

UniprotKB ID	Entry name	organism	full name	oglcnacscore	oglcnac sites	phosphorylation sites	PMIDS	sequence
P42859	HD_MOUSE	Mus musculus	Huntingtin	9.108089	NaN	S396;S398;S411;S620;S623;S1159;S1179;S1853	33300544	MATLEKLMKAFESLKSFQQQQQQQ PPPQAPPPPPPPPPQPPQPPQGP PPPPPLPGPAEEPLHRPKKELSATKK DRVNHCLTICENIVAQSLRNSPEFQ KLLGIAMELFLLCSDAESDVRMVA DECLNKVIKALMDSNLPRLQLELYK EIKKNGAPRSLRAALWRFAELAHLV RPQKCRPYLVNLLPCLTRTSKRPEES VQETLAAAVPKIMASFGNFANDNEI KVLLKAFIANLKSSSPTVRRTAAGSA VSICQHSRRTQYFYNWLLNVLLGLL VPMEEHSTLLILGVLLTLRCLVPLL QQQVKDTSLKGSFGVTRKEMEVSPTS TEQLVQVYELTLHHTQHGDHNVVT GALELLQQLFRTPPELLQALTPPGG LGQLTLVQEEARGRGRSGSIVELLAG GGSSCSPVLSRKQKGVLLGEEAL EDDSESRSDVSSAFAASVKSEIGGE LAASSGVSTPGSVGHDIITEQPRSQH TLQADSVDLSGCDLTAATDGDEED ILSHSSSQFSAVPSDPAMDLDNGTQ ASSPISDSSQTTTEGPD SAVTPSDSS EIVLDGADSQYLGMQIGQPQEDDEE GAAGVLSGEVSDVFRNSSLALQQA LLERMGHSRQPSDSSIDKYVTRDEV AEASDPESKPCRIKGDIGQPNDDDS APLVHCVRLLSASFLLTGEKKALVPD RDVRVSVKALALSCIGAVALHPESF FSRLYKVLNNTTESTEEQYVSDILNYI DHGDPQVRGATAILCGTLVYSILSR RLRVGDWLGNI RTLTGNTFSLVDCIP LLQKTLKDESSVTCKLACTAVRHCV LSLCSSSYSDLGLQLLIDMLPLKNSS YWLVRTELLDTLAEIDFRLVSFLEAK AESLHRGAHHYTGFLKLQERVLNNV VIYLLGDEDPRVRHVAATSLTRLVPK LFYKCDQGQADPVVAVARDQSSVYL KLLMHETQPPSHFSVSTITRIYRGYS LLPSITDVTMENNLSRVVAAVSHEL TSTTRALTFGCCEALCLLSAAFPVCT WSLGWHCGVPPLSASDESRSKCTV GMASMILTLLSSAWFPLDLSAHQDA LILAGNLLAASAPKSLRSSWTSEEEA NSAATRQEEIWPALGDRTLVPVLEQ LFSHLLKVINICAHVLDVTPGPAIK AALPSLTNPPSLSPIRRKGKEKEPGE QASTPMSPKKVGEASAASRQSDTSG PVTASKSSSLGSFYHLP SYLKLHDVL KATHANYKVTLDLQNSTEFKGGFLR

SALDVLSQILELATLQDIGKCVEEVL  
GYLKSCFSREPMMATVCVQQLKTL  
FGTNLASQFDGLSSNPSKSCRAQR  
LGSSSVRPGLYHYCFMAPYTHFTQA  
LADASLRNMVQAEQERDASGWFDV  
LQKVSAQLKTNLTSVTKNRADKNAI  
HNHIRLFEPLVIKALKQYTTTTSVQL  
QKQVLDLLAQLVQLRVNYCLLSDQ  
VFIGFVLKQFEYIEVGQFRESEAIIPN  
IFFFLVLLSYERYHSKQIIGIPKIIQLC  
DGIMASGRKAVTHAIPALQPIVHDLF  
VLRGTNKADAGKELETQKEVVVSMML  
LRLIQYHQVLEMFILVLQQCHKENE  
DKWKRLSRQVADIILPMLAKQQMHI  
DSHEALGVLNTLFEILAPSSLRPVDM  
LLRSMFITPSTMASVSTVQLWISGIL  
AILRVLISQSTEDIVLCRIQELSFSPH  
LLSCPVINRLRGGGGNVTLGECSEG  
KQKSLPEDTFSRFLQLVGILLEDIVT  
KQLKVDMEQHTFYCQELGTLML  
CLIHIFKSGMFRRITAAATRLFTSDG  
CEGSFYTLES LNARVRSMVPTH PAL  
VLLWCQILLINHTDHRWWAEVQQ  
TPKRHSLSCTKSLNPQKS GEEEDSG  
SAAQLGMCNREIVRRGALILFCDYV  
CQNLHDSEHLTWLIVNHIQDLISLS  
HEPPVQDFISAIHRNSAASGLFIQAI  
QSRCENLSTPTTLKKTLCLEGIHLS  
QSGAVLTLYVDRLLGTPFRALARMV  
EELNRIQEHLQNSGLAQRHQRLYSL  
LDRFRLSTVQDSLSPVPPVTSHPDGD  
DGHTSLETVSPDKDWYLQLVRSQC  
WTRSDSALLEGAELVNRIPAEDMND  
FMMSSEFNLSLLAPCLSLGMSEIAN  
GQKSPLFEAARGVILNRVTSVVQQLP  
AVHQVFQPFPLPIEPTAYWNKLNDDL  
GDTTSYQSLTILARALAQYLVVLSKV  
PAHLHLPPEKEGDTVKFVVMVTVEAL  
SWHLIHEQIPLSLDLQAGLDCCCLA  
LQVPGLWGVLSPEYVTHACSLIHC  
VRFILEAIAVQPGDQLLGPE SRSHTP  
RAVRKEEVDSIQNL SHVTSACEMV  
ADMVESLQSVLALGHKRNSTLPSFL  
TAVLKNIVISLARLPLVNSYTRVPPLV  
WKLGWSPKPGDFGTVPFPEIPVEFL  
QEKEILKEFIYRINTLGWTNRTQFEE  
TWATLLGVLVTQPLVMEQEESPPEE  
DTERTQIHVLAVQAITSVLVSAMTVP  
VAGNPAVSCLEQQPRNKPLKALDTR  
FGRKLSMIRGIVEQEIQEMVSRQEN  
TATHSHQAWDPVPSLLPATTGALI

SHDKLLLQINPEREPGNMSYKLGQV  
SIHSVWLGNITPLREEEWDEEEEE  
ESDVPAPTSPVSPVNSRKHRAGVDI  
HSCSQFLELYSRWILPSSAARRTPV  
ILISEVVRSLLVSDLFTERTQFEMM  
YLTTELRRVHPSEDEILIQYLVPATC  
KAAAVLGMDKTVAEPVSRLLLESTLR  
SSHLP SQIGALHGILYVLECDLLDDT  
AKQLIPVSDYLLSNLKGIAHCVNIH  
SQQHVLVMCATAFYLMENYPLDVG  
PEFSASVIQMCGVMLSGSEESTPSII  
YHCALRGLERLLLSEQLSRLDTESLV  
KLSVDRVNVQSPHRAMAALGLMLT  
CMYTGKEKASPGRASDPSPATPDSE  
SVIVAMERVS VLFDRIRKGFPCEARV  
VARILPQFLDDFFPPQDVMNKVIGE  
FLSNQQPYPQFMATVVYKVFQTLHS  
AGQSSMVRDWVMLSLSNFTQRTPV  
AMAMWLSLSCFLVSASTSPWVSAILP  
HVISRMGKLEQVDVNLFLVATDFY  
RHQIEEEFDRRAFQSVFEVVAAPGS  
PYHLLACLQNVHKVTTC