SUPPORTING INFORMATION

Characterization of the internal translation initiation region in monoclonal antibodies expressed in *Escherichia coli*

Erik M. Leith, William B. O'Dell, Na Ke, Colleen McClung, Mehmet Berkmen, Christina Bergonzo, Robert G. Brinson, Zvi Kelman

Supporting Experimental Procedures

Supporting Figure S1

Supporting Figure S2

Supporting Figure S3

Supporting Figure S4

Supporting Figure S5

Supporting Experimental Procedures

Normalization of fluorescence from plate reader assay

Cell density-normalized GFP fluorescence data (arbitrary units) for each sample were obtained by dividing the observed sample GFP fluorescence by the sample OD_{600} (blanked against uninoculated LB) at each time point and then subtracting cell density-normalized culture autofluorescence from a concurrent untransformed BL21(DE3) control culture at the same time point as shown in Eqn. S1:

$$\text{(S1)}\quad normalized\ fluorescence = \left[\frac{Fluorescence_{GFP,sample}}{\left(\frac{OD_{600,sample} - OD_{600,LB}}{b}\right)}\right] - \left[\frac{Fluorescence_{GFP,control}}{\left(\frac{OD_{600,control} - OD_{600,LB}}{b}\right)}\right]$$

The optical density path length of 1 mL culture in one well of a 24-well plate, *b*, was determined according to Eqn. S2 using the absorbance of uninoculated LB media at 977 nm and 900 nm and the constant 1 cm absorbance of pure water at 977 nm, 0.18 cm^{-1}:

$$\text{(S2)}\quad b = \frac{(A_{LB,977\ nm} - A_{LB,900\ nm})}{0.18\ cm^{-1}}$$

Supporting Figures

A) eNISTmAb

Full-length

DNA sequence:	AGC	AAT	ACC	AAA	GTG	GAT	AAA	CGT	GTT	GAA	CCG	AAA	AGC
Amino acid:	S_{210}	N_{211}	T_{212}	K_{213}	V_{214}	D_{215}	K_{216}	R_{217}	V_{218}	E_{219}	P_{220}	K_{221}	S_{222}

Truncated

DNA sequence:	AGC	AAT	ACC	AAA	GTG	GAT	AAA	CGT	GTT	GAA	CCG	AAA	AGC
Amino acid:					M_1	D_2	K_3	R_4	V_5	E_6	P_7	K_8	S_9

B) Adalimumab

Full-length

DNA sequence:	AGT	AAC	ACG	AAA	GTG	GAT	AAA	AAA	GTT	GAA	CCG	AAA	AGC
Amino acid:	S_{211}	N_{212}	T_{213}	K_{214}	V_{215}	D_{216}	K_{217}	K_{218}	V_{219}	E_{220}	P_{221}	K_{222}	S_{223}

Truncated

DNA sequence:	AGT	AAC	ACG	AAA	GTG	GAT	AAA	AAA	GTT	GAA	CCG	AAA	AGC
Amino acid:					M_1	D_2	K_3	K_4	V_5	E_6	P_7	K_8	S_9

Supporting Figure S1. Region of internal initiation in the heavy chain of eNISTmAb (A) and adalimumab (B). The full-length mAb include a Val (highlighted in blue) encoded by GTG (highlighted in yellow), while the truncated heavy chain includes a Met (highlighted in blue) at position 1 encoded by GTG (highlighted in yellow).

A) DNA sequence

```
              -85                                                              -40                                                 +1
eNISTmAb    GAGCAGCGTTGTTACCGTTCCGAGCAGCAGCCTGGGCACCCAGACCTATATTTGTAATGTTAATCATAAACCGAGCAATACCAAAGTGGA  90
adalimumab  GAGTTCCGTGGTTACCGTCCCGTCATCGAGCCTGGGCACCCAAACGTACATCTGCAACGTTAATCACAAACCGAGTAACACGAAAGTGGA  90

                                            +27                                                                              +93
eNISTmAb    TAAACGTGTTGAACCGAAAAGCTGCGATAAAACCCATACCTGTCCGCCTTGTCCGGCACCGGAACTGCTGGGTGGTCCGTCAGTTTTT  178
adalimumab  TAAAAAAGTTGAACCGAAAAGCTGTGACAAAACTCACACATGCCCACCGTGCCCAGCACCTGAACTCCTGGGGGGACCGTCAGTCTTC  178
```

B) Amino acid sequence

```
             186            201          214
eNISTmAb    SSVVTVPSSSLGTQTYICNVNHKPSNTKVDKRVEPKSCDKTHTCPPCPAPELLGGPSVF  59
adalimumab  SSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVF  59
             187            202          215
```

C) Percent identities between eNISTmAb and adalimumab

Construct	DNA percent identity	Amino acid percent identity
Full-length	78% (138/178)	98% (58/59)
Short	84% (64/76)	95% (21/22)

Supporting Figure S2. Alignments and identities of the DNA and amino acid sequences around the regions of internal initiation in the eNISTmAb and adalimumab.

A) DNA sequence alignment of the -85 to +93 region comprising the "full-length" constructs tested for internal initiation. Identical bases between the genes are highlighted in yellow, while mismatches are highlighted in gray. The internal initiation site is highlighted blue. The red bar covers the shortened region (-40, +27) identified that retains the same GFP fluorescence of the "full-length" eNISTmAb reporter construct. The DNA sequences shown were obtained by back-translation of the desired amino acid sequence followed by codon optimization for expression in *E. coli* by respective DNA synthesis vendors as described in "Methods".

B) Amino acid sequence alignment encoded by the DNA shown in A with similar color scheme.

C) Summary of DNA and amino acid identities between eNISTmAb and adalimumab in the region considered for characterizing internal initiation.

Bgl II

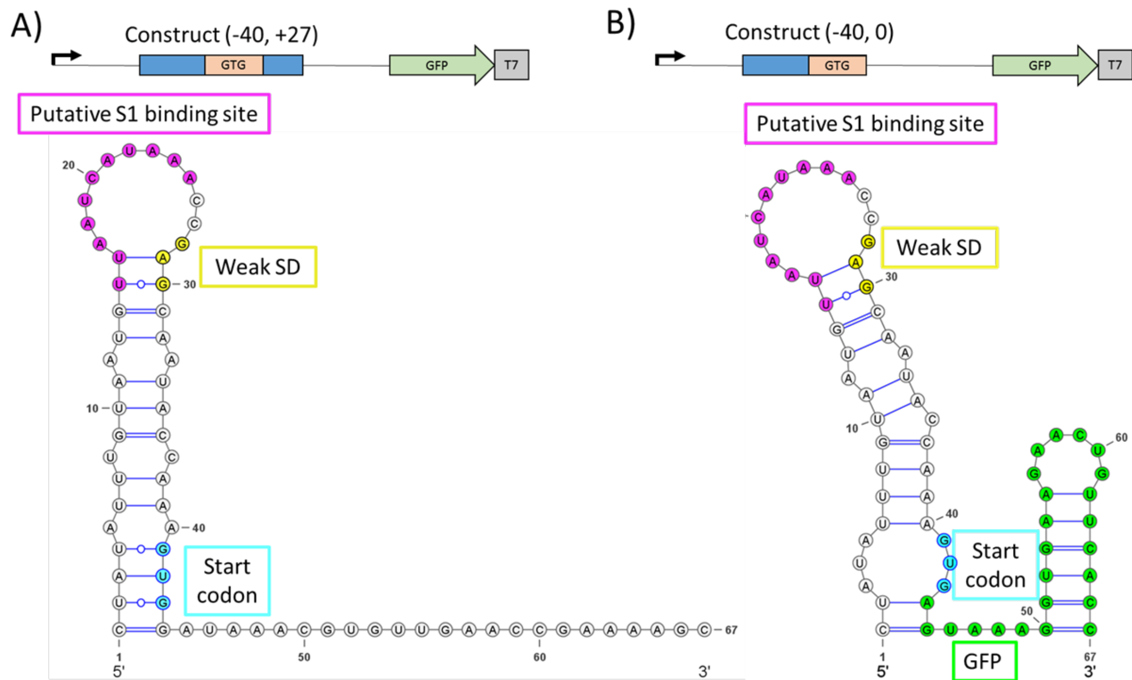
```
agatctatcccgcgaaattaatacgactcactatagggagcagcgttgttaccgttccgagca
gcagcctgggcacccagacctatatttgtaatgttaatcataaaccgagcaataccaaagtgga
taaacgtgttgaaccgaaaagctgcgataaaacccatacctgtccgccttgtccggcaccggaa
ctgctgggtggtccgtcagttttAGTAAAGGTGAAGAACTGTTCACCGGTGTTGTTCCGATCC
TGGTTGAACTGGATGGTGATGTTAACGGCCACAAATTCTCTGTTCGTGGTGAAGGTGAAGGTGA
TGCAACCAACGGTAAACTGACCCTGAAATTCATCTGCACTACCGGTAAACTGCCGGTTCCATGG
CCGACTCTGGTGACTACCCTGACCTATGGTGTTCAGTGTTTTTCTCGTTACCCGGATCACATGA
AGCAGCATGATTTCTTCAAATCTGCAATGCCGGAAGGTTATGTACAGGAGCGCACCATTTCTTT
CAAAGACGATGGCACCTACAAAACCCGTGCAGAGGTTAAATTTGAAGGTGATACTCTGGTGAAC
CGTATTGAACTGAAAGGCATTGATTTCAAAGAGGACGGCAACATCCTGGGCCACAAACTGGAAT
ATAACTTCAACTCCCATAACGTTTACATCACCGCAGACAAACAGAAGAACGGTATCAAAGCTAA
CTTCAAAATTCGCCATAACGTTGAAGACGGTAGCGTACAGCTGGCGGACCACTACCAGCAGAAC
ACTCCGATCGGTGATGGTCCGGTTCTGCTGCCGGATAACCACTACCTGTCCACCCAGTCTAAAC
TGTCCAAAGACCCGAACGAAAAGCGCGACCACATGGTGCTGCTGGAGTTCGTTACTGCAGCAGG
TATCACGCACGGCATGGATGAACTCTACAAATAATAAagttggctgctgccaccgctgagcaat
aactagcataaccccttggggcctctaaacgggtcttgaggggttttttgctcgag
```

Xho I

Red – restriction enzyme sites
Yellow – T7 promoter
Purple – sequences around the GTG initiation in the eNISTmAb
Green – GFP
Gray – T7 terminator
Blue – GTG putative initiation site, mutated to GTA, GTC, GTT and ATG

Supporting Figure S3. DNA sequence of the GFP reporter construct.

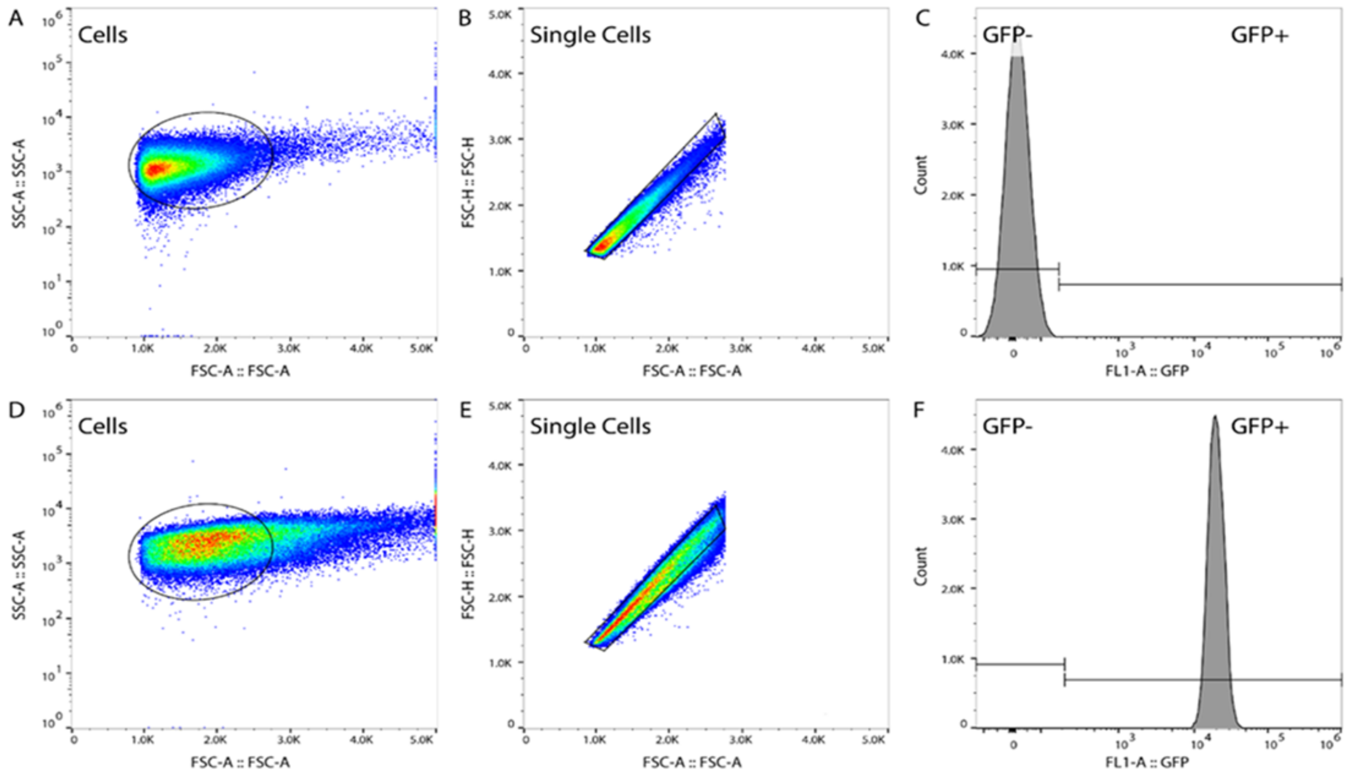
The purple region of the reporter construct is derived from the eNISTmAb expression vector with GTG as the initiator codon. The sequence is nearly identical to the full-length (-85,+93) construct I shown in Figure 4 but the GTG codon is replaced with ATG. The construct was subcloned into a pET-21a vector at the restriction sites indicated in red.



Supporting Figure S4. Predictions of secondary structure of the shortened (-40, +27) and the (-40, 0) sequences with GTG as the initiator codon.

A) Secondary structure prediction of the mRNA from shortened (-40, +27) sequence.

B) Secondary structure prediction of the mRNA from (-40, 0) sequence. Twenty-seven nucleotides of the GFP are also included and predicted to form a secondary structure.



Supporting Figure S5. Flow cytometry gating strategy.

Gates applied to all samples are shown for A-C: untransformed BL21(DE3) and D-F: ATG start codon construct. Intensity axes in histograms (C, F) are shown with biexponential scaling.