Reticulum (ER) and Golgi Apparatus: The presence of signal peptides and transmembrane domains may also indicate localization to the ER and Golgi apparatus for proteins that are involved in secretion or reside in the membranes of these organelles.

5. Post-translational Modifications: The sequence has regions rich in serine (S) and threonine (T), suggesting potential sites for phosphorylation, which could further influence its localization, possibly to signaling pathways associated with the plasma membrane or nuclear functions.

Prediction Summary: Based on the sequence motifs and composition: Nucleus: The presence of multiple NLS motifs suggests nuclear localization. Membrane: The hydrophobic regions and potential transmembrane domains suggest that the protein could also be associated with the plasma membrane or other intracellular membranes like the ER or Golgi apparatus. Given these characteristics, the protein may have dual localization in both the nucleus and membrane-bound compartments (plasma membrane or internal membranes).



To forecast the intracellular localization of the protein, I'll analyze its amino acid composition and sequence motifs.

1. Amino Acid Composition: The protein has a high percentage of hydrophobic amino acids (e.g., Leu, Ile, Val, Phe) and a moderate percentage of charged amino acids (e.g., Asp, Glu, Lys, Arg). This suggests that the protein may be localized to the plasma membrane or an intracellular organelle with a hydrophobic environment.

2. Sequence Motifs: The protein contains several sequence motifs that are known to be involved in protein localization:

a. The MEDEAVLDRGASFLKHVCDEEEVEGHHTIYIGVHVPKSYRRRRHKRKTGHREKKEKERISENYSDKSDVENADESSSSILKPLISPAAERIRFILGEEDDSPAPPQLFTELDELLAVDQGEMEWKETARWIKFEEKVEQGGERWSKPHVATLSLHSLFELRTCMEKGSIMLDREASSLPQLVEMIVDHQIETGLLKPDLDKDKVITYTLRKHRRHQTKKSNLRSLADIGKTVSSASRMFTNPDNGSPAMTHRNLTSSSLNDISDKPEKDQLKNKFMKKLPRDAEASNVLVGEVDFLDSPFI AFVRLQQA VMLGALTEVPVPT RFLFILLGPKGKA KSYHEIGRAIATLMSDEVFHDIA YKAKDRQDLIAGIDEFLDEVIVLPPGEWDPAIRIEPPKSLPSSDKRKNMYSGGENVQMNGDTPPDGGHGGGGGHADCEELQRTGRFCGGLIKDIKRKAPFFASDFYDALNIQALSAILFIYLATVTN AITFGGLLGDATDNMQGVLESFLGTAVSGAIFCLFAGQPLTILSSTGPVLVFERLLFNFSKDHNF DYLEFRLWIGLWSAFLCLILVATDASFLVQYFTRFTEEGFSSLSIFIFYDAFKKMIKLADYYPINSNFKVG YNTQFSCVCMPPDPVNISVSNDTTLAPEDLPTISSSNMYHNATFDWAFLT TKECLKYGGKLVGN NCGFVPDITLMSFILFLGTYTSSMALKKFKTSPYFPTTARKLISDFAILPILIFCVIDALVGVDTPKLIVPSEFKPTSPNRGWFVAPFGGNPWVWYLA AAPALLVTILIFMDQQITAVIVNRKEHKLKKGAGYHLDLFWVAILMVVCSFMALPWYVAATVISIAHIDSLKMETETSAPGEQPKFLGVREQ RVTGTLVFILTGLSVFMAPILKFIMPVLYGVFLYMGVASLNGVQFMDRLKLLMPLKHQP DFIYLRHVPLRRVHLFTFLQVLC LALLWILKSTVAAIIFPV MILALVAVRKGMDYLF SQHDL SFLDDVIPEKD KKKKKEDEK KKKKKKGSVDS DNDSDC PYSEK VPSIKIPMDIMEQQPFLSDSKPSDRERSPTFLERHTSC) sequence contains several stretches of hydrophobic amino acids, which could facilitate membrane association.

b. The protein contains a potential transmembrane domain (TMD) around residues 150-170 (MEDEAVLDRGASFLKHVCDEEEVEGHHTIYIGVHVPKSYRRRRHKRKTGHREKKEKERISENYSDKSDVENADESSSSILKPLISPAAERIRFILGEEDDSPAPPQLFTELDELLAVDQGEMEWKETARWIKFEEKVEQGGERWSKPHVATLSLHSLFELRTCMEKGSIMLDREASSLPQLVEMIVDHQIETGLLKPDLDKDKVITYTLRKHRRHQTKKSNLRSLADIGKTVSSASRMFTNPDNGSPAMTHRNLTSSSLNDISDKPEKDQLKNKFMKKLPRDAEASNVLVGEVDFLDSPFI AFVRLQQA VMLGALTEVPVPT RFLFILLGPKGKA KSYHEIGRAIATLMSDEVFHDIA YKAKDRQDLIAGIDEFLDEVIVLPPGEWDPAIRIEPPKSLPSSDKRKNMYSGGENVQMNGDTPPDGGHGGGGGHADCEELQRTGRFCGGLIKDIKRKAPFFASDFYDALNIQALSAILFIYLATVTN AITFGGLLGDATDNMQGVLESFLGTAVSGAIFCLFAGQPLTILSSTGPVLVFERLLFNFSKDHNF DYLEFRLWIGLWSAFLCLILVATDASFLVQYFTRFTEEGFSSLSIFIFYDAFKKMIKLADYYPINSNFKVG YNTQFSCVCMPPDPVNISVSNDTTLAPEDLPTISSSNMYHNATFDWAFLT TKECLKYGGKLVGN NCGFVPDITLMSFILFLGTYTSSMALKKFKTSPYFPTTARKLISDFAILPILIFCVIDALVGVDTPKLIVPSEFKPTSPNRGWFVAPFGGNPWVWYLA AAPALLVTILIFMDQQITAVIVNRKEHKLKKGAGYHLDLFWVAILMVVCSFMALPWYVAATVISIAHIDSLKMETETSAPGEQPKFLGVREQ RVTGTLVFILTGLSVFMAPILKFIMPVLYGVFLYMGVASLNGVQFMDRLKLLMPLKHQP DFIYLRHVPLRRVHLFTFLQVLC LALLWILKSTVAAIIFPV MILALVAVRKGMDYLF SQHDL SFLDDVIPEKD KKKKKEDEK KKKKKKGSVDS DNDSDC PYSEK VPSIKIPMDIMEQQPFLSDSKPSDRERSPTFLERHTSC), which is a characteristic feature of transmembrane proteins.

c. The protein contains several potential phosphorylation sites, including several serine and threonine residues, which could be involved in regulating protein localization or activity. Based on these features, I predict that the protein is likely



To forecast the intracellular localization of this protein in eukaryotic cells, I'll analyze its amino acid composition and sequence motifs. Here's a step-by-step analysis:

1. Sequence length: The protein is quite long, with over 1000 amino acids, suggesting it's a complex, multi-domain protein.

2. Charged residues: There are many charged residues (K, R, D, E) throughout the sequence, which is common in water-soluble proteins.

3. Hydrophobic regions: There are several stretches of hydrophobic amino acids, particularly in the middle and latter parts of the sequence, suggesting possible transmembrane domains.

4. N-terminal analysis: - No clear signal peptide at the very beginning - Presence of many charged residues (D, E, R, K) in the first 50 amino acids

5. Nuclear localization signals (NLS): - Strong NLS-like motif: RRRRRHKRKT near the N-terminus - Several lysine-rich regions throughout the sequence

6. Transmembrane domain predictions: - Multiple hydrophobic stretches that could form transmembrane helices

7. Specific motifs: - "KKKKK" repeat near the C-terminus, which could be a nuclear retention signal - "DSDNDDSD" sequence near the C-terminus, potentially a phosphorylation site or protein interaction motif

8. Overall composition: - Rich in glycine (G) and proline (P) in some regions, suggesting flexible linkers - Presence of cysteine (C) residues that could form disulfide bonds

Based on this analysis, I predict that this protein could have multiple localizations:

1. Primary localization: Nuclear Reasoning: Strong NLS-like motif near the N-terminus, lysine-rich regions, and the "KKKKK" repeat near the C-terminus are strong indicators of nuclear localization.

2. Secondary localization: Membrane-associated (possibly nuclear membrane or endoplasmic reticulum) Reasoning: The presence of multiple hydrophobic regions suggests transmembrane domains, which could anchor the protein in a membrane.

3. Potential for nucleocytoplasmic shuttling Reasoning: The combination of nuclear localization signals and membrane-association features suggests this protein might shuttle between the nucleus and the cytoplasm, possibly in response to cellular signals. This protein appears to be a complex, multi-domain protein that may play a role in nuclear functions, possibly including transcription regulation or nuclear envelope interactions. Its potential membrane association suggests it might be involved in signaling pathways or transport processes between the nucleus and cytoplasm. To confirm this prediction, experimental verification would be necessary, such as fluorescent tagging and microscopy, or subcellular fractionation followed by Western blotting.

