AT1G01040.1

Full length CDS

ATGGTAATGGAGGATGAGCCTAGAGAAGCCACAATAAAGCCTTCTTATTGGCTAGATGCTTGCGAGGACATCTCTTGTGATCTTATCGATGATCTCGTGTCTGAATTTGATCCTTCCTCTGTTGCTGTCAATGAATCCACTGATGAAAACGGCGTCATCAATGATTTTTTCGGTGGGATTGATCACATTTTAGATAGTATCAAGAACGGTGGAGGCTTACCAAACAATGGCGTTTCTGATACCAATTCTCAAATCAACGAGGTTACTGTAACTCCTCAGGTTATTGCTAAGGAGACAGTGAAGGAGAATGGGTTGCAAAAGAATGGCGGTAAGAGAGACGAATTCTCGAAAGAGGAAGGAGACAAGGATAGGAAGAGAGCTAGGGTTTGTAGTTATCAGAGTGAAAGGAGTAACCTTTCAGGTAGAGGGCATGTTAATAATTCTAGGGAGGGAGATAGGTTTATGAATAGGAAACGTACTCGTAATTGGGACGAGGCGGGTAACAATAAGAAGAAAAGGGAATGTAACAATTACAGAAGAGATGGTAGAGATAGAGAAGTTAGGGGTTATTGGGAGAGGGATAAAGTTGGTTCCAATGAGTTGGTTTATAGGTCAGGGACTTGGGAAGCTGATCATGAAAGAGATGTTAAGAAAGTGAGTGGTGGAAACCGCGAATGCGATGTCAAGGCAGAGGAGAACAAGAGTAAGCCTGAAGAACGTAAAGAGAAGGTTGTGGAAGAGCAAGCAAGGCGATACCAGTTGGATGTTCTTGAACAAGCTAAAGCGAAAAACACGATTGCTTTCCTTGAGACCGGTGCTGGAAAGACACTTATCGCGATTCTTCTTATTAAAAGTGTTCATAAGGATCTGATGAGCCAGAACAGAAAAATGCTCTCGGTGTTCTTGGTTCCCAAAGTGCCTTTGGTTTATCAGCAAGCAGAAGTGATCCGTAATCAAACTTGTTTTCAAGTTGGACATTATTGTGGTGAGATGGGACAGGACTTTTGGGATTCTCGAAGGTGGCAACGAGAGTTTGAGTCTAAGCAGGTTCTAGTTATGACAGCACAAATTCTGTTGAATATACTGAGACACAGTATCATTAGAATGGAAACAATTGATCTTCTTATTCTCGACGAGTGTCACCACGCTGTCAAGAAACATCCATACTCTTTAGTGATGTCAGAGTTTTACCATACAACTCCTAAAGATAAAAGACCTGCCATCTTTGGAATGACTGCTTCGCCTGTTAATTTAAAGGGTGTTTCAAGCCAAGTAGATTGTGCGATAAAGATACGTAACCTCGAGACCAAGTTGGATTCTACGGTTTGTACTATAAAAGATCGAAAAGAATTAGAGAAACATGTGCCTATGCCTTCAGAGATAGTCGTCGAGTATGACAAAGCTGCTACTATGTGGTCTCTTCATGAGACAATAAAGCAAATGATTGCAGCTGTTGAAGAAGCGGCACAAGCAAGTTCAAGGAAAAGCAAGTGGCAATTTATGGGGGCTAGGGATGCTGGAGCAAAGGATGAATTGAGACAGGTTTATGGCGTCTCTGAAAGAACGGAGAGCGATGGTGCTGCCAATTTGATTCATAAACTTAGAGCTATCAATTATACTCTTGCTGAATTGGGTCAATGGTGTGCTTACAAGGTGGGACAATCATTCTTGTCTGCTTTGCAAAGTGATGAGAGGGTGAATTTCCAAGTCGACGTGAAGTTTCAAGAATCATACCTCAGTGAGGTGGTGTCACTCTTGCAATGTGAGCTTCTGGAAGGCGCTGCTGCTGAAAAAGTCGCGGCGGAAGTTGGCAAACCAGAAAATGGTAATGCACATGACGAGATGGAGGAGGGAGAGCTCCCTGATGATCCTGTGGTCTCGGGAGGGGAGCACGTTGATGAAGTAATAGGCGCCGCAGTGGCTGATGGGAAAGTTACTCCAAAAGTACAATCATTGATCAAACTACTCCTCAAATATCAGCACACAGCTGATTTTCGAGCTATTGTTTTCGTTGAGAGGGTGGTTGCTGCTTTGGTTCTTCCTAAGGTTTTTGCGGAGCTGCCTTCGCTTAGTTTTATACGGTGTGCCAGCATGATTGGACACAATAACAGCCAGGAGATGAAATCATCTCAAATGCAGGATACAATTTCCAAATTCCGAGATGGGCATGTGACACTGTTAGTTGCCACAAGCGTTGCTGAGGAAGGACTTGATATTAGGCAATGTAACGTTGTTATGCGTTTCGACCTTGCAAAGACGGTGCTGGCATACATTCAGTCTCGTGGCCGGGCAAGAAAGCCTGGATCAGACTACATACTCATGGTTGAGAGAGGAAATGTATCTCACGCAGCGTTCCTAAGGAATGCTAGGAACAGTGAGGAGACACTTCGAAAAGAAGCAATAGAAAGGACTGATCTTAGTCATCTCAAAGATACATCGAGATTAATCTCAATTGATGCTGTGCCTGGTACAGTTTATAAGGTGGAGGCAACTGGTGCCATGGTTAGCTTGAATTCCGCGGTTGGTCTTGTACATTTCTACTGCTCTCAGCTTCCTGGTGACAGGTATGCAATCCTTCGTCCTGAGTTTAGCATGGAGAAGCATGAAAAGCCTGGGGGCCACACGGAATATTCATGTAGGCTTCAGCTTCCTTGCAATGCACCGTTTGAAATACTTGAGGGTCCTGTTTGCAGTTCAATGCGTCTTGCACAACAGGCTGTATGTTTAGCTGCTTGCAAGAAACTGCATGAGATGGGTGCATTTACCGATATGCTATTACCGGACAAAGGAAGTGGTCAAGACGCTGAGAAGGCTGACCAAGATGATGAAGGTGAGCCTGTTCCTGGAACTGCTAGACATAGAGAGTTCTATCCTGAAGGTGTGGCGGATGTACTTAAGGGAGAATGGGTTTCATCTGGAAAGGAAGTTTGTGAGAGCTCAAAGCTATTCCATTTATACATGTATAATGTCAGATGTGTAGATTTTGGCTCTTCAAAAGATCCATTCCTAAGCGAAGTTTCAGAGTTCGCGATTCTTTTTGGCAATGAGCTGGATGCAGAGGTTGTATTATCGATGTCTATGGATCTTTATGTTGCTCGGGCCATGATCACTAAAGCATCTCTTGCTTTCAAGGGATCACTTGATATTACAGAAAACCAGCTATCATCTCTAAAAAAGTTTCATGTGAGGTTAATGAGTATCGTGTTGGATGTTGATGTTGAACCCTCCACGACACCATGGGATCCTGCAAAGGCCTACCTGTTTGTCCCTGTTACTGACAATACGTCTATGGAACCCATAAAAGGGATCAACTGGGAATTGGTTGAAAAGATTACGAAAACCACAGCGTGGGACAACCCTCTTCAGAGAGCTCGTCCCGATGTATATCTCGGGACTAATGAGAGAACTCTTGGTGGGGACAGAAGGGAATATGGGTTTGGTAAACTTCGTCACAACATTGTATTTGGGCAGAAATCTCACCCAACTTATGGTATTAGAGGAGCTGTTGCATCCTTCGATGTTGTGAGAGCTTCTGGATTGTTACCTGTGAGAGATGCTTTTGAGAAGGAAGTAGAAGAGGATTTATCAAAAGGAAAATTGATGATGGCTGATGGGTGCATGGTTGCAGAAGATCTTATTGGGAAAATAGTGACAGCCGCACATTCCGGGAAGCGGTTTTACGTAGATTCAATTTGTTATGACATGAGTGCAGAAACATCTTTCCCTAGGAAAGAGGGATATCTTGGTCCCCTAGAGTACAACACGTACGCTGACTATTACAAGCAAAAGTATGGAGTTGATTTGAACTGTAAGCAACAACCTTTGATTAAAGGACGTGGTGTTTCGTATTGCAAGAACCTTCTTTCTCCTCGGTTTGAACAGTCAGGTGAATCTGAGACAGTCCTTGATAAGACATATTACGTGTTTCTTCCACCTGAACTATGCGTTGTGCATCCGCTTTCGGGTTCACTTATCCGAGGTGCTCAGAGGTTACCCTCTATAATGAGAAGAGTTGAGAGCATGTTACTCGCTGTTCAACTCAAAAATTTGATTAGTTATCCTATTCCCACATCAAAGATTCTTGAAGCCTTGACTGCCGCCTCGTGCCAGGAAACGTTCTGCTACGAGAGAGCTGAGCTTTTAGGAGATGCGTATCTAAAATGGGTTGTTAGTCGTTTTCTGTTTCTCAAGTATCCTCAAAAGCACGAGGGTCAGCTTACAAGGATGAGGCAACAAATGGTTAGTAATATGGTTCTTTATCAGTTTGCTCTGGTTAAAGGGCTTCAGTCATATATCCAGGCGGATCGATTCGCCCCGTCTAGGTGGTCTGCTCCTGGTGTGCCTCCGGTTTTCGACGAGGACACAAAAGATGGAGGATCTTCGTTTTTCGATGAAGAGCAAAAACCTGTTTCCGAGGAAAACAGCGATGTGTTTGAAGATGGGGAGATGGAGGATGGTGAACTAGAGGGTGATTTGAGTTCGTACCGAGTTTTATCTAGCAAAACGTTAGCTGATGTTGTTGAGGCTTTGATTGGTGTTTATTACGTCGAAGGGGGTAAGATTGCAGCTAATCATTTGATGAAATGGATTGGGATTCACGTGGAGGATGATCCTGATGAAGTCGATGGAACATTGAAAAATGTTAATGTTCCAGAGAGTGTGCTCAAGAGCATCGACTTTGTTGGTCTTGAGAGAGCTCTTAAATATGAGTTTAAAGAGAAAGGTCTTCTTGTTGAAGCTATAACACATGCTTCAAGACCATCTTCAGGTGTTTCGTGTTACCAGAGATTGGAATTTGTTGGTGACGCGGTCTTGGATCATCTCATCACAAGACATCTATTTTTCACATACACAAGCCTTCCTCCTGGTCGGTTAACAGATCTTCGAGCTGCAGCGGTTAACAACGAGAATTTTGCTCGCGTTGCGGTTAAACATAAACTCCACTTGTACCTTCGTCACGGTTCAAGCGCCCTCGAAAAACAGATTCGGGAATTTGTGAAGGAGGTTCAAACCGAGTCATCGAAACCGGGGTTTAACTCTTTTGGTTTGGGAGACTGCAAAGCACCAAAAGTTCTTGGAGACATTGTTGAATCTATTGCAGGTGCTATTTTTCTTGATAGTGGAAAAGATACAACTGCTGCTTGGAAGGTTTTTCAACCTTTGCTTCAGCCCATGGTGACACCAGAGACACTTCCAATGCATCCGGTGCGAGAGCTACAAGAGCGGTGCCAGCAACAAGCAGAAGGGTTAGAATACAAAGCGAGTAGGAGTGGTAACACAGCGACTGTGGAAGTTTTCATCGACGGTGTTCAAGTTGGAGTAGCGCAAAACCCGCAGAAGAAAATGGCTCAAAAGCTAGCTGCGAGGAACGCACTTGCAGCTTTGAAAGAGAAAGAAATAGCAGAATCAAAGGAGAAGCATATCAACAACGGTAATGCGGGAGAGGATCAAGGCGAGAATGAGAATGGGAACAAGAAGAATGGGCATCAGCCGTTTACGAGACAAACGTTGAATGATATTTGTTTGAGGAAGAATTGGCCAATGCCTTCTTACAGATGTGTGAAAGAAGGAGGACCGGCTCATGCAAAGAGATTTACGTTTGGGGTAAGAGTTAATACGAGCGATAGAGGATGGACCGATGAGTGTATTGGCGAGCCAATGCCGAGTGTTAAGAAAGCTAAGGATTCAGCTGCGGTTCTTCTACTTGAGCTTTTAAATAAAACTTTTTCTTGA

Nt position: 2294

2294/3 = 764.67,

Or 765, which is an R

Protein sequence =

MVMEDEPREATIKPSYWLDACEDISCDLIDDLVSEFDPSSVAVNESTDENGVINDFFGGIDHILDSIKNGGGLPNNGVSDTNSQINEVTVTPQVIAKETVKENGLQKNGGKRDEFSKEEGDKDRKRARVCSYQSERSNLSGRGHVNNSREGDRFMNRKRTRNWDEAGNNKKKRECNNYRRDGRDREVRGYWERDKVGSNELVYRSGTWEADHERDVKKVSGGNRECDVKAEENKSKPEERKEKVVEEQARRYQLDVLEQAKAKNTIAFLETGAGKTLIAILLIKSVHKDLMSQNRKMLSVFLVPKVPLVYQQAEVIRNQTCFQVGHYCGEMGQDFWDSRRWQREFESKQVLVMTAQILLNILRHSIIRMETIDLLILDECHHAVKKHPYSLVMSEFYHTTPKDKRPAIFGMTASPVNLKGVSSQVDCAIKIRNLETKLDSTVCTIKDRKELEKHVPMPSEIVVEYDKAATMWSLHETIKQMIAAVEEAAQASSRKSKWQFMGARDAGAKDELRQVYGVSERTESDGAANLIHKLRAINYTLAELGQWCAYKVGQSFLSALQSDERVNFQVDVKFQESYLSEVVSLLQCELLEGAAAEKVAAEVGKPENGNAHDEMEEGELPDDPVVSGGEHVDEVIGAAVADGKVTPKVQSLIKLLLKYQHTADFRAIVFVERVVAALVLPKVFAELPSLSFIRCASMIGHNNSQEMKSSQMQDTISKFRDGHVTLLVATSVAEEGLDIRQCNVVMRFDLAKTVLAYIQSRGRA**R**KPGSDYILMVERGNVSHAAFLRNARNSEETLRKEAIERTDLSHLKDTSRLISIDAVPGTVYKVEATGAMVSLNSAVGLVHFYCSQLPGDRYAILRPEFSMEKHEKPGGHTEYSCRLQLPCNAPFEILEGPVCSSMRLAQQAVCLAACKKLHEMGAFTDMLLPDKGSGQDAEKADQDDEGEPVPGTARHREFYPEGVADVLKGEWVSSGKEVCESSKLFHLYMYNVRCVDFGSSKDPFLSEVSEFAILFGNELDAEVVLSMSMDLYVARAMITKASLAFKGSLDITENQLSSLKKFHVRLMSIVLDVDVEPSTTPWDPAKAYLFVPVTDNTSMEPIKGINWELVEKITKTTAWDNPLQRARPDVYLGTNERTLGGDRREYGFGKLRHNIVFGQKSHPTYGIRGAVASFDVVRASGLLPVRDAFEKEVEEDLSKGKLMMADGCMVAEDLIGKIVTAAHSGKRFYVDSICYDMSAETSFPRKEGYLGPLEYNTYADYYKQKYGVDLNCKQQPLIKGRGVSYCKNLLSPRFEQSGESETVLDKTYYVFLPPELCVVHPLSGSLIRGAQRLPSIMRRVESMLLAVQLKNLISYPIPTSKILEALTAASCQETFCYERAELLGDAYLKWVVSRFLFLKYPQKHEGQLTRMRQQMVSNMVLYQFALVKGLQSYIQADRFAPSRWSAPGVPPVFDEDTKDGGSSFFDEEQKPVSEENSDVFEDGEMEDGELEGDLSSYRVLSSKTLADVVEALIGVYYVEGGKIAANHLMKWIGIHVEDDPDEVDGTLKNVNVPESVLKSIDFVGLERALKYEFKEKGLLVEAITHASRPSSGVSCYQRLEFVGDAVLDHLITRHLFFTYTSLPPGRLTDLRAAAVNNENFARVAVKHKLHLYLRHGSSALEKQIREFVKEVQTESSKPGFNSFGLGDCKAPKVLGDIVESIAGAIFLDSGKDTTAAWKVFQPLLQPMVTPETLPMHPVRELQERCQQQAEGLEYKASRSGNTATVEVFIDGVQVGVAQNPQKKMAQKLAARNALAALKEKEIAESKEKHINNGNAGEDQGENENGNKKNGHQPFTRQTLNDICLRKNWPMPSYRCVKEGGPAHAKRFTFGVRVNTSDRGWTDECIGEPMPSVKKAKDSAAVLLLELLNKTFS

Pfam: (this is a slow search because protein is 1910 AA long)

Start at 12:01 pm; finished within 2 minutes today.

(Hard to look at this structure. . .)

Alphafold

DCL1, Arabidopsis

https://alphafold.ebi.ac.uk/entry/Q9SP32

**residue 765 (So, R765K)**

Is this in a region with a high level of model confidence?

Yes, high confidence (76)

Is this in a region with a predicted secondary structure? If so, what type?

Looks like a loop between a beta sheet and an alpha-helix

Is the wild-type amino acid usually found at this site involved in hydrogen bond interactions with another amino acid? If so, which one(s)?

R765 has H bond to

Ile739, Asp738, Cys742,

<https://biosig.lab.uq.edu.au/dynamut/results_prediction/16969023601>

“A chain with more than 800 residues can take several hours to be analysed.”

Pfam: (residue 765)

<https://www.ebi.ac.uk/interpro/result/InterProScan/#table>

helicase: 250 – 415; **helicase, 648 – 812**; Dicer dimerization, 840 – 935; PAZ domain, 1181 – 1343; RNaseIII, 1358 – 1539 and 1560 – 1731; and two DBRM at the end