

ELECTRONIC SUPPLEMENTARY INFORMATION (ESI)

**Characterization of a Randomized FRET Library for Protease
Specificity Determination.**

Jonathan F. Fretwell¹, Shams M. K. Ismail², and Thomas L. Selby^{*1}

¹Department of Chemistry, The University of Central Florida, Orlando, Florida 32816

²Immunogenetics Department, Human Genetics and Genome Research Division,
National Research Center, Cairo, Egypt.

*Corresponding author address: Thomas L. Selby; 4000 Central Florida Blvd, Orlando,
Fl 32816-2366, Telephone: (407) 823-6752, FAX: (407) 823-2252, [tselby@mail.ucf.edu](mailto:t.selby@mail.ucf.edu)

Linker Sequence

TGTTCATGGAATTTGTGCTTTGCTCC
 GTATCATGATTAATTTCTCAAGAAAGA
 CCCGGGAGATTTAGTCAAATGACTCGG
 TTGATATTATCGGCGTTACGGATGTG
 TATGTACGAAAAAAGAAATGGTTTGT
 TTGACAACATTTTAGGAGGAGAAAGGC
 AATACAAAATTTCTAAGTATCTCAAGC
 CAAATAGGCAATTTCTAAACTTTTAAA
 AGCCCTACAGAAATACCAGAACCATT
 ACTAACCGGATACAGCTAATGGTGAAC
 AAGAACATTCCTCATTTGTATGGCAATA
 GGATCTGAAGAAGTCCAGGGTGGATG
 AGTCGTAATTTATAGCTTCCAGCTGCAG
 GGTGTACGGGGTGGTCTTATATGC
 ATCACTGTCTTGTCTGCCCTGAAAGAAG
 GTAGAATGGAGGGATTTGGTTCATCTAC
 TTGTGGACAGCACACAGGCAAGACTA
 TGGGCTATTCAATCCCTGCCCTGAAAT
 CGCGAGAGGTACCTACATTGCCACTTG
 GGCTATGGTCTATGTAAGGCCATCTCG
 CAATTTGTTAAGGAGGCAACCGGTACC
 GGCTATGGTCTATGTAAGGCCATCTCG
 CAGTACCCGATTGATTGTATATTTACC
 AAAGTCCATATATCTCGGCAGGTTTCAG
 GTTTGGACTAAGCAGCGCCATTTCTAT
 TTTCTAAATCCCGCGTCTCTTATCAT
 TTTCAATTCGCGCGGTTCTTATGTAA
 AATATTCGATTCCGGTTCAGCGACCAG
 GCTATAAAAAGCGTATCTGGGAGTCTT
 AGTTAAACACGCTACGAAACGATAGAAT
 CTCTTACCGAAGCAATTCACGCAAGA
 CGGACGAATTTAAGCCATAAAGACAGT
 ACTTGGTCTTGGGCGACAGATGTCCGG
 TAATTATTTAGCAGGTTCTGTCCATG
 TCTACGATGACGTCAACCGATACCCTT
 TGCTGGGTAAATCAAATAAATATGAT
 TGCTGGGTAAATCAAATAAATATGAT
 AGAAAAAAGTAAAGACTGTCCAGG
 ACATGTAACCGGTAGCTCATTTGCTAGG
 ACTATTACCGATTCTTATTTGCTGACT
 GCTTACGATACAGTTCTTACTTAAAGC
 ATTGTGTAATGCTAGACATCCGTTTG
 GAGTCACTATCCGAATCCTTTTAAAGA
 GATAGGGCTCTTACAGATATCTAACGAA
 GCCATTAACCAATATCACGCGATAACT
 GAGTCACTATTCGAATCCTTTTAAAGA
 TACAAATGGCCCATAGGAGTCAATTTGC
 TACTCTACGCTTTCCGGTGCCTCCATA
 CAAACGGGGTGTCTTAACTTGAAC
 GTTCTACAATTTGCTACGATACACGTG
 GGGAGACTTCTCAGGGAGGAGTAAA
 CCAGAACAGTCTAGCCGGGAGCTTAG
 GCCATAACTTTGGCCCTCCTAAGACTG
 ACGATCGTGTCTTTTAGTATCATAGTA
 CAAATAGGCAATTTCTAAACTTTTAAA
 TTAGCATCTTTTATATGTATTTGCCCG
 GCACCAACAGAACTACCTCACATTCGC
 CTAGGCGCTTGGACAGAACCGGAGCGT
 GCCCTATTTCCGGCATAATCAGAGATGA
 TTGGAACAACCTTGATCAGATCAACA
 GAGTGTATACTTGGTATTTGGCCTTGTG
 TAAAACCTTTATTTGTTTGCATTTACTC
 TTAAGAAGTAATCGCGTAAACAGACGAG
 ACCGCAGATCCGGTAAAGCAATCCAT
 TTTATGCGACGTTACGCCCTCTGATACA
 TGGCGAATCCAGAATAAGACGATATAA
 TAATCTGCCTCTGATTGTAACCTATAA
 CAGGCAGATGAATCAGAATGTCCACTC
 TGTAACTTAGAATGCGTCAACAAATAA
 CATATAGTGTCTTACACAATGTAGTA
 GGAGCAAAGCACAAATTCATGTAAACA
 CCGAGTCAATTTGACTAAATCTCCCGGG
 ACAAACTTCTTTTTCGATACATA
 GCTTGAAGATACCTAGAATTTGTATT
 TTTAAAAGTTAGAAATTTGCCATTTTG
 CATCCACCCTGGGACTTCTTCAGATCC
 GTAGATGACCCAATCCCTCCATTTCTAC
 ATTTACAGGCGAGGATTTGAATAGCCCA

Protein Sequence

CYMEFVLCS
VS*LI SQER
 PGRFSQMTR
 LILSAVTDV
 YVRKKRMVC
L TTF*EEKG
N TKF*VSSS
Q IGNF*TFK
 SPTEIPEPF
 TNAIQLMVN
 KNIPHCMAI
 GSEEVPGWM
 SRNYSFQLQ
 GVRGWSLIC
I TVLLP*KK
 VEWRDWVIY
 LWTAHEARL
W AIQSLP*N
 RERYLHCHL
 GYGLCKAIS
 QLFKEANGT
 GYGLCKAIS
 QYPIDCIFT
 KVHISRQVQ
 VWTQKRFHY
 FLNFPSPYH
F QFRAVLM*
 NIRFRFSDQ
 AIKSVSGSL
S *TRYER*N
 LLTEAIHAR
 RTNLSHKDS
 TWSWATRCR
***L FSRFLSM**
 STMTSTDTL
 CWVNQNKYD
 CWVNQNKYD
 RKSKDCPR
TCNA*LIAR
 TITDSYCLT
 AYDTVLYLS
 IVVMLDIRL
 ESLFRILLR
 DRALQISNE
 AINQYHAIT
 ESLFRILLR
 YNGPIGVIC
 YSTLSGASI
 QTGCFLTLN
 VLQFATHV
 GRLLREEDK
PEQSSPGA*
 AHNFGLLRL
 TIVLFSIIV
Q IGNF*TFK
 LASFMYCYP
 APTELPHIR
 LGAWTERER
ALFRHNQR*
 LETTLIRST
 ECILGIGLV
***NFICLHLL**
 LRSNRVTDE
 TARSGKQIH
 FMRRYASDT
W RIQNKTI*
SASDCNL
 QADESECP
CNLECVNK*
 HIVFLHNVV
 GAKHKFHV
 PSHLTKSPG
 TNHSFFSYI
A *RYLEFCI
 FKSLEIAYL
 HPPWDFFRS
 VDDPIPPFY
 ISGQGLNSP

Table 1. List of initial sequences in the randomized region from the first set of plasmids analyzed from the library. The corresponding amino acids for each region is also given based on the in-frame translation of the DNA in the cassette region (see Figure 2 in the text). This list was used to construct Table 1 in the main text. Stop codons are denoted using * and are colored in red to highlight the number of stop codons that were found in this initial set. Each of the colonies containing these plasmids turned green as described in the main text and as shown in Supplemental figure 1.

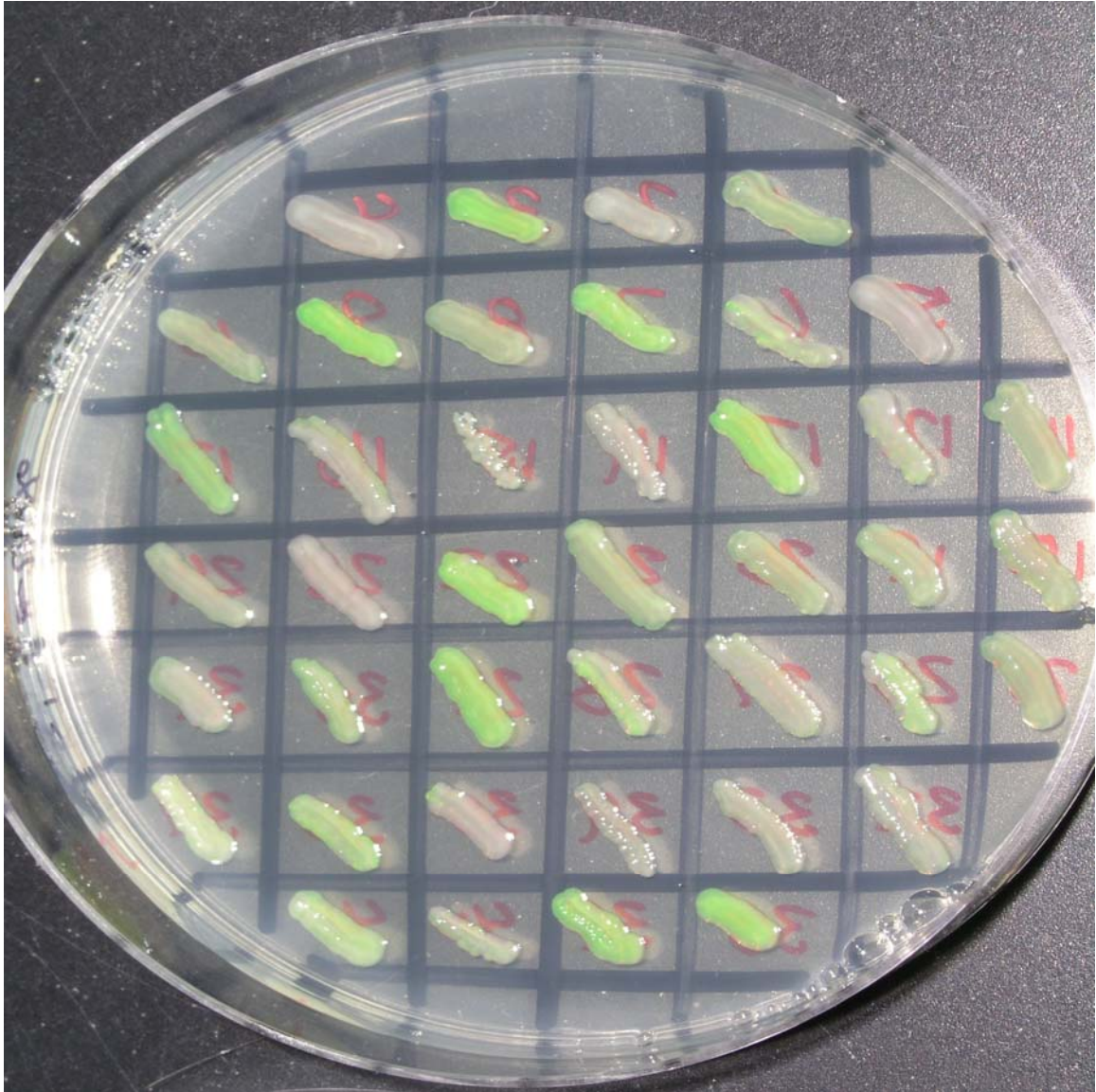


Figure 1. Example of the plate showing the green *E. coli* colonies for those cells that contain plasmids with stop codons in the randomized linker region. The cell line used was Novablue(DE3) and incubation was performed for two days, with day 1 at 37 °C and day 2 at room temperature. This image was captured with plain light (no UV irradiation) on a standard bench top with nothing added to the plate to increase the green color intensity. The darker green colonies, numbered from right to left, then down are # 3, 7, 9, 13, 17, 22, 29, 36, 38, and 39, which correspond to the first ten sequences shown in red in Supplementary Table 1 containing stop codons. Longer incubation periods will yield additional green colonies that turn a slight orange color after 1 week at room temperature. This screening step (initial green colony elimination) is now used to eliminate colonies that would not yield useful fusion proteins and are eliminated from the library.