

Examples of descriptive text protein-QA

Protein id: Q2V4L8

Protein

<seq>MASPFFFFVLLSALSLENTYASPNYREALSKSLLFFQGQRSGRLPSDQQLSWRSSSSGLSDGSSAHVDLTGGYYDAGDNVKFNFPMAFTTT
MLSWSSELYGKKMGPELQNSRVAIRWATDYLLKCARATPGKLYVGVGDPNGDHKCWERPEDMDTPRTVYSVSPSNPGSDVAAETAALAAS
SMVFRKVDPKYSRLLLATAKKVMQFAIQYRGAYSNSLSSSVCPFYCSYSGYKDELLWGAAWLHRATNDPYYTNFIKSLGGGDQPDIFSWDN
KYAGAYVLLSRRAVLNKDNNFELYKQAAENFMCKILPNSPSSSTKYTKGGLMYKLPQSNLQYVTSITFLLTTYAKYMKSTKQTFNCGNSLIV
PNALINLSKRQVDYVLGVNPMKMSYMGVFSSNFPKRIHHRGSSLPSRAVRSNSLGCNGGFQSFRTQNPNNPILTGAIVGGPNQNDEYPDQRD
DYTRSEPATYINAAFVGPLAYFAASRSP</seq>

Question

In one coherent paragraph, describe the protein represented by the given amino-acid sequence, covering: (1) its main function and any bound cofactors/ligands, (2) subcellular localization and the biological process involved, (3) key domains or motifs, and (4) the specific reaction catalyzed if it is an enzyme.

Answer

A brief overview of the protein with the provided amino acid sequence (Q2V4L8, GUN3_ARATH) is as follows: This enzyme, an endoglucanase 3 (EC 3.2.1.4) from *Arabidopsis thaliana*, catalyzes the endohydrolysis of (1→4)-β-D-glucosidic linkages in cellulose, lichenin, and cereal β-D-glucans, facilitating polysaccharide degradation without requiring bound cofactors. It is secreted to the extracellular region, particularly localized to the plant-type cell wall and Golgi apparatus, where it participates in cell wall organization and cellulose catabolic processes. Key features include its classification within the glycosyl hydrolase 9 family, a catalytic nucleophile at position 77, additional active sites at residues 402, 453, and 462, an N-linked glycosylation site at Asn-370, and an N-terminal signal peptide (residues 1–21) for secretion. Functionally, it mediates root cap cell sloughing through cell wall breakdown, is specifically expressed in root cap cells, and is down-regulated by auxin and abscisic acid. The mature chain spans residues 22–484 and may undergo alternative splicing.

Protein id: Q39056

Protein

<seq>MISTLRRAVFLRRFPVVSPIKRAFSSRIDDEFDPQIMNINELNQEMQSIFGQEPSPDGPGTMDFSELKSSKIEPLRSKNIDFRQQIEYHKS
THSSKNDSQLAIEQYAKVASDMSKLTHTVGIAGEAQMVVDVSSKDNSKRTALACCKVILGKRVFDDLVLANQMGKGDVLGVAKIAGINGAKQTSS
LIPLCHNIALTHVRVDLRLNPEDFSVDIEGEASCTGKTGVEMEAMTAVSVAGLTVYDMCKAASKDISITDVRLERKTGGKSGSWSL</seq>

Question

Determine whether this statement about the given protein is true or false: although this protein's SUMOylation at specific lysine residues supports its role in mRNA processing complexes, and mutations causing an autosomal recessive neurodevelopmental disorder implicate its functional importance, structural studies reveal that these covalent modifications occur exclusively at lysine-359 within its third RNA recognition motif domain

Answer

A concise description of the protein with the specified amino acid sequence (Q39056, CNX3_ARATH) is as follows:\n\nThis mitochondrial protein, localized specifically within the mitochondrial matrix via an N-terminal transit peptide (residues 1-32), functions as a cyclic pyranopterin monophosphate synthase (EC 4.6.1.17). It is essential for molybdenum cofactor biosynthesis, catalyzing the specific lyase reaction that converts (8S)-3',8-cyclo-7,8-dihydroguanosine 5'-triphosphate into cyclic pyranopterin monophosphate (cPMP) and diphosphate (RHEA:49580). Key residues form an active site (including residue 240) and substrate binding sites (residues 187-189 and 225-226). Belonging to the MoaC protein family, it assembles as a homohexamer and is predominantly expressed in the roots of *Arabidopsis thaliana*. Its primary biological role is in the \"Mo-molybdopterin cofactor biosynthetic process\" (GO:0006777).

Examples of knowledge-based protein-QA

Protein id: O35854

Protein

<seq>MSAAILGQVWTRKLLPIPWRLCVPGRCVSSNFKAADLQVQVTREPQKKPAPSQPLLFGKTFTDHMLMV E WNSKTGWGPPRIQPPQNLT
LHPACSGLHYSLQLFEGLKAYKGRDKQVRLFRPWLNM DRMLRSARRLC LPDFDKQELLE CIRQLIEVDKDWVPDGN GTSLYVRPVLIGNEPS
LGVG MVTQALLFVILCPVGSYFPGDSMT PVSLLADPSFVRAWIGGVGDCKLGGNYGPTVAVQQEAQKKGCEQVLWLYGPDHQLTEVGTMNI
FVYWTHEDGELELATPPLDGIILPGVVRQSLLDLARTWGEFRVAERKVTM KELKRALEEGRVREVFV GSGTACQVCPVHQILYEGKQLHIPTM
ENGPELILRFQKELKAIQYGTSAHDWMLRV</seq>

Question

You are a molecular biology professor writing a protein description for undergraduates. Based on the characteristics of the given protein amino acid sequence, briefly answer the following questions: 1. What is the primary catalytic function of this protein in branched-chain amino acid metabolism?n2. Where is this protein localized within the cell?n3. What cofactor is essential for its enzymatic activity?n4. How does its substrate affinity (KM) vary among the branched-chain amino acids?n5. What is the quaternary structure of this protein?n6. Which post-translational modifications are known to occur on this protein?n7. What is the physiological direction of its catalytic reactions?n8. In which tissues is this protein expressed?n9. What family of enzymes does this protein belong to?

Answer

1. It catalyzes the transamination of leucine, isoleucine, and valine with 2-oxoglutarate to form their respective α-keto acids and glutamate.n2. It is localized in the mitochondrion.n3. Pyridoxal 5'-phosphate is the essential cofactor.n4. The KM values are 1 mM for L-leucine and L-isoleucine, and 5 mM for L-valine.n5. It forms a homodimer.n6. It undergoes N6-(pyridoxal phosphate)lysine modification at position 230 and N6-acetyllysine at position 322.n7. The reactions proceed physiologically from left to right (branched-chain amino acids to α-keto acids).n8. It is expressed in all tissues.n9. It belongs to the class-IV pyridoxal-phosphate-dependent aminotransferase family.

Protein id: Q9LXN8

Protein

<seq>MAFSMLFTGHAECVKFVKKFNLPLLVTGGGGYTKENVARCWTVETGILLDTelpNEISENDYIKYFAPDFSLKIPGGHIENLNTKSYISSI
KVQILENLRYIQHAPSVQM QEVPPDFYIPDFDEDEQNP DVRVDQRSRDKQIQRDDEYFDGDNDNDAS</seq>

Question

You are a molecular biology professor writing a protein description for undergraduates. Based on the characteristics of the given protein amino acid sequence, briefly answer the following questions: 1. What is the enzymatic reaction catalyzed by this protein?n2. Where is this protein localized within the cell?n3. What is the role of this protein in chromatin organization?n4. Which histone proteins are substrates for its deacetylation activity?n5. How does its activity contribute to transcriptional regulation?n6. What are the functional implications of its disordered region (residues 120-158)?n7. Does it form part of any multiprotein complexes?n8. What is the significance of its basic and acidic residue compositional biases?n9. How does it compare to other members of the histone deacetylase family?

Answer

1. It catalyzes the hydrolysis of N(6)-acetyl-L-lysyl-[histone] to produce L-lysyl-[histone] and acetate.n2. It is localized in the nucleus.n3. It is involved in chromatin organization through histone deacetylation, leading to epigenetic repression.n4. It deacetylates lysine residues on the core histones H2A, H2B, H3, and H4.n5. Its deacetylation activity represses transcription by modifying chromatin structure.n6. The disordered region may contribute to protein flexibility or interaction with other molecules.n7. Yes, it acts via the formation of large multiprotein complexes.n8. These biases may influence protein-protein interactions or substrate binding.n9. It belongs to the histone deacetylase family and shares similar functional domains.