

Molecular Biology

Entry Clone Accession: NM_199420.3

Entry Clone Source: Remko Prevo (ROB, ORCRB), originally from Richard Wood, Uni of Texas

SGC Construct ID: POLQA-c221

Protein Region: M134-M1029

Vector: pFB-LIC-Bse. This is a baculovirus transfer vector (Bac-to-bac), with N-terminal 6 His tag followed by a TEV cleavage site

Host: DH10Bac

Sequence (with tag(s)):

MGHHHHHHSSGVDLGTENLYFQSMNLLRRSGKRRRSESGSDSFSGSGGDSSASPQFLS
GSVLSPPPGLGRCLKAAAAGECKPTVPDYEIDKLLLANWGLPKAVLEKYHSFGVKKMFE
WQAECLLLGQVLEGKNLVYSAPTSAGKTLVAELLILKRVLEMRRKALFILPFVSVAKEKK
YYLQSLFQEVGIKVDGYMGSTSPSRHFSSLDIAVCTIERANGLINRLIEENKMDLLGMVV
VDELHMLGDSHRGYLLELLLTIKICYITRKSASCQADLASSLSNAVQIVGMSATLPNLELVA
SWLNAELYHTDFRPVPLLESVKVGNSIYDSSMKLVREFEPMLQVKGDEDHVVSLEYETI
CDNHSVLLFCPSKKWCEKLADIIAREFYNLHHQAEGLVKPSECPPVILEQKELLEVM DQL
RRLPSGLDSVLQKTPWGVAFHHAGLTFEERDIIEGA FRQGLIRVLAATSTLSSGVNLPAR
RVIIRTPIFGGRPLDILTYKQMVGRAGRKGVDTVGESILICKNSEKSKGIAL LQGS LKP VRS
CLQRREGEEVTGSMIRAILEIIVGGVASTSQDMHTYAACTFLAASMKEGKQGIQRNQESV
QLGAIEACVMWLLNEFIQSTEASDGTEGKVYHPTH LGSATLSSSLSPADTLDIFADLQRA
MKG FVLENDLHILYLVTPMFEDWTTIDWYRFFCLWEKLPTSMKRVAELVGVEEGFLARC
VKGKVVAR TERQHRQMAIHKRFFTS LVLLDLISEVPLREINQKYGCN R GQIQSLQQSAAV
YAGMITVFSNRLGWHNMELLLSQFQKRLTFGIQRELCDLVRVSL LNAQRARVLYASGFHT
VADLARANIVEVEVILKNAV PFKSARKAVDEEEEEAVEERRNMRTIWVTGRKGLTEREAA
ALIVEEARMILQQDLVEM

Sequence after tag cleavage:

SMMNLLRRSGKRRRSESGSDSFSGSGGDSSASPQFLSGSVLSPPPGLGRCLKAAAAGECK
PTVPDYEIDKLLLANWGLPKAVLEKYHSFGVKKMFEWQAECLLLGQVLEGKNLVYSAP
TSAGKTLVAELLILKRVLEMRRKALFILPFVSVAKEKKYYLQSLFQEVGIKVDGYMGSTS
PSRHFSSLDIAVCTIERANGLINRLIEENKMDLLGMVVVDELHMLGDSHRGYLLELLLTIKI
CYITRKSASCQADLASSLSNAVQIVGMSATLPNLELVASWLNAELYHTDFRPVPLLESVK
VGNSIYDSSMKLVREFEPMLQVKGDEDHVVSLEYETICDNHSVLLFCPSKKWCEKLADII
AREFYNLHHQAEGLVKPSECPPVILEQKELLEVM DQLRRLPSGLDSVLQKTPWGVAFH
HAGLTFEERDIIEGA FRQGLIRVLAATSTLSSGVNLPARRVIIRTPIFGGRPLDILTYKQMV
GRAGRKGVDTVGESILICKNSEKSKGIAL LQGS LKP VRSCLQRREGEEVTGSMIRAILEIIV
GVASTSQDMHTYAACTFLAASMKEGKQGIQRNQESVQLGAIEACVMWLLNEFIQSTEAS
SDGTEGKVYHPTH LGSATLSSSLSPADTLDIFADLQRAMKG FVLENDLHILYLVTPMFED
WTTIDWYRFFCLWEKLPTSMKRVAELVGVEEGFLARCVKGKVVAR TERQHRQMAIHKR
FFTS LVLLDLISEVPLREINQKYGCN R GQIQSLQQSAAVYAGMITVFSNRLGWHNMELLLS
QFQKRLTFGIQRELCDLVRVSL LNAQRARVLYASGFHTVADLARANIVEVEVILKNAV PFK
SARKAVDEEEEEAVEERRNMRTIWVTGRKGLTEREAAALIVEEARMILQQDLVEM

DNA Sequence:

ATGGGCCACCATCATCATCATCATTCTTCTGGTGTAGATCTGGGTACCGAGAACCTGTA
CTTCCAATCCATGATGAATCTTCTGCGTCGGAGTGGGAAACGGCGGCGTTCAGAAATCA
GGCTCAGATTCGTTCTCGGGAAGCGGCGGTGACAGCAGTGCCAGCCCCCAGTTCCTC
TCCGGGTCCGTGCTGAGCCCCGCCGCCCGGCCTTGGTCGCTGCCTGAAGGCCGCAGCT
GCAGGAGAATGCAAGCCTACAGTTCCTGACTACGAAATAGACAAGCTACTATTGGCA
AACTGGGGACTTCCTAAAGCAGTTCTGGAAAAATACCACAGTTTTTGGTGTA AAAAAG
ATGTTTGAATGGCAGGCAGAGTGCCTTTTGCTTGGACAAGTCCTGGAAGGAAAGAAT

TTAGTTTATTCAGCTCCTACAAGTGCTGGGAAGACTCTTGTGGCAGAATTACTTATTTT
GAAGCGGGTTTTGGAAATGCGGAAGAAAGCTTTGTTTATTCTTCCCTTTGTTTCTGTG
GCTAAAGAGAAGAAATACTACCTCCAGAGTCTGTTTCAGGAAGTAGGAATAAAAGTA
GACGGTTATATGGGCAGCACCTCTCCATCAAGGCATTTCTCTTCATTGGATATTGCAGT
CTGCACAATTGAGAGAGCCAATGGTCTGATCAATCGCCTCATAGAGGAAAATAAGATG
GATCTGTTAGGAATGGTGGTTGTGGATGAATTACATATGCTGGGAGACTCTCACCGAG
GGTATCTGCTGGAACTTTTGCTGACCAAGATTTGCTATATACTCGGAAATCAGCATCT
TGTCAGGCAGATCTAGCCAGTTCTCTGTCTAATGCTGTGCAAATCGTTGGCATGAGTG
CTACCCTTCCTAATTTGGAGCTTGTGGCTTCCTGGTTGAATGCTGAACTCTACCATAACC
GACTTTCGCCCTGTACCGCTTTTGGAGTCAGTAAAAGTTGGAAATTCCATATATGACTC
TTCAATGAAACTTGTGAGGGAATTTGAGCCCATGCTTCAAGTGAAGGGAGATGAGGA
CCATGTTGTTAGCTTATGTTATGAGACGATTTGTGATAACCATTCAGTATTACTTTTTTG
TCCATCAAAGAAATGGTGTGAGAAGCTGGCAGATATCATTGCTCGAGAGTTTTATAAT
CTACATCATCAAGCTGAGGGATTGGTGAAACCCTCTGAATGCCCACCAGTAATTCTGG
AACAAAAAGAACTCCTGGAAGTGATGGATCAGTTAAGACGGTTGCCTTCAGGACTGG
ACTCTGTATTACAGAAAAGTGTACCATGGGGAGTAGCATTTTCATCATGCAGGTCTTACT
TTTGAGGAGAGGGATATCATTGAAGGAGCCTTTCGTCAAGGTCTCATTTCGAGTCTTGG
CGGCAACTTCTACTCTTTCTTCTGGGGTGAATTTACCTGCACGTCGTGTGATTATTCGA
ACCCCTATTTTTTGGTGGTCGACCTCTAGATATTCTTACTTATAAGCAGATGGTTGGTCGT
GCTGGCAGGAAAGGAGTGGACACAGTAGGCGAGAGTATCTTAATTTGTAAGAACTCT
GAGAAATCAAAGGCATAGCTCTCCTTCAGGGTTCTCTAAAGCCTGTTTCGCAGCTGTC
TGCAAAGACGAGAAGGAGAAGAAGTAACTGGCAGCATGATACGAGCTATTCTGGAGA
TAATAGTTGGTGGAGTGGCAAGTACATCACAAGATATGCATACTTATGCTGCCTGCACA
TTTTTGGCTGCAAGTATGAAAGAAGGGAAGCAAGGAATTCAGAGAAATCAAGAGTCT
GTTTCAGCTTGGAGCGATTGAGGCCTGTGTGATGTGGCTACTAGAAAATGAATTCATCC
AGAGTACAGAAGCCAGTGATGGAACAGAAGGAAAGGTGTATCATCCAACACATCTTG
GTTTCGGCCACTCTTTCTTCTTCACTTTCTCCAGCTGATACTTTAGATATTTTTGCTGACC
TGCAAAGAGCAATGAAGGGCTTTGTTTTAGAGAATGATCTTCATATTCTCTATCTGGTT
ACACCTATGTTTGAGGATTGGACTACTATTGATTGGTATCGATTTTTCTGTTTATGGGAG
AAGTTGCCAACTTCAATGAAAAGGGTGGCAGAGCTAGTGGGAGTTGAAGAGGGGTT
CTTGGCCCGTTGTGTGAAAGGAAAAGTAGTAGCCAGAACTGAGAGACAGCATCGACA
AATGGCCATCCATAAAAGGTTTTTCACCAGTCTTGTGCTATTAGATTTAATCAGTGAAG
TTCCCTTAAGGGAAATAAATCAGAAATATGGATGCAATCGTGGGCAGATTCAATCTTTG
CAACAGTCAGCTGCTGTTTATGCAGGGATGATTACAGTATTTTCCAACCGTCTGGGCT
GGCACAACATGGAATACTACTTTCCCAATTTTCAAGAGCGTCTTACGTTTGGCATCCA
GAGGGAGCTGTGTGACCTGGTTCGGGTATCCTTACTAAATGCTCAGAGAGCCAGGGTT
CTCTATGCTTCTGGCTTTCATACTGTGGCAGACCTTGCTAGAGCAAATATTGTGGAGGT
GGAGGTGATTCTGAAAAATGCTGTGCCTTTCAAAAGTGCCCGGAAGGCAGTGGATGA
GGAAGAGGAAGCAGTTGAAGAACGTCGCAATATGCGAACTATCTGGGTGACTGGCAG
AAAAGGTTTAACTGAAAGGGAAGCAGCAGCCCTTATAGTGGAAGAAGCCAGAATGAT
TCTGCAGCAGGACTTAGTTGAAATGTGA

Protein Expression

Medium: SF900II

Antibiotics: Ampicillin

Procedure: Baculoviruses were generated by recombination in *E. coli* DH10Bac (Life Technologies) followed by transfection and two rounds of amplification in SF9 cells. POLQ was expressed in 1-L cultures of SF9 cells in 4-L shaker flasks at 27°C, infected at 2×10^6 cells/ml with 3 ml of P2 virus, and incubated for further 70 h. The cells were collected by centrifugation, suspended in 15 ml/l of lysis buffer (50 mM HEPES, pH 7.5, 0.5 M NaCl, 5% v/v glycerol, 10 mM imidazole, and 1 mM TCEP) and frozen at -80°C.

Protein Purification

Procedure: Cell pellets were thawed and resuspended in buffer A (50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 10 mM imidazole, 0.5 mM Tris (2-carboxyethyl) phosphine (TCEP)), with the addition of 1x protease inhibitor set VII (Merck, Darmstadt, Germany). Cells were lysed using sonication and cell debris pelleted by centrifugation. Lysates were applied to a Ni-IDA IMAC gravity flow column, washed with 2 column volumes of wash buffer (buffer A supplemented with 30 mM imidazole), and eluted with the addition of 300 mM imidazole in buffer A. The purification tag was cleaved with the addition of 1:20 mass ratio of His-tagged TEV protease during overnight dialysis into buffer B (20 mM HEPES, pH 7.5, 500 mM NaCl, 5 % glycerol, 0.5 mM TCEP). TEV was removed by IMAC column rebinding and final protein purification was performed by size exclusion chromatography using a HiLoad 16/60 Superdex 200 column at 1 ml/min in buffer B. Protein concentrations were determined by measurement at 280nm (Nanodrop) using the calculated molecular mass and extinction coefficients.

The protein was analyzed by ESI-TOF intact mass spectrometry which revealed a major peak at 99858.0 Da which corresponds closely to the expected mass of (99852 Da) which includes a single R66I point mutant that was present in initial entry clones (reference PMID 14576298).

Structure Determination

Crystallization: For crystallization POLQ was concentrated to 17 mg/ml using a 50,000 mwco centrifugal concentrator and exchanged to a buffer containing 10 mM HEPES pH 7.5, 250 mM NaCl, 0.5 mM TCEP. Crystals were obtained at 20°C from conditions containing 20% PEG6000, 10% ethylene glycol, 0.1M HEPES pH 7.0, 0.2M sodium chloride. Crystals were cryo-protected by transferring to a solution of mother liquor supplemented with 25 % ethelene glycol and flash-cooled in liquid nitrogen.

Data Collection: Data were collected to 3.4 Å resolution at a wavelength of 0.97Å at Diamond light source beamline I03. Diffraction data were processed with the program XDS.

Data Processing: The structures were solved by molecular replacement using the program PHASER with PDBid 2ZJA as a starting model. Model building and real space refinement were performed in COOT and the structures refined using PHENIX REFINE to a final Rfactor= 22.6%, Rfree= 26.5%.