**Howarth lab sequence validation**

Your name: Anthony Keeble

Gene: MGSS-SpyTag3-sfGFP

Also known as MGSS-ST3-sfGFP

Plasmid (including clone no.): pET28 MGSS SpyTag003-sfGFP clone1

Purpose: Expresses SpyTag003-sfGFP in bacterial cytoplasm. The protein has a modified N-terminus such that the soluble protein has the residues GSS between the N-terminal methionine and the first residue of SpyTag003 in order to aid improved expression. The SpyTag003 allows the protein to bind and react with SpyCatcher proteins. The sfGFP allows the protein to be identified at low protein concentrations on SDS-PAGE if the protein is only heated to 50 oC in loading buffer.

Date: 2022-08-24

Predicted nucleotide sequence (from ATG to stop codon):

**atgggcagcagc**cgtggcgttcctcatattgttatggtggacgcctacaaacgctataaatcgggaggtggttcaggc**cgtaaaggcgaagagctgttcactggtgtcgtccctattctggtggaactggatggtgatgtcaacggtcataagttttccgtgcgtggcgagggtgaaggtgacgcaactaatggtaaactgacgctgaagttcatctgtact**actggtaaactgccggtaccttggccgactctggtaacgacgctgacttatggtgttcagtgctttgctcgttatccggaccatatgaagcagcatgacttcttcaagtccgccatgccggaaggctatgtgcaggaacgcacgatttcctttaaggatgacggcacgtacaaaacgcgtgcggaagtgaaatttgaaggcgataccctggtaaaccgcattgagctgaaaggcattgactttaaagaagacggcaatatcctgggccataagctggaatacaattttaacagccacaatgtttacatcaccgccgataaacaaaaaaatggcattaaagcgaattttaaaattcgccacaacgtggaggatggcagcgtgcagctggctgatcactaccagcaaaacactccaatcggtgatggtcctgttctgctgccagacaatcactatctgagcacgcaaagcgttctgtctaaagatccgaacgagaaacgcgatcatatggttctgctggagttcgtaaccgcagcgggcatcacgcatggtatggatgaactgtacaaaggctccggctccggctcc**caccatcaccatcaccat**tgataa

No. of nucleotides (including stop codon): 831

Predicted amino acid sequence from translating the DNA sequence above:

**MGSS**RGVPHIVMVDAYKRYKSGGGSG**RKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYK**GSGSGS**HHHHHH\*\***

**MGSS** initial sequence N-terminal to SpyTag003

SpyTag003

**sfGFP**

**His6-tag**

# ProtParam

## User-provided sequence:

10 20 30 40 50 60   
GSSRGVPHIV MVDAYKRYKS GGGSGRKGEE LFTGVVPILV ELDGDVNGHK FSVRGEGEGD   
  
 70 80 90 100 110 120   
ATNGKLTLKF ICTTGKLPVP WPTLVTTLTY GVQCFARYPD HMKQHDFFKS AMPEGYVQER   
  
 130 140 150 160 170 180   
TISFKDDGTY KTRAEVKFEG DTLVNRIELK GIDFKEDGNI LGHKLEYNFN SHNVYITADK   
  
 190 200 210 220 230 240   
QKNGIKANFK IRHNVEDGSV QLADHYQQNT PIGDGPVLLP DNHYLSTQSV LSKDPNEKRD   
  
 250 260 270   
HMVLLEFVTA AGITHGMDEL YKGSGSGSHH HHHH

[References](https://web.expasy.org/protparam/protpar-ref.html) and [documentation](https://web.expasy.org/protparam/protparam-doc.html) are available.

**Number of amino acids:** 274

**Molecular weight:** 30478.19

**Theoretical pI:** 6.65

Top of Form

**Amino acid composition:**   
Ala (A) 10 3.6%

Arg (R) 10 3.6%

Asn (N) 13 4.7%

Asp (D) 19 6.9%

Cys (C) 2 0.7%

Gln (Q) 8 2.9%

Glu (E) 16 5.8%

Gly (G) 31 11.3%

His (H) 17 6.2%

Ile (I) 12 4.4%

Leu (L) 20 7.3%

Lys (K) 22 8.0%

Met (M) 5 1.8%

Phe (F) 12 4.4%

Pro (P) 11 4.0%

Ser (S) 15 5.5%

Thr (T) 18 6.6%

Trp (W) 1 0.4%

Tyr (Y) 11 4.0%

Val (V) 21 7.7%

Pyl (O) 0 0.0%

Sec (U) 0 0.0%

(B) 0 0.0%

(Z) 0 0.0%

(X) 0 0.0%

Bottom of Form

**Total number of negatively charged residues (Asp + Glu):** 35

**Total number of positively charged residues (Arg + Lys):** 32

**Atomic composition:**

Carbon C 1352

Hydrogen H 2088

Nitrogen N 382

Oxygen O 410

Sulfur S 7

**Formula:** C1352H2088N382O410S7

**Total number of atoms:** 4239

**Extinction coefficients:**

Extinction coefficients are in units of M-1 cm-1, at 280 nm measured in water.

Ext. coefficient 22015

Abs 0.1% (=1 g/l) 0.722, assuming all pairs of Cys residues form cystines

Ext. coefficient 21890

Abs 0.1% (=1 g/l) 0.718, assuming all Cys residues are reduced

**Estimated half-life:**

The N-terminal of the sequence considered is G (Gly).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).

>20 hours (yeast, in vivo).

>10 hours (Escherichia coli, in vivo).

**Instability index:**

The instability index (II) is computed to be 24.60

This classifies the protein as stable.

**Aliphatic index:** 71.42

**Grand average of hydropathicity (GRAVY):** -0.608

Primer: T7R

Date: 2022-08-20

Read: (txt file from Source)

NNNNNNNNNNCNNCTTTCGGGCTTTGTTAGCAGCCGGATCTCAGTGGTGCTAGCTTATCAATGGTGATGGTGATGGTGGGAGCCGGAGCCGGAGCCTTTGTACAGTTCATCCATACCATGCGTGATGCCCGCTGCGGTTACGAACTCCAGCAGAACCATATGATCGCGTTTCTCGTTCGGATCTTTAGACAGAACGCTTTGCGTGCTCAGATAGTGATTGTCTGGCAGCAGAACAGGACCATCACCGATTGGAGTGTTTTGCTGGTAGTGATCAGCCAGCTGCACGCTGCCATCCTCCACGTTGTGGCGAATTTTAAAATTCGCTTTAATGCCATTTTTTTGTTTATCGGCGGTGATGTAAACATTGTGGCTGTTAAAATTGTATTCCAGCTTATGGCCCAGGATATTGCCGTCTTCTTTAAAGTCAATGCCTTTCAGCTCAATGCGGTTTACCAGGGTATCGCCTTCAAATTTCACTTCCGCACGCGTTTTGTACGTGCCGTCATCCTTAAAGGAAATCGTGCGTTCCTGCACATAGCCTTCCGGCATGGCGGACTTGAAGAAGTCATGCTGCTTCATATGGTCCGGATAACGAGCAAAGCACTGAACACCATAAGTCAGCGTCGTTACCAGAGTCGGCCAAGGTACCGGCAGTTTACCAGTAGTACAGATGAACTTCAGCGTCAGTTTACCATTAGTTGCGTCACCTTCACCCTCGCCACGCACGGAAAACTTATGACCGTTGACATCACCATCCAGTTCCACCAGAATAGGGACGACACCAGTGAACAGCTCTTCGCCTTTACGGCCTGAACCACCTCCCGATTTATAGCGTTTGTAGGCGTCCACCATAACAATATGAGGAACGCCACGGCTGCTGCCCATATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTCTAGAGGGGAATTGTTATCCGCTCACAATTCCCCTATAGTGAGTCGTATTAATTTCGCGGGATCGAGATCTCGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAACGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCNNGGCCGGGGGAATGTTGGGCGCCATCTCCTTGCATGNNCATTCCTTGGGGNNGGGGGGNNNCAANGGGCTCAAC

Alignment using BLAST (pairwise with dots for identities):

Query 1 ATGGGCAGCAGCCGTGGCGTTCCTCATATTGTTATGGTGGACGCCTACAAACGCTATAAA 60

Sbjct 885 ............................................................ 826

Query 61 TCGGGAGGTGGTTCAGGCCGTAAAGGCGAAGAGCTGTTCACTGGTGTCGTCCCTATTCTG 120

Sbjct 825 ............................................................ 766

Query 121 GTGGAACTGGATGGTGATGTCAACGGTCATAAGTTTTCCGTGCGTGGCGAGGGTGAAGGT 180

Sbjct 765 ............................................................ 706

Query 181 GACGCAACTAATGGTAAACTGACGCTGAAGTTCATCTGTACTACTGGTAAACTGCCGGTA 240

Sbjct 705 ............................................................ 646

Query 241 CCTTGGCCGACTCTGGTAACGACGCTGACTTATGGTGTTCAGTGCTTTGCTCGTTATCCG 300

Sbjct 645 ............................................................ 586

Query 301 GACCATATGAAGCAGCATGACTTCTTCAAGTCCGCCATGCCGGAAGGCTATGTGCAGGAA 360

Sbjct 585 ............................................................ 526

Query 361 CGCACGATTTCCTTTAAGGATGACGGCACGTACAAAACGCGTGCGGAAGTGAAATTTGAA 420

Sbjct 525 ............................................................ 466

Query 421 GGCGATACCCTGGTAAACCGCATTGAGCTGAAAGGCATTGACTTTAAAGAAGACGGCAAT 480

Sbjct 465 ............................................................ 406

Query 481 ATCCTGGGCCATAAGCTGGAATACAATTTTAACAGCCACAATGTTTACATCACCGCCGAT 540

Sbjct 405 ............................................................ 346

Query 541 AAACaaaaaaaTGGCATTAAAGCGAATTTTAAAATTCGCCACAACGTGGAGGATGGCAGC 600

Sbjct 345 ............................................................ 286

Query 601 GTGCAGCTGGCTGATCACTACCAGCAAAACACTCCAATCGGTGATGGTCCTGTTCTGCTG 660

Sbjct 285 ............................................................ 226

Query 661 CCAGACAATCACTATCTGAGCACGCAAAGCGTTCTGTCTAAAGATCCGAACGAGAAACGC 720

Sbjct 225 ............................................................ 166

Query 721 GATCATATGGTTCTGCTGGAGTTCGTAACCGCAGCGGGCATCACGCATGGTATGGATGAA 780

Sbjct 165 ............................................................ 106

Query 781 CTGTACAAAGGCTCCGGCTCCGGCTCCCACCATCACCATCACCATTGATAA 831

Sbjct 105 ................................................... 55

Image of abi of any nucleotides in question: none

(say what N you are claiming)

Nucleotides validated by this read:

1-831 (end)

Summary of all reads:

T7F gives 1-831 (end)

Complete gene covered