

**Text S1** Complete DNA and amino acid sequences of TALEN ORF. The RVDs are highlighted in yellow.

>B2-L DNA sequence

ATGAGATCTGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTAC  
AAGGATGACGATGACAAGATGGCCCCAAGAAGAAGAGGAAGGTGGGCATTCATGGG  
GTACCCATGATGTCTCGTACTCGCTTGCCTTCGCCTCCTGCTCCTTCTCCCGCCTTC  
TCCGCTGGTTCATTTTCTGACTTACTTCGCCAATTCGACCCGAGCTTATTCAACACT  
AGCCTTTTTGATTCCCTGCCTCCATTCGGTGCTCACCATACAGAAGCTGCCACTGGT  
GAATGGGACGAGTTCAATCCGACTGAGGGCAGCAGATGCACCACCACCTACCATG  
AGAGTGGCTGTTACGGCTGCCAGGCCACCGAGAGCTAAACCAGCACCAAGAAGGCGT  
GCAGCACAGCCATCAGACGCTTCTCCGGCTGCCAAGTGGATCTCAGGACCTTAGGT  
TACTCACAACAGCAACAGGAAAAAATCAAGCCAAAAGTTAGATCTACAGTCGCCAG  
CACCATGAGGCATTGGTGGGTACGGATTTACTCACGCGCATATTGTTGCTCTGTCTG  
CAGCATCCGGCAGCGTTGGGAACCGTCGCTGTAAAGTATCAAGACATGATAGCTGCC  
CTCCCCGAAGCCACACACGAGGCAATCGTGGGCGTTGGTAAACAATGGTCAGGCGCT  
AGGGCCTTGGAAGCGCTGTTGACAGTGGCTGGAGAGCTCAGAGGCCCCCTTTACAA  
CTTGATAACGGTCAGCTCTTAAAGATAGCTAAACGCGGTGGAGTCACGGCAGTAGAA  
GCGGTGCACGCTTGGCGTAACGCGTTAACAGGAGCTCCCCTGAATTTGACTCCTGAA  
CAGGTGGTTGCAATCGCGTCGAACAATGGCGGTAAACAAGCCCTTGAGACCGTACAG  
CGCCTTCTGCCAGTGTTGTGCCAAGCACACGGACTCACGCCGCAACAGGTCGTAGCT  
ATTGCCAGTAACAATGGAGGCAAACAAGCGTTAGAAACAGTTCAGCGTTTGCTCCCC  
GTCTTATGTCAAGCTCATGGCCTTACTCCTCAGCAAGTGGTTGCAATCGCGTCAAAC  
ATTGGTGGAAAGCAAGCCTTGAGACCGTACAGAGGTTACTTCCAGTGCTGTGCCAA  
GCACACGTTTTGACGCCGGAACAGGTCGTAGCTATAGCCTCTAATGGCGGTGGAAAA  
CAAGCGCTCGAGACAGTTCAGAGACTGTTGCCCGTCCTCTGTCAAGCACACGGATTA

ACACCAGAGCAAGTTGTCGCAATCGCGAGCCATGACGGCGGTAAACAAGCCCTGGAG  
ACCGTACAGGCACTCTTACCAGTGCTTTGCCAAGCACACGGACTGACGCCAGAGCAA  
GTCGTAGCTATTGCCTCCCATGACGGTGGCAAACAAGCGTTGGAGACAGTTCAGGCT  
CTTCTGCCCGTCTTGTGTCAAGCTCACGGTCTCACACCAGAACAGGTGGTTGCAATA  
GCGTCGAACATCGGTGGAAAGCAAGCCCTCGAGACCGTACAGGCATTGCTCCCAGTG  
TTATGCCAAGCTCACGGACTTACTCCTGAGCAGGTTGTAGCTATTGCCAGTCACGAT  
GGCGGTAAACAAGCTCTGGAGACAGTTCAGCGCTTACTTCCCGTCTGTGTCAAGCC  
CATGGCCTGACACCCGAGCAGGTGGTTGCAATTGCGTCAAACATAGGAGGCAAGCAA  
GCATTGGAGACCGTACAGCGTCTGTTGCCAGTGCTCTGCCAGGCTCACGGTTTAACT  
CCTCAGCAAGTTGTTGCTATAGCCTCTAACGGTGGAGGCAAACAAGCCCTCGAAACA  
GTTCAGAGGCTCTTACCCGTCCTTTGTCAAGCTCATGGACTTACACCCGAACAAGTT  
GTGGCAATCGCGAGCAACAATGGTGGAAAGCAAGCGCTGGAGACCGTACAGGCTCTT  
CTGCCAGTCTTGTGCCAAGCTCACGGCCTCACTCCTGAACAGGTCGTAGCTATCGCC  
TCCAACATTGGCGGTAAACAAGCCTTGGAGACAGTGCAGAGATTGCTCCCCGTTTTA  
TGTCAAGCTCATGGTCTTACACCCGAACAAGTCGTTGCAATTGCGTCGAATGGAGGT  
GGTAAACAGGCCCTGGAGACCGTCCAGCGCTTGTGTTGCCAGTGCTGTGCCAAGCTCAC  
GGACTGACTCCAGAGCAGGTTGTTGCTATAGCCAGTCATGACGGAGGCAAACAGGCT  
TTGGAGACAGTTCAGCGTTTGCTGCCTGTTCTCTGTCAAGCTCACGGCTTAACACCC  
GAACAAGTTGTTGCAATCGCGTCACATGATGGTGGAAAGCAAGCCCTGGAGACCGTG  
CAGAGACTCTTACCAGTTCTTTGCCAAGCTCACGGTCTGACTCCTCAGCAAGTTGTC  
GCTATTGCATCTAATGGAGGTGGAAGGCCAGCACTGGAAACAGTCCAACGTCTTCTG  
CCTGTGTTGTGTCAGGCTCATGGACTCACTCCAGAACAAGTGGTTGCAATAGCGAGC  
AATAACGGTGGTAAACAGGCTTTGGAGACCGTTCAGCGCTTGCTCCCCGTCTTATGC  
CAGGCTCACGGCCTTACGCCCCAGCAAGTTGTTGCTATTGCATCCAACGGAGGTGGT  
AGGCCTGCACTCGAATCGATAGTGGCACAATTAAGTCGTCCCGACCCTGCCCTTGCA  
GCGCTGACTAATGATCACTTGGTTCGCACTCGCGTGCTTAGGAGGCAGACCTGCCCTT

GATGCAGTGAAAAAGGGTCTGCCACATGCTCCCGCTCTCATCAAAAGAACCAATCGC  
CGTATTCCCGAACGCACCTCGCACCGTGTGGCTGGATCCCAGCTGGTGAAGAGCGAG  
CTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTAC  
ATCGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGACCGCATCCTGGAGATGAAG  
GTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGAAAGCACCTGGGCGGAAGC  
AGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGATTACGGCGTGATC  
GTGGACACAAAGGCCTACAGCGGGCTACAATCTGCCTATCGGCCAGGCCGACGAG  
ATGCAGAGATACGTGAAGGAGAACCAGACCCGGAATAAGCACATCAACCCCAACGAG  
TGGTGAAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCTGTGAGCGGC  
CACTTCAAGGGCAACTACAAGGCCAGCTGACCAGGCTGAACCACAAAACCAACTGC  
AATGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAGATGATCAAAGCC  
GGCACCTGACACTGGAGGAGGTGCGGGCAAGTTCAACAACGGCGAGATCAACTTC  
TGATAA

>B2-L amino acid sequence

MRSDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPMMSRTRLPSPPAPSPAF  
SAGSFSDLLRQFDPSLNFNTSLFDSLPPFGAHTTEAATGEWDEVQSGLRAADAPPPTM  
RVAVTAARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSSQQQEKIKPKVRSSTVAQ  
HHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGA  
RALEALLTVAGELRGPPLQLDTGQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTP  
QVVAIAS **NN**GGKQALETVQRLLPVLCQAHGLTPQQVVAIAS **NN**GGKQALETVQRLLP  
VLCQAHGLTPQQVVAIAS **NI**GGKQALETVQRLLPVLCQAHGLTPEQVVAIAS **NG**GGK  
QALETVQRLLPVLCQAHGLTPEQVVAIAS **HD**GGKQALETVQALLPVLCQAHGLTPEQ  
VVAIAS **HD**GGKQALETVQALLPVLCQAHGLTPEQVVAIAS **NI**GGKQALETVQALLPV  
LCQAHGLTPEQVVAIAS **HD**GGKQALETVQRLLPVLCQAHGLTPEQVVAIAS **NI**GGKQ  
ALETVQRLLPVLCQAHGLTPQQVVAIAS **NG**GGKQALETVQRLLPVLCQAHGLTPEQV

VAIAS<sup>NN</sup>GGKQALETVQALLPVLCQAHGLTPEQVVAIAS<sup>NI</sup>GGKQALETVQRLLPVL  
CQAHGLTPEQVVAIAS<sup>NG</sup>GGKQALETVQRLLPVLCQAHGLTPEQVVAIAS<sup>HD</sup>GGKQA  
LETVQRLLPVLCQAHGLTPEQVVAIAS<sup>HD</sup>GGKQALETVQRLLPVLCQAHGLTPQQVV  
AIAS<sup>NG</sup>GGRPALETVQRLLPVLCQAHGLTPEQVVAIAS<sup>NN</sup>GGKQALETVQRLLPVLC  
QAHGLTPQQVVAIAS<sup>NG</sup>GGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPAL  
DAVKKGLPHAPALIKRTNRRIPERTSHRVAGSQLVKSELEEKSELRHKLKYPHEY  
IELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVI  
VDTKAYSGGYNLPIGQADEMQRYVKENQTRNKHINPNEWKVPSSVTEFKFLFVSG  
HFKGNYKAQLTRLNHKTNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF

>B2-R DNA sequence

ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGAT  
GACGATGACAAGATGGCCCCAAGAAGAAGAGGAAGGTGGGCATCCACGGGGTACCC  
ATGATGTCGCGCACAAAGATTGCCAAGCCCACCAGCACCTTCACCAGCATTCTCAGCG  
GGTTCCTTTTCCGATTTACTCCGTCAGTTCGACCCGTCGTTATTCAACACTTCGCTT  
TTTGATAGTCTGCCTCCATTTCGGTGCTCACCATACAGAAGCTGCCACTGGAGAATGG  
GACGAGGTTCAAAGTGGACTGAGGGCAGCAGATGCACCACCACCTACCATGAGAGTG  
GCTGTTACGGCTGCAAGGCCACCAAGAGCTAAGCCAGCACCAAGAAGGCGTGCAGCA  
CAGCCAAGCGACGCTTCCCCAGCTGCACAAGTGGATCTCAGGACCTTAGGTTACAGC  
CAACAGCAACAGGAAAAAATCAAGCCAAAAGTTAGATCCACAGTCGCCCAGCACCAT  
GAGGCATTGGTGGGACACGGCTTTACTCACGCGCATATTGTTGCTCTGTCACAGCAT  
CCGGCAGCGTTGGGAACCGTCGCTGTAAAGTATCAAGACATGATAGCTGCCCTCCCC  
GAAGCCACACACGAGGCAATCGTGGGTGTTGGAAAACAATGGTCAGGTGCTAGGGCC  
TTGGAAGCGCTGTTGACAGTGGCTGGCGAGCTCAGAGGTCCCCCTTTACAACCTTGAT  
ACCGGACAGCTCTTAAAGATAGCTAAACGCGGTGGAGTCACGGCAGTAGAAGCGGTG  
CACGCTTGGCGTAACGCGTTAACAGGAGCTCCCCTGAATTTGACTCCTGAACAGGTG

GTTGCAATCGCGTCAAACGGCGGTGGAAAGCAAGCCCTTGAGACCGTACAGCGCCTT  
CTGCCAGTGTGTGCCAAGCACACGGCCTCACGCCGCAACAGGTCGTAGCTATTGCC  
TCTAATGGCGGTGGAAAACAAGCGTTAGAAACAGTTCAGCGTTTGCTCCCCGTCTTA  
TGTC AAGCTCATGGTCTTACTCCTCAGCAAGTGGTTGCAATCGCGAGCATTGGTGGA  
GGAAAGCAAGCCTTGGAGACCGTACAGAGGTTACTTCCAGTGCTGTGCCAAGCACAC  
GGATTGACGCCGGAACAGGTCGTAGCTATAGCCTCCAACGGTGGAGGAAAACAAGCG  
CTCGAGACAGTTCAGAGACTGTTGCCCGTCCTCTGTCAAGCACATGGCTTAACACCA  
GAGCAAGTTGTGCAATCGCGTCGAATGGTGGCGGAAAGCAAGCCCTGGAGACCGTA  
CAGGCACTCTTACCAGTGCTTTGCCAAGCCCACGGTCTTACTCCTGAGCAAGTTGTA  
GCTATTGCCAGTAACGGTGGCGGAAAACAAGCGTTGGAGACAGTTCAGGCTCTTCTG  
CCCGTCTTGTGTCAAGCTCATGGACTCACACCAGAACAGGTGGTTGCAATAGCGTCA  
AATATCGGCGGTAAACAAGCCCTCGAGACCGTACAGGCATTGCTCCAGTGTTATGC  
CAAGCTCACGGCCTTACTCCTGAGCAAGTGGTTGCTATTGCCTCTAACATAGGAGGC  
AAACAAGCTCTGGAGACAGTTCAGCGCTTACTTCCCGTCCTGTGTCAAGCCCATGGA  
CTGACACCAGAGCAAGTTGTGGCAATCGCGAGCAATATTGGTGGAAAGCAAGCATTG  
GAGACCGTACAGCGTCTGTTGCCAGTGCTCTGCCAGGCTCACGGATTAACCTCCTCAG  
CAAGTTGTTGCTATAGCCTCCAACAATGGTGGTAAACAGGCCCTCGAAACAGTTCAG  
AGGCTCTTACCCGTCTTTGTCAAGCTCATGGCCTTACACCCGAGCAGGTTGTGCA  
ATCGCGTCAAACCTCTGGAGGCAAGCAAGCGCTGGAGACCGTACAGGCTCTTCTGCCA  
GTCTTGTGCCAAGCTCACGGTCTCACTCCTGAGCAAGTCGTAGCTATTGCCTCGAAT  
AAGGGTGGAAAACAAGCCTTGGAGACAGTGCAGAGATTGCTCCCCGTTTTATGTCAA  
GCTCACGGACTTACACCCGAGCAGGTCGTTGCAATAGCGAGTCATGATGGTGGTAAA  
CAAGCCCTGGAGACCGTCCAGCGCTTGTGGCCAGTGCTGTGCCAAGCTCATGGCCTG  
ACTCCTGAGCAGGTCGTAGCTATCGCCTCAAACGGAGGCGGTAAACAGGCTTTGGAG  
ACAGTTCAGCGTTTGCTGCCTGTTCTCTGTCAAGCTCACGGTTTAACACCCGAACAG  
GTGGTTGCAATTGCGTCTAACAATGGAGGCAAACAGGCTCTGGAGACCGTGCAGAGA

CTCTTACCAGTTCTTTGCCAAGCTCATGGACTGACTCCTCAGCAAGTCGTAGCTATA  
GCAAGCAATGGTGGAGGAAGGCCAGCACTGGAAACAGTCCAACGTCTTCTGCCTGTG  
TTGTGTCAGGCTCACGGCCTCACTCCAGAACAAGTGGTTGCAATTGCGTCCAACAAT  
GGTGGAAAACAGGCACTGGAGACCGTTCAGCGCTTGCTCCCGGTCTTATGCCAGGCT  
CACGGACTTACGCCCCAGCAAGTGGTCGCTATTGCATCGAACGGAGGTGGAAGGCCT  
GCACTCGAATCAATAGTGGCACAATTATCTCGTCCCGACCCTGCCCTTGCAGCGCTG  
ACTAATGATCACTTGGTCGCACTCGCGTGCTTAGGCGGTAGACCTGCCCTTGATGCA  
GTGAAAAGGGTCTGCCACATGCTCCCGCACTGATAAAACGCACCAACCGTCGTATT  
CCTGAACGCACCTCACATCGTGTGCGCAGGATCCCAGCTGGTGAAGAGCGAGCTGGAG  
GAGAAGAAGTCCGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTACATCGAG  
CTGATCGAGATCGCCAGGAACAGCACCCAGGACCGCATCCTGGAGATGAAGGTGATG  
GAGTTCTTCATGAAGGTGTACGGCTACAGGGGAAAGCACCTGGGCGGAAGCAGAAAG  
CCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGATTACGGCGTGATCGTGGAC  
ACAAAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAGGCCGACGAGATGGAG  
AGATACGTGGAGGAGAACCAGACCCGGAATAAGCACCTCAACCCCAACGAGTGGTGG  
AAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGCGGCCACTTC  
AAGGGCAACTACAAGGCCAGCTGACCAGGCTGAACCACATCACCAACTGCAATGGC  
GCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAGATGATCAAAGCCGGCACC  
CTGACACTGGAGGAGGTGCGGCGCAAGTTCACAACGGCGAGATCAACTTC

>B2-R amino acid sequence

MDYKDHDGDYKDHDIDYKDDDDKMAPKKRKVGIHGVPMMSRTRLPSPPAPSPAFSA  
GSFSDLRLRQFDPSLNFNTSLFDSLPPFGAHHTEAATGEWDEVQSGLRAADAPPPTMRV  
AVTAARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSSQQQEKIKPKVIRSTVAQHH  
EALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARA  
LEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPQV

VAIASNGGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNGGGKQALETVQRLLPVL  
CQAHGLTPQQVVAIASIGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQA  
LETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQALETVQALLPVLCQAHGLTPEQV  
AIASNGGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLC  
QAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQAL  
ETVQRLLPVLCQAHGLTPQQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVA  
IASNSGGKQALETVQALLPVLCQAHGLTPEQVVAIASNKGKQALETVQRLLPVLCQ  
AHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQALE  
TVQRLLPVLCQAHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPQQVVAI  
ASNGGGRPALETVQRLLPVLCQAHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQA  
HGLTPQQVVAIASNGGGRPALESIVAQLSRPDPALAALTNDDLVALACLGGRPALDA  
VKKGLPHAPALIKRTNRRIPERTSHRVAGSQLVKSELEEKSELRHKLKYVPHEYIE  
LIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVD  
TKAYSGGYNLPIGQADEMERYVEENQTRNKHLNPNEWKVPSSVTEFKFLFVSGHF  
KGNKYAQLTRLNHIITNCNGAVLSVEELLIIGGEMIKAGTLTLEEVRKFNNGEINF

>B3-L DNA sequence

ATGAGATCTGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTAC  
AAGGATGACGATGACAAGATGGCCCCAAGAAGAAGAGGAAGGTGGGCATTCATGGG  
GTACCCATGATGAGCAGAACCAGACTCCCCAGCCCACCAGCGCCCTCGCCAGCGTTT  
TCGGCAGGCAGCTTTTCGGACTTGTTACGTCAGTTTGACCCAAGCTTATTCAACACT  
AGCCTTTTTGATTCCCTGCCTCCATTCGGTGCTCACCATACAGAAGCTGCCACTGGA  
GAATGGGACGAGGTTCAATCCGACTGAGGGCAGCAGATGCACCACCACCTACCATG  
CGTGTGGCAGTTACGGCTGCCAGGCCACCGAGAGCTAAACCAGCACCAAGAAGGCGT  
GCAGCACAACCATCAGACGCTTCTCCGGCTGCCAGGTGGATCTCCGCACCTTAGGT  
TACTCACAACAGCAACAGGAAAAAATTAAGCCGAAAGTTCGTTCTACAGTCGCCCAA

CACCATGAGGCATTGGTGGGACACGGCTTTACTCACGCGCATATAGTTGCTCTGTCTG  
CAACATCCGGCAGCGTTGGGAACCGTCGCTGTAAAGTATCAGGACATGATCGCTGCC  
CTCCCTGAAGCCACACACGAGGCAATTGTGGGTGTTGGAAAACAGTGGTCAGGTGCT  
CGCGCCTTGGAAGCGCTGTTGACAGTGGCTGGAGAGCTCCGTGGTCCACCTTTACAA  
CTTGATACCGGACAGCTCTTAAAGATCGCTAAAAGGGGTGGAGTCACGGCAGTAGAA  
GCGGTGCACGCTTGGAGAAACGCGTTAACAGGAGCTCCCCTGAATTTGACTCCTGAA  
CAAGTGGTTGCAATTGCGTCGCACGACGGCGGTAAACAAGCTCTTGAGACCGTACAG  
AGGCTTCTGCCAGTGTGTTGTGCCAAGCACATGGACTCACGCCAGCACAGGTCGTAGCT  
ATCGCCAGTAATATTGGAGGCAAACAAGCGTTAGAAACAGTTCAGAGATTGCTCCCC  
GTCTTATGTCAAGCTCACGGTCTTACTCCTGACCAGGTGGTTGCAATAGCGTCACAT  
GATGGTGGAAAGCAAGCTTTGGAGACCGTACAGCGCTTACTTCCAGTGCTGTGCCAA  
GCGCACGGATTGACGCCGGCTCAAGTTGTAGCTATCGCCTCTCATGATGGCGGTAAA  
CAAGCCCTCGAAACAGTTCAGCGTCTGTTGCCCGTCCTCTGTCAAGCACACGGCTTA  
ACTCCTGAACAGTTGTGCAATAGCGAGCAACATCGGAGGCAAGCAAGCGCTGGAG  
ACCGTACAGAGGCTCTTACCAGTGCTTTGCCAAGCTCATGGTCTGACGCCGGACCAG  
GTCGTAGCTATTGCCTCCAATATAGGTGGAAAACAAGCCTTGGAGACAGTTCAGAGA  
CTTCTGCCCGTCTTGTGTCAAGCACACGGACTCACTCCTGATCAGGTTGTGCAATT  
GCGTCGAACAATGGCGGTAAACAAGCGCTCGAAACCGTACAGCGCTTGCTCCCAGTG  
TTATGCCAAGCACATGGCCTTACGCCGGAACAGGTCGTAGCTATAGCCAGTAACGGA  
GGCGGTAAACAAGCCCTGGAGACAGTTCAGCGTTTACTTCCCGTCCTGTGTCAAGCT  
CACGGTTTGACACCAGACCAGGTGGTTGCAATCGCGTCAAATATTGGAGGCAAGCAA  
GCCTTAGAAACCGTACAGAGGCTGTTGCCAGTGCTCTGCCAAGCTCACGGATTAACT  
CCTGAGCAAGTCGTTGCAATCGCCTCTCATGATGGTGGAAAACAAGCCCTCGAGACA  
GTTCAGAGACTCTTACCCGTCCTTTGTCAAGCTCACGGCCTGACTCCTCAACAGGTG  
GTTGCAATTGCGAGCAATAACGGTGGTAAACAGGCCCTGGAAACCGTACAGCGCCTT  
CTGCCAGTCTTGTGCCAAGCTCACGGTCTCACTCCTGAGCAAGTCGTAGCTATAGCC



TCCCATGACGGAGGCAAACAGGCTTTGGAGACAGTGCAGCGTTTGCTCCCCGTTTTA  
TGTC AAGCTCATGGACTTACTCCTGATCAAGTTGTTGCAATAGCGTCGAACATCGGT  
GGAAAGCAAGCTCTGGAGACCGTCCAGAGGTTGTTGCCAGTGCTGTGCCAAGCACAC  
GGATTGACTCCTGCACAAGTTGTAGCTATTGCCAGTAATAACGGCGGTAAACAAGCC  
TTGGAAACAGTTCAGCGCTTGCTGCCTGTTCTCTGTCAAGCTCATGGTTTAACTCCT  
GAGCAAGTTGTCGCAATTGCGTCAAACATAGGTGGCAAGCAGGCCCTGGAGACCGTG  
CAGCGTCTCTTACCAGTTCTTTGCCAAGCTCACGGACTGACGCCGCAACAAGTGGA  
GCTATTGCCTCTAATGGTGGAGGCAAACAAGCCCTGGAAACAGTCCAGAGACTTCTG  
CCCGTGTGTGTCAAGCTCACGGCCTCACTCCTGAACAGGTTGTGGCAATCGCGAGC  
CATGACGGTGGAAAGCAGGCTTTGGAGACCGTTCAGCGCTTGCTCCAGTCTTATGC  
CAAGCTCACGGATTGACTCCTGAGCAGGTCGTAGCTATAGCATCCAACGGAGGTGGA  
AGGCCAGCACTCGAGTCGATCGTGGCTCAATTAAGTAGACCCGACCCTGCCCTTGCA  
GCGCTGACTAATGATCACTTGGTTCGCACTCGCGTGCTTAGGCGGTAGACCCGCCCTT  
GATGCAGTGAAAAAGGTTCTGCCACATGCTCCAGCACTCATCAAAGAACCAATCGT  
CGTATTCCCGAGAGGACATCACACAGGGTGGCGGGATCCCAGCTGGTGAAGAGCGAG  
CTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTAC  
ATCGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGACCGCATCCTGGAGATGAAG  
GTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGAAAGCACCTGGGCGGAAGC  
AGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGATTACGGCGTGATC  
GTGGACACAAAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAGGCCGACGAG  
ATGCAGAGATACGTGAAGGAGAACCAGACCCGGAATAAGCACATCAACCCCAACGAG  
TGGTGGAAAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTTCGTGAGCGGC  
CACTTCAAGGGCAACTACAAGGCCAGCTGACCAGGCTGAACCACAAAACCAACTGC  
AATGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAGATGATCAAAGCC  
GGCACCCCTGACACTGGAGGAGGTGCGGCGCAAGTTCAACAACGGCGAGATCAACTTC  
TGATAA

>B3-L amino acid sequence

MRSDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIIHGVPMMSTRRLPSPAPSPAF  
SAGSFSDLLRQFDPSLNFNTSLFDSLPPFGAHTTEAATGEWDEVQSGLRAADAPPPTM  
RVAVTAARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSSQQQEKIKPKVIRSTVAQ  
HHEALVGHGFTHAHIVALSQHPAALGTAVVKYQDMIAALPEATHEAIVGVGKQWSGA  
RALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPE  
QVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLP  
VLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGK  
QALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQ  
VVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPV  
LCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQ  
ALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPQQV  
VAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLC  
QAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQA  
LETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPQQVV  
AIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLC  
QAHGLTPEQVVAIASNGGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPAL  
DAVKKGLPHAPALIKRTNRRIPERTSHRVAGSQLVKSELEKKSELRHKLKYVPHEY  
IELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVI  
VDTKAYSGGYNLPIGQADEMQRYVKENQTRNKHINPNEWKVPSSVTEFKFLFVSG  
HFKGNYKAQLTRLNHTNCNGAVLSVEELLIGGEMIKAGTLTLEEVRKFNNGEINF

>B3-R DNA sequence

ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGAT  
GACGATGACAAGATGGCCCCAAGAAGAAGAGGAAGGTGGGCATCCACGGGGTACCC

ATGATGTCTCGCACTCGTCTCCCTAGCCCTCAGCACCGTCACCAGCATTCTCGGCA  
GGTTCCTTCTCGGACTTATTACGCCAGTTTGACCCGAGCTTATTCAATACTAGCCTT  
TTTGATTCCCTGCCTCCATTTCGGTGCTCACCATACAGAAGCTGCCACTGGCGAATGG  
GACGAGGTTCAATCCGGTCTGAGGGCAGCAGATGCACCACCACCTACCATGAGAGTG  
GCTGTTACGGCTGCCAGGCCACCGAGAGCTAAACCAGCACCAAGAAGGCGTGCAGCA  
CAGCCATCAGACGCTTCTCCGGCTGCCAAGTGGATCTCAGGACCTTAGGTTACTCA  
CAACAGCAACAGGAAAAAATCAAACCAAAGTTAGATCTACAGTCGCCCAGCACCAT  
GAGGCATTGGTGGGCCACGGTTTTACTCACGCGCATATCGTTGCTCTGTGCAGCAT  
CCGGCAGCGTTGGGAACCGTCGCTGTAAAGTATCAAGACATGATTGCTGCCCTCCCC  
GAAGCCACACACGAGGCAATAGTGGGAGTTGGCAAACAATGGTCAGGAGCTAGGGCC  
TTGGAAGCGCTGTTGACAGTGGCTGGAGAGCTCAGAGGACCCCTTTACAACCTTGAT  
ACCGGCCAGCTCTTAAAGATTGCTAAACGCGGTGGAGTCACGGCAGTAGAAGCGGTG  
CACGCTTGGCGTAACGCGTTAACAGGAGCTCCCCTGAATTTGACTCCTGAACAGGTG  
GTTGCAATAGCGTCGCACGACGGCGGTAAACAAGCCCTTGAGACCGTACAGCGCCTT  
CTGCCAGTGTTGTGCCAAGCACATGGTCTCACGCCGCAACAGGTCGTAGCTATCGCC  
AGTAACATTGGAGGCAAACAAGCGTTAGAAACAGTTCAGCGTTTGCTCCCCGTCTTA  
TGTCAAGCTCACGGACTTACTCCTCAACAGGTGGTTGCAATCGCGTCAAACAATGGT  
GGAAAGCAAGCCTTGAGACCGTACAGAGGTTACTTCCAGTGCTGTGCCAAGCACAT  
GGCTTGACGCCGGAACAGGTCGTAGCTATTGCCTCTAATAAGGGCGGTAAACAAGCG  
CTCGAGACAGTTCAGAGACTGTTGCCCGTCTCTGTCAAGCACACGGTTTAAACACCA  
GAGCAAGTGGTTGCAATAGCGAGCAACAATGGAGGCAAACAAGCCCTGGAGACCGTA  
CAGGCACTCTTACCAGTGCTTTGCCAAGCCCATGGACTTACTCCTGAGCAAGTTGTA  
GCTATCGCCTCCAACGGTGGAGGCAAGCAAGCGTTGGAGACAGTTCAGGCTCTTCTG  
CCCGTCTTGTGTCAAGCTCACGGCCTCACACCAGAACAGGTGGTTGCAATTGCGTCG  
CATGATGGTGGAAAACAAGCCCTCGAGACCGTACAGGCATTGCTCCAGTGTTATGC  
CAAGCTCACGGTCTTACTCCTGAACAGGTCGTAGCTATAGCCAGTAACGGCGGTGGA

AAGCAAGCTCTGGAGACAGTTCAGCGCTTACTTCCCGTCCTGTGTCAAGCCCATGGA  
CTGACACCCGAGCAAGTTGTTGCAATCGCGTCAAATGGCGGTGGAAAACAAGCATTG  
GAGACCGTACAGCGTCTGTTGCCAGTGCTCTGCCAGGCTCACGGCTTAACTCCTCAG  
CAAGTTGTTGCTATTGCCTCTAACAATGGCGGTAAACAAGCCCTCGAAACAGTTCAG  
AGGCTCTTACCCGTCTTTTGTCAAGCTCATGGTCTCACACCCGAGCAGGTTGTGCGA  
ATAGCGAGCAACGGAGGCGGTAAACAAGCGCTGGAGACCGTACAGGCTCTTCTGCCA  
GTCTTGTGCCAAGCTCACGGACTCACTCCAGAGCAAGTCGTAGCTATCGCCTCCAAT  
GGAGGCGGTAAACAAGCCTTGGAGACAGTGCAGAGATTGCTCCCCGTTTTATGTCAA  
GCTCACGGCCTTACACCCGAGCAGGTTGTGGCTATTGCGTCGCATGACGGTGGCAAA  
CAGGCCCTGGAGACCGTCCAGCGCTTGTGTCAGTGCTGTGCCAAGCTCATGGTCTG  
ACTCCAGAGCAGGTTGTTGCTATAGCCAGTAACATCGGAGGAAAGCAGGCCTTGGAG  
ACAGTTCAGCGTTTGCTGCCTGTTCTCTGTCAAGCTCACGGATTAACACCCGAGCAG  
GTTGTTGCAATAGCGTCAAATAACGGTGGTAAACAGGCTTTGGAGACCGTGCAGAGA  
CTCTTACCAGTCTTTTGCCAAGCTCATGGCCTGACTCCTCAGCAAGTGGTTGCTATC  
GCCTCTAACAAGGAGGCCCGCCCCGCCCTGGAAACAGTCCAACGTCTTCTGCCTGTG  
TTGTGTCAGGCTCACGGTCTCACTCCAGAACAAGTGGTTGCAATTGCGAGCAATGGA  
GGTGGCAAACAAGCTTTGGAGACCGTTCAGCGCTTGCTCCCGGTCTTATGCCAGGCT  
CACGGACTTACGCCCCAGCAAGTGGTTCGCTATCGCATCCAACGGTGGAGGAAGGCCT  
GCACTCGAATCGATTGTGGCACAATTAAGTCGTCCCGACCCTGCCCTTGCAGCGCTG  
ACTAATGATCACTTGGTCGCACTCGCGTGCTTAGGTGGAAGACCTGCCCTTGATGCA  
GTGAAAAGGGTCTGCCACATGCTCCCGCACTCATCAAGAGAACTAATAGACGCATA  
CCCGAGAGAACCAGCCACCGTGTGCTGGATCCCAGCTGGTGAAGAGCGAGCTGGAG  
GAGAAGAAGTCCGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTACATCGAG  
CTGATCGAGATCGCCAGGAACAGCACCCAGGACCGCATCCTGGAGATGAAGGTGATG  
GAGTTCTTCATGAAGGTGTACGGCTACAGGGGAAAGCACCTGGGCGGAAGCAGAAAG  
CCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGATTACGGCGTGATCGTGGAC

ACAAAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAGGCCGACGAGATGGAG  
AGATACGTGGAGGAGAACCAGACCCGGAATAAGCACCTCAACCCCAACGAGTGGTGG  
AAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCTGTGAGCGGCCACTTC  
AAGGGCAACTACAAGGCCAGCTGACCAGGCTGAACCACATCACCAACTGCAATGGC  
GCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAGATGATCAAAGCCGGCACC  
CTGACACTGGAGGAGGTGCGGCGCAAGTTCACAACGGCGAGATCAACTTC

>B3-R amino acid sequence

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIVGVPMMSRTRLPSPPAPSPAFSA  
GSFSDLLRQFDP SLFN TSLFDSLPPFGAHHTEAATGEWDEVQSGLRADAPPPTMRV  
AVTAARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSSQQQEKIKPKVRSSTVAQHH  
EALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARA  
LEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPEQV  
VAIAS<sup>HD</sup>GGKQALETVQRLLPVLCQAHGLTPQQVVAIAS<sup>NI</sup>GGKQALETVQRLLPVL  
CQAHGLTPQQVVAIAS<sup>NN</sup>GGKQALETVQRLLPVLCQAHGLTPEQVVAIAS<sup>NK</sup>GGKQA  
LETVQRLLPVLCQAHGLTPEQVVAIAS<sup>NN</sup>GGKQALETVQALLPVLCQAHGLTPEQVV  
AIAS<sup>NG</sup>GGKQALETVQALLPVLCQAHGLTPEQVVAIAS<sup>HD</sup>GGKQALETVQALLPVLC  
QAHGLTPEQVVAIAS<sup>NG</sup>GGKQALETVQRLLPVLCQAHGLTPEQVVAIAS<sup>NG</sup>GGKQAL  
ETVQRLLPVLCQAHGLTPQQVVAIAS<sup>NN</sup>GGKQALETVQRLLPVLCQAHGLTPEQVVA  
IAS<sup>NG</sup>GGKQALETVQALLPVLCQAHGLTPEQVVAIAS<sup>NG</sup>GGKQALETVQRLLPVLCQ  
AHGLTPEQVVAIAS<sup>HD</sup>GGKQALETVQRLLPVLCQAHGLTPEQVVAIAS<sup>NI</sup>GGKQALE  
TVQRLLPVLCQAHGLTPEQVVAIAS<sup>NN</sup>GGKQALETVQRLLPVLCQAHGLTPQQVVAI  
AS<sup>NK</sup>GGRPALETVQRLLPVLCQAHGLTPEQVVAIAS<sup>NG</sup>GGKQALETVQRLLPVLCQA  
HGLTPQQVVAIAS<sup>NG</sup>GGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDA  
VKKGLPHAPALIKRTNRRIPERTSHRVAGSQLVKSELEEKSELRHKLKYVPHEYIE  
LIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVD

TKAYSGGYNLP I G Q A D E M E R Y V E E N Q T R N K H L N P N E W W K V Y P S S V T E F K F L F V S G H F  
K G N Y K A Q L T R L N H I T N C N G A V L S V E E L L I G G E M I K A G T L T L E E V R R K F N N G E I