




Task 4: Function Keywords Prediction

Instruction	What are the UniProtKB keywords for this specific protein sequence?
Input	MRGSFFSRLPPQLSLLLLLLLLLSWRRVWTQEHIGTDPSPVAPVCPEACSCSPGGKANCSALALPAVPAGLSWQVRSLLLDNRNRVSTLPPGAFADAGALLYLVLRENRLRSV HARAFWGLGVLQRLDLSSNQLETLSPGTFTPLRALSFLSLAGNRLALLEPSILGPLLLRVLSLQDNSLSALEAGLLNSLPALDVLRLHGNPWACSCALRPLCTWLRKHPRPTSET ETLLCVSPKLQTLNLLTDFPDNAFKQCTQSLAARDLAVVYALGPASFLASLAICLALGSVLTACGARRRRRRRTTVRHLIRRQPDPEGPASLEDVGSPTTTAIQA
Ground Truth	Cell membrane; Cytoplasm; Cytoskeleton; Disulfide bond; Ion channel; Ion transport; Leucine-rich repeat; Membrane; Reference proteome; Repeat; Signal; Transmembrane; Transmembrane helix; Transport
OPI-tuned model	Cell membrane; Cytoplasm; Cytoskeleton; Disulfide bond; Ion channel; Ion transport; Leucine-rich repeat; Membrane; Reference proteome; Repeat; Signal; Transmembrane; Transmembrane helix; Transport
 GPT-4o	<p>To determine the UniProtKB keywords for this specific protein sequence, you would typically need to identify the protein based on its sequence. The UniProtKB keywords are standard terms used in the UniProt database to describe specific features, functions, and characteristics of a protein.</p> <p>However, without direct access to the UniProt database or specific annotation tools, it isn't possible to definitively identify the UniProtKB keywords for this sequence directly from the sequence alone.</p> <p>If you have access to a sequence alignment tool or a database search tool (like BLAST), you could input the sequence to find the closest matching protein in the UniProt database. Once you have identified the corresponding protein entry, you could then check the UniProtKB keywords associated with that entry.</p> <p>Would you like guidance on how to use these tools to find this information?</p>
 Llama-3.1-8B-Instruct	<p>I can help you with that! The UniProtKB keywords for this specific protein sequence are: * Cell membrane * Cytoplasm * Lipid anchor * Lipid-binding * Lipid metabolism * Membrane protein * Phospholipid * Protein binding * Protein modification * Protein processing * Signal peptide * Transmembrane region Please note that UniProtKB keywords are not always exhaustive and may not cover all possible functions or features of a protein.</p>
 Claude 3.5 Sonnet	<p>To provide accurate UniProtKB keywords for this specific protein sequence, I would need to search a protein database, which I don't have direct access to. However, I can analyze the sequence and provide some general observations that might hint at potential keywords:</p> <ol style="list-style-type: none"><li>The sequence starts with a methionine (M), which is typical for the start of a protein.</li><li>There's a high concentration of leucine (L) near the N-terminus, which could suggest a signal peptide or transmembrane region.</li><li>The sequence contains several repeating patterns, particularly of leucine (L) and serine (S), which might indicate leucine-rich repeat (LRR) domains.</li><li>There are multiple cysteines (C) present, which could be involved in disulfide bond formation.</li><li>The C-terminal end has a stretch of hydrophobic residues, which might suggest a transmembrane region.</li><li>The sequence ends with AIQA, which doesn't appear to be a common motif but could be functionally significant.</li></ol> <p>Based on these observations, some potential UniProtKB keywords might include:</p> <ul style="list-style-type: none"><li>- Signal peptide</li><li>- Transmembrane</li><li>- Leucine-rich repeat</li><li>- Disulfide bond</li><li>- Membrane protein</li></ul> <p>However, to get the actual UniProtKB keywords, you would need to search this sequence in the UniProtKB database or use tools like BLAST to find similar proteins with known annotations. Would you like me to explain any of these observations in more detail?</p>