

## Qualitative examples of Protein2Text-QA question answering task

### User Query

Amino acid sequence to analyze:<seq> MCTKTKEECEKHYMKHFINNPLFASTLLNLKQAEAAKTADTAIPFHSTDDPPRPTFDSLRSRDMAGYMPARADFI  
EEFDNYAEWDLRDIDFVEDSDILHALKMAVVDIYHSRLKERQRRKKIIRDHGLINLRKFQLMERRYKPEVQDLYETMRRFARIVGPPVEHD  
KFIESHALEFELRREIKRLQEYRTAGITNFCASARTYDHLKKTREEERLKRITMLSEVLQYIQDSSACQQWLRQADIDSGLSPSPIMASNSGR  
SAPPLNLTGLPGTEKLNEKEKELCQMVRLVPGAYLEYKSALLNECNKQGGLRLAQARALIKIDVKNKTRKIYDFLIREGYITKG </seq>  
What is unique about the structure of the transcriptional adapter?

### Model Answers

#### Zero-shot

The transcriptional adapter protein sequence contains multiple lysine (K) and arginine (R) residues, indicating potential nuclear localization and involvement in DNA binding. The presence of acidic residues (E, D) suggests roles in protein-protein interactions, crucial for transcriptional regulation.

ROUGE-L: 11.76  
Human Rating: 2

#### Context Engineering

The transcriptional adapter has a unique structure characterized by **multiple low-complexity and disordered regions**, which likely facilitate **dynamic interactions** with other transcription factors and regulatory proteins, enhancing its role in transcriptional regulation.

ROUGE-L: 22.22  
Human Rating: 4

### Ground Truth

The transcriptional adapter is **intrinsically disordered, meaning its structure is not fixed**.

## Qualitative examples of Mol-Instructions question answering task

### User Query

Amino acid sequence to analyze: <seq> MTPTAPITQDMLTITRKLDPGPTNLVGLTRDGLRDVLIAAGTPEKQAKMRVGGQIWQWIYQKGVDRDFDAMTNLS  
KSYRAELAEKFVIEVPEVVSQRQVSADGTRKYLVRVAGGHEVETVYIPEESRGTLCTISSQVGCTLTCSFCHTGTQKLVNLTAGEIIGQVMIARD  
DLGEWPEPGRAPKDETRLLSNIVLMGMGEPLYNFEGVRDAMKIAMDGEGISLSRRRITLSTSGVVPEIARTAEIIGCLLAVSFHGTDTDEIRDK  
LVPINKKWNIAALLNALREYPRLSNSERITFEYVMLKGVNDSDEDARRLVKLIKIPAKINLIPFNEWPGAPYERSDWDRIESFANIIYKAGYA  
SPIRTPRGEDIMAACGQLKSETERARKSRKVIAAEAGLT </seq>  
Please examine the protein encoded by the amino acid sequence and describe its functional role, potential involvement in cellular processes, and its subcellular location.

### Model Answers

#### Zero-shot

The protein encoded by the provided amino acid sequence appears to be **a member of the GTPase family, characterized by motifs involved in binding and hydrolyzing GTP**. These proteins often function as molecular switches in various cellular processes, including **signal transduction, cell division, and vesicle trafficking**. The presence of sequences such as "GTP" and "Ras-like" motifs suggest a role in regulating pathways like the Ras-MAPK signaling cascade, which is crucial for cell proliferation and differentiation.

ROUGE-L: 19.20  
Human Rating: 1

#### Context Engineering

The protein encoded by the given amino acid sequence functions as **a dual-specificity RNA methyltransferase**, similar to RlmN\_AZOC5. It **catalyzes methylation of adenosine at specific positions in 23S rRNA and tRNA**, aiding in ribosomal fidelity and tRNA stability. It relies on a **[4Fe-4S] cluster** and S-adenosyl-L-methionine (SAM) for its activity. The protein **is localized in the cytoplasm**.

ROUGE-L: 28.83  
Human Rating: 4

### Ground Truth

The protein characterized by the amino acid sequence demonstrates **4 iron, 4 sulfur cluster binding, metal ion binding, rRNA (adenine-C2-)-methyltransferase activity, rRNA binding, tRNA (adenine-C2-)-methyltransferase activity, tRNA binding and is implicated in the rRNA base methylation**. Its subcellular localization is primarily **within the cytoplasm**.