pACYC-GST Vector

Source	Constructed by Farrell MacKenzie
Company	Structural Genomics Consortium, Toronto

Description	pACYC-GST is a derivative of the pACYC-LIC+ vector (SGC) and contains the p15A origin, cat gene and argU, ileY and leuW tRNA genes, as well as lacI, T7 promoter and T7 terminator regions contained in pACYC-LIC+. The plasmid is compatible with pET based plasmids. It is used for co-expression of recombinant proteins with the addition of a 231 amino acid N-terminal fusion tag containing a GST-tag followed by a TEV protease cleavage site. Two stop codons are included in the vector at the C-terminal cloning site.

Antibiotic resistance	Chloramphenicol
Promoter	T7 - lacO
Cloning Methods	Insertion of a DNA sequence into the cloning/expression region is performed using Clontech's In-fusion enzyme-mediated directional recombination between complementary 15 nucleotide DNA sequences at the ends of the insert (PCR product) and BsaI linearized vector. Insertion of a target sequence involves replacement of a SacB gene stuffer sequence, which provides for negative selection of the original plasmid on 5% sucrose.
N – terminal fusion sequence	MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEG-DKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYIAD-KHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDF-ETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDF-MLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKY-LKSSKYIAWPLQGWQATFGGGDHPPKSDSSGRENLYF-QG
5' primer tail for amplification of insert	5' TTGTATTTCCAGGGC --- 3'
3' primer tail for amplification of insert	5' CAAGCTTCGTCATCA --- 3'
5' sequencing primer T7	5' AATTAATACGACTCACTATAGGG 3'
3' sequencing primer T7term	5' ATGCTAGTTATTGCTCAGCGG 3'

pACYC-GST sequence (8,362 bp):

CGCCGCTAGGTGTGCGAGTTCAAGTCTCGCCTCCCGCACCATTCACCAGAAAGCG
TTGATCGGATGCCCTCGAGTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGCAT
GATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGT
TAGCAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAA
CGTCTGCGACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGT
CTGGAAACGCGGAAGTCCCCTACGTGCTGCTGAAGTTGCCCGCAACAGAGAGTGG
AACCAACCGGTGATACCACGATACTATGACTGAGAGTCAACGCCATGAGCGGCCTC
ATTTCTTATTCTGAGTTACAACAGTCCGCACCGCTGCCGGTAGCTCCTTCCGGTGG
GCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACT
CGTAGGACAGGTGCCGGCAGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCAC
CATACCCACGCCGAAACAAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAG
GATGCTGCTGGCTACCCTGTGGAACACCTACATCTGTATTAACGAAGCGCTAACCG
TTTTTATCAGGCTCTGGGAGGCAGAATAAATGATCATATCGTCAATTATTACCTCCA
CGGGGAGAGCCTGAGCAAACTGGCCTCAGGAGTCATGCCCCGCGCCCACCGGAA
GGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAA
TGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGG
AAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGT
TTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCT
GATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGT
TTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATG
AGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGC
CCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCA
GCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCG
GACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGT
GAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGG
CCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCA
GTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAG
ACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATC
CTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGAT
TGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACC
ACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACG
GCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCC
CGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTT
CCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAA
ACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTT
CACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAA
AGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTC
CTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCA
AGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCT
GCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGAT
CTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCC
GGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCG
AAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAA
ATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGTCCCCTATACTAGGTTATTG
GAAAATTAAGGGCCTTGTGCAACCCACTCGACTTCTTTTGGAATATCTTGAAGAAAA
ATATGAAGAGCATTTGTATGAGCGCGATGAAGGTGATAAATGGCGAAACAAAAAGT
TTGAATTGGGTTTGGAGTTTCCCAATCTTCCTTATTATATTGATGGTGATGTTAAATT

AACACAGTCTATGGCCATCATACGTTATATAGCTGACAAGCACAACATGTTGGGTG
GTTGTCCAAAAGAGCGTGCAGAGATTTCAATGCTTGAAGGAGCGGTTTTGGATATT
AGATACGGTGTTTCGAGAATTGCATATAGTAAAGACTTTGAAACTCTCAAAGTTGAT
TTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAA
CATATTTAAATGGTGATCATGTAACCCATCCTGACTTCATGTTGTATGACGCTCTTG
ATGTTGTTTTATACATGGACCCAATGTGCCTGGATGCGTTCCCAAAATTAGTTTGTT
TTAAAAAACGTATTGAAGCTATCCCACAAATTGATAAGTACTTGAAATCCAGCAAGT
ATATAGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCC
TCCAAAATCGGATAGCAGCGGCAGAGAAAACTTGTATTTCCAGGGCAGAGACCCTG
AAAGATCCATAACTTCGTATAGCATACATTATACGAAGTTATGCGGCCGCGACGTCC
ACATATACCTGCCGTTCACTATTATTTAGTGAAATGAGATATTATGATATTTTCTGAA
TTGTGATTAAAAAGGCAACTTTATGCCCATGCAACAGAAACTATAAAAAATACAGAG
AATGAAAAGAAACAGATAGATTTTTTAGTTCTTTAGGCCCGTAGTCTGCAAATCCTTT
TATGATTTTCTATCAAACAAAAGAGGAAAATAGACCAGTTGCAATCCAAACGAGAGT
CTAATAGAATGAGGTCGAAAAGTAAATCGCGCGGGTTTGTTACTGATAAAGCAGGC
AAGACCTAAAATGTGTAAAGGGCAAAGTGTATACTTTGGCGTCACCCCTTACATATT
TTAGGTCTTTTTTTATTGTGCGTAACTAACTTGCCATCTTCAAACAGGAGGGCTGGA
AGAAGCAGACCGCTAACACAGTACATAAAAAAGGAGACATGAACGATGAACATCAA
AAAGTTTGCAAAACAAGCAACAGTATTAACCTTTACTACCGCACTGCTGGCAGGAG
GCGCAACTCAAGCGTTTGCGAAAGAAACGAACCAAAAGCCATATAAGGAAACATAC
GGCATTTCCCATATTACACGCCATGATATGCTGCAAATCCCTGAACAGCAAAAAAAT
GAAAAATATCAAGTTCCTGAGTTCGATTCGTCCACAATTAAAAATATCTCTTCTGCAA
AAGGCCTGGACGTTTGGGACAGCTGGCCATTACAAAACGCTGACGGCACTGTCGC
AAACTATCACGGCTACCACATCGTCTTTGCATTAGCCGGAGATCCTAAAAATGCGG
ATGACACATCGATTTACATGTTCTATCAAAAAGTCGGCGAAACTTCTATTGACAGCT
GGAAAAACGCTGGCCGCGTCTTTAAAGACAGCGACAAATTCGATGCAAATGATTCT
ATCCTAAAAGACCAAACACAAGAATGGTCAGGTTCAGCCACATTTACATCTGACGG
AAAAATCCGTTTATTCTACACTGATTTCTCCGGTAAACATTACGGCAAACAAACACT
GACAACTGCACAAGTTAACGTATCAGCATCAGACAGCTCTTTGAACATCAACGGTG
TAGAGGATTATAAATCAATCTTTGACGGTGACGGAAAAACGTATCAAAATGTACAGC
AGTTCATCGATGAAGGCAACTACAGCTCAGGCGACAACCATACGCTGAGAGATCCT
CACTACGTAGAAGATAAAGGCCACAAATACTTAGTATTTGAAGCAAACACTGGAACT
GAAGATGGCTACCAAGGCGAAGAATCTTTATTTAACAAAGCATACTATGGCAAAAGC
ACATCATTCTTCCGTCAAGAAAGTCAAAAACTTCTGCAAAGCGATAAAAAACGCACG
GCTGAGTTAGCAAACGGCGCTCTCGGTATGATTGAGCTAAACGATGATTACACACT
GAAAAAAGTGATGAAACCGCTGATTGCATCTAACACAGTAACAGATGAAATTGAACG
CGCGAACGTCTTTAAAATGAACGGCAAATGGTACCTGTTCACTGACTCCCGCGGAT
CAAAAATGACGATTGACGGCATTACGTCTAACGATATTTACATGCTTGGTTATGTTT
CTAATTCTTTAACTGGCCCATACAAGCCGCTGAACAAAACTGGCCTTGTGTTAAAAA
TGGATCTTGATCCTAACGATGTAACCTTTACTTACTCACACTTCGCTGTACCTCAAG
CGAAAGGAAACAATGTCGTGATTACAAGCTATATGACAAACAGAGGATTCTACGCA
GACAAACAATCAACGTTTGCGCCTAGCTTCCTGCTGAACATCAAAGGCAAGAAAAC
ATCTGTTGTCAAAGACAGCATCCTTGAACAAGGACAATTAACAGTTAACAAATAAAA
ACGCAAAAGAAAATGCCGATATCCTATTGGCATTGACGTCAGGTGGCACTTTTCGG
TCTCGTGATGACGAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGA
GATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTG
AGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG
CTGAAACCTCAGGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAA
ACCGGTAAACCAGCAATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGA
CGACCGGGTCGAATTTGCTTTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATT
CAGGCGTAGCACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATTACGCC

CCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATG
GAAGCCATCACAGACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTG
TCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCAT
ATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGA
AAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACG
CCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCC
AGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACA
CTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGATGAGCA
TTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTC
TTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT
TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCA
ACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATC
TCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGG
AACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTC
CCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACA
GGTATTTATTCGGCGCAAAGTGCGTCGGGTGATGCTGCCAACTTACTGATTTAGTG
TATGATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTATCAGCTGTCCCTCCT
GTTCAGCTACTGACGGGGTGGTGCGTAACGGCAAAAGCACCGCCGGACATCAGCG
CTAGCGGAGTGTATACTGGCTTACTATGTTGGCACTGATGAGGGTGTCAGTGAAGT
GCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATATGTGAT
ACAGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTG
CGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGG
AAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCT
CCGCCCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAAC
CCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCT
CTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATTCCGCTGTTATGGCCGGTTTGTC
TCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGT
ATGCACGAACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCT
TGAGTCCAACCCGGAAAGACATGCAAAAGCACCACTGGCAGCAGCCACTGGTAATT
GATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAAACTGAAAGGAC
AAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAG
CTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTTTTCAGAGCAA
GAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATCAGATAAAA
TATTTCTAGATTTCAGTGCAATTTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCA
TACGATATAAGTTGTAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTGCATGCA
TCCATCAAAAAAATATTGACAACATAAAAAACTTTGTGTTATACTTGTAACGCTACAT
GGAGATTAACTCAATCTAGCTAGAGAGGCTTTACACTTTATGCTTCCGGCTCGTATA
ATGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGA
TTACGGATTCACTGGAACTCTAGACCAAAGAGAGGACACAATGCAGGGTTCTGTGA
CAGAGTTTCTAAAACCGCGCCTGGTTGATATCGAGCAAGTGAGTTCGACGCACGCC
AAGGTGACCCTTGAGCCTTTAGAGCGTGGCTTTGGCCATACTCTGGGTAACGCACT
GCGCCGTATTCTGCTCTCATCGATGCCGGGTTGCGCGGTGACCGAGGTTGAGATT
GATGGTGTACTACATGAGTACAGCACCAAAGAAGGCGTTCAGGAAGATATCCTGGA
AATCCTGCTCAACCTGAAAGGGCTGGCGGTGAGAGTTCAGGGCAAAGATGAAGTT
ATTCTTACCTTGAATAAATCTGGCATTGGCCCTGTGACTGCAGCCGATATCACCCAC
GACGGTGATGTCGAAATCGTCAAGCCGCAGCACGTGATCTGCCACCTGACCGATG
AGAACGCGTCTATTAGCATGCGTATCAAAGTTCAGCGCGGTCGTGGTTATGTGCCG
GCTTCTACCCGAATTCATTCGGAAGAAGATGAGCGCCCAATCGGCCGTCTGCTGGT
CGACGCATGCTACAGCCCTGTGGAGCGTATTGCCTACAATGTTGAAGCAGCGCGT
GTAGAACAGCGTACCGACCTGGACAAGCTGGTCATCGAAATGGAAACCAACGGCA
CAATCGATCCTGAAGAGGCGATTCGTCGTGCGGCAACCATTCTGGCTGAACAACTG

G

This sequence incorporates regions of vector pACYC-LIC+ (SGC) that remain unconfirmed.

Coding sequence for the N-terminal tag
BsaI recognition sites
SacB cassette
Stop codons

CGCCGCTAGGTGTGCGAGTTCAAGTCTCGCCTCCCGCACCATTCACCAGAAAGCGTTGATC
GGATGCCCTCGAGTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGCATGATCGTGCTCCT
GTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAGCAGAATGAATCACCG
ATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCGACCTGAGCAACAACAT
GAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGGAAGTCCCCTACGTGCTGC
TGAAGTTGCCCGCAACAGAGAGTGGAACCAACCGGTGATACCACGATACTATGACTGAGAG
TCAACGCCATGAGCGGCCTCATTTCTTATTCTGAGTTACAACAGTCCGCACCGCTGCCGGTA
GCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTAT
CATGCAACTCGTAGGACAGGTGCCGGCAGCGCCCAACAGTCCCCCGGCCACGGGGCCTGC
CACCATACCCACGCCGAAACAAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGA
TGCTGCTGGCTACCCTGTGGAACACCTACATCTGTATTAACGAAGCGCTAACCGTTTTTATC
AGGCTCTGGGAGGCAGAATAAATGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTG
AGCAAACTGGCCTCAGGAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAG
GCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGC
GTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATC
GGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCAC
CAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAA
GCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGG
ATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCG
CAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAG
CATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGG
CACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGC
CAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATT
TGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGA
AAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTG
CAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCAC
TGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTC
TACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACA
ATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGT
TTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTT
CCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGT
CTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCA
CCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCG
ATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGT
AGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCG
CCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATG
AGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCA
ACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATC
TCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCT
CTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGTCCCCTATACTAGGTTATTG
GAAAATTAAGGGCCTTGTGCAACCCACTCGACTTCTTTTGGAATATCTTGAAGAAAAATATGA
AGAGCATTTGTATGAGCGCGATGAAGGTGATAAATGGCGAAACAAAAAGTTTGAATTGGGTT
TGGAGTTTCCCAATCTTCCTTATTATATTGATGGTGATGTTAAATTAACACAGTCTATGGCCAT
CATACGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTGTCCAAAAGAGCGTGCAGAGA
TTTCAATGCTTGAAGGAGCGGTTTTGGATATTAGATACGGTGTTTCGAGAATTGCATATAGTA
AAGACTTTGAAACTCTCAAAGTTGATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCG

AAGATCGTTTATGTCATAAAACATATTTAAATGGTGATCATGTAACCCATCCTGACTTCATGTT
GTATGACGCTCTTGATGTTGTTTTATACATGGACCCAATGTGCCTGGATGCGTTCCCAAAATT
AGTTTGTTTTAAAAAACGTATTGAAGCTATCCCACAAATTGATAAGTACTTGAAATCCAGCAA
GTATATAGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCA
AAATCGGATAGCAGCGGCAGAGAAAACTTGTATTTCCAGGGCAGAGACCCTGAAAGATCCA
TAACTTCGTATAGCATACATTATACGAAGTTATGCGGCCGCGACGTCCACATATACCTGCCG
TTCACTATTATTTAGTGAAATGAGATATTATGATATTTTCTGAATTGTGATTAAAAAGGCAACT
TTATGCCCATGCAACAGAAACTATAAAAAATACAGAGAATGAAAAGAAACAGATAGATTTTTT
AGTTCTTTAGGCCCGTAGTCTGCAAATCCTTTTATGATTTTCTATCAAACAAAAGAGGAAAAT
AGACCAGTTGCAATCCAAACGAGAGTCTAATAGAATGAGGTCGAAAAGTAAATCGCGCGGG
TTTGTTACTGATAAAGCAGGCAAGACCTAAAATGTGTAAAGGGCAAAGTGTATACTTTGGCG
TCACCCCTTACATATTTTAGGTCTTTTTTTATTGTGCGTAACTAACTTGCCATCTTCAAACAGG
AGGGCTGGAAGAAGCAGACCGCTAACACAGTACATAAAAAAGGAGACATGAACGATGAACA
TCAAAAAGTTTGCAAAACAAGCAACAGTATTAACCTTTACTACCGCACTGCTGGCAGGAGGC
GCAACTCAAGCGTTTGCGAAAGAAACGAACCAAAAGCCATATAAGGAAACATACGGCATTTC
CCATATTACACGCCATGATATGCTGCAAATCCCTGAACAGCAAAAAAATGAAAAATATCAAGT
TCCTGAGTTCGATTCGTCCACAATTAAAAATATCTCTTCTGCAAAAGGCCTGGACGTTTGGGA
CAGCTGGCCATTACAAAACGCTGACGGCACTGTCGCAAACTATCACGGCTACCACATCGTCT
TTGCATTAGCCGGAGATCCTAAAAATGCGGATGACACATCGATTTACATGTTCTATCAAAAAG
TCGGCGAAACTTCTATTGACAGCTGGAAAAACGCTGGCCGCGTCTTTAAAGACAGCGACAAA
TTCGATGCAAATGATTCTATCCTAAAAGACCAAACACAAGAATGGTCAGGTTCAGCCACATTT
ACATCTGACGGAAAAATCCGTTTATTCTACACTGATTTCTCCGGTAAACATTACGGCAAACAA
ACACTGACAACTGCACAAGTTAACGTATCAGCATCAGACAGCTCTTTGAACATCAACGGTGT
AGAGGATTATAAATCAATCTTTGACGGTGACGGAAAAACGTATCAAAATGTACAGCAGTTCAT
CGATGAAGGCAACTACAGCTCAGGCGACAACCATACGCTGAGAGATCCTCACTACGTAGAA
GATAAAGGCCACAAATACTTAGTATTTGAAGCAAACACTGGAACTGAAGATGGCTACCAAGG
CGAAGAATCTTTATTTAACAAAGCATACTATGGCAAAAGCACATCATTCTTCCGTCAAGAAAG
TCAAAAACTTCTGCAAAGCGATAAAAAACGCACGGCTGAGTTAGCAAACGGCGCTCTCGGTA
TGATTGAGCTAAACGATGATTACACACTGAAAAAAGTGATGAAACCGCTGATTGCATCTAACA
CAGTAACAGATGAAATTGAACGCGCGAACGTCTTTAAAATGAACGGCAAATGGTACCTGTTC
ACTGACTCCCGCGGATCAAAAATGACGATTGACGGCATTACGTCTAACGATATTTACATGCT
TGGTTATGTTTCTAATTCTTTAACTGGCCCATACAAGCCGCTGAACAAAACTGGCCTTGTGTT
AAAAATGGATCTTGATCCTAACGATGTAACCTTTACTTACTCACACTTCGCTGTACCTCAAGC
GAAAGGAAACAATGTCGTGATTACAAGCTATATGACAAACAGAGGATTCTACGCAGACAAAC
AATCAACGTTTGCGCCTAGCTTCCTGCTGAACATCAAAGGCAAGAAAACATCTGTTGTCAAA
GACAGCATCCTTGAACAAGGACAATTAACAGTTAACAAATAAAAACGCAAAAGAAAATGCCG
ATATCCTATTGGCATTGACGTCAGGTGGCACTTTTCGGTCTCGTGATGACGAAGCTTGCGGC
CGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAA
GCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAAC
GGGTCTTGAGGGGTTTTTTGCTGAAACCTCAGGCATTTGAGAAGCACACGGTCACACTGCTT
CCGGTAGTCAATAAACCGGTAAACCAGCAATAGACATAAGCGGCTATTTAACGACCCTGCCC
TGAACCGACGACCGGGTCGAATTTGCTTTCGAATTTCTGCCATTCATCCGCTTATTATCACTT
ATTCAGGCGTAGCACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCG
CCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATC
ACAGACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAA
TATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAA
ACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAG
GGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGC
CGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAAC
GGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGA
ATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTG
CTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGT
ACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAAC
GGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAAC
TCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGC
CGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACAC

CAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCGCAAAGTGCGT
CGGGTGATGCTGCCAACTTACTGATTTAGTGTATGATGGTGTTTTTGAGGTGCTCCAGTGGC
TTCTGTTTCTATCAGCTGTCCCTCCTGTTCAGCTACTGACGGGGTGGTGCGTAACGGCAAAA
GCACCGCCGGACATCAGCGCTAGCGGAGTGTATACTGGCTTACTATGTTGGCACTGATGAG
GGTGTCAGTGAAGTGCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAA
TATGTGATACAGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGAC
TGCGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGGAAGA
TACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCCCCCT
GACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAAA
GATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGTT
TACCGGTGTCATTCCGCTGTTATGGCCGGTTTGTCTCATTCCACGCCTGACACTCAGTTCCG
GGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCACGAACCCCCCGTTCAGTCCGACCGCTG
CGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAAGACATGCAAAAGCACCACTGG
CAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTA
AACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGT
TGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTTTTCAGAGCAA
GAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATCAGATAAAATATTTC
TAGATTTCAGTGCAATTTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAG
TTGTAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTGCATGCATCCATCAAAAAAATATT
GACAACATAAAAAACTTTGTGTTATACTTGTAACGCTACATGGAGATTAACTCAATCTAGCTA
GAGAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAA
TTTCACACAGGAAACAGCTATGACCATGATTACGGATTCACTGGAACTCTAGACCAAAGAGA
GGACACAATGCAGGGTTCTGTGACAGAGTTTCTAAAACCGCGCCTGGTTGATATCGAGCAA
GTGAGTTCGACGCACGCCAAGGTGACCCTTGAGCCTTTAGAGCGTGGCTTTGGCCATACTC
TGGGTAACGCACTGCGCCGTATTCTGCTCTCATCGATGCCGGGTTGCGCGGTGACCGAGGT
TGAGATTGATGGTGTACTACATGAGTACAGCACCAAAGAAGGCGTTCAGGAAGATATCCTGG
AAATCCTGCTCAACCTGAAAGGGCTGGCGGTGAGAGTTCAGGGCAAAGATGAAGTTATTCTT
ACCTTGAATAAATCTGGCATTGGCCCTGTGACTGCAGCCGATATCACCCACGACGGTGATGT
CGAAATCGTCAAGCCGCAGCACGTGATCTGCCACCTGACCGATGAGAACGCGTCTATTAGC
ATGCGTATCAAAGTTCAGCGCGGTCGTGGTTATGTGCCGGCTTCTACCCGAATTCATTCGGA
AGAAGATGAGCGCCCAATCGGCCGTCTGCTGGTCGACGCATGCTACAGCCCTGTGGAGCG
TATTGCCTACAATGTTGAAGCAGCGCGTGTAGAACAGCGTACCGACCTGGACAAGCTGGTC
ATCGAAATGGAAACCAACGGCACAATCGATCCTGAAGAGGCGATTCGTCGTGCGGCAACCA
TTCTGGCTGAACAACTGG

This sequence uses Peter's GeneTool pACYC-LIC+ sequence, which has regions of unconfirmed sequence.