SUPPLEMENTAL TABLE LEGENDS

Table S1. Compilation of useful websites.

Table S2. Compilation of all small proteins whose synthesis has been verified thus far. The table will periodically be updated at https://www.nichd.nih.gov/about/org/dir/affinity-groups/CSB/storz/data-protocols#RNAs. Please direct corrections to Gisela Storz at storzg@mail.nih.gov.

Column A = Protein name.
Column B = Alternative names.
Column C = Number of amino acids in protein.
Column D = Identified functions.
Column E = Left coordinate for small protein gene with respect to sense (clockwise or Watson strand) based on *E. coli* MG1655 genome version NC_000913.3.
Column F = Right coordinate for small protein gene with respect to sense (clockwise or Watson strand) based on *E. coli* MG1655 genome version NC_000913.3.
Column G = Orientation of gene with respect to sense (clockwise or Watson strand) based on *E. coli* MG1655 genome version NC_000913.3.
Column H = Adjacent genes. sORFs encoded within larger genes are noted, as well as their orientation relative to the larger gene.
Column I = Orientation of the sORF and adjacent genes. For the orientation, ">" corresponds to the sense (clockwise or Watson) strand and "<" corresponds to the antisense strand. The sORF arrow is in bold. ")" indicates the sORF overlaps with the adjacent gene. "[]" indicates the sORF is internal to a larger gene, with the sORF orientation being designated first and the larger gene orientation designated second.
Column J = Method by which small protein was detected.
Column K = Predicted transmembrane helix.
Column L = Localization determined.
Column M = Amino acid sequence. "*" corresponds to stop codon.
Column N = Nucleotide sequence.
Column O = Sequence of start codon (red) and 30 nucleotides upstream. Stretches of A and G residues of 4 or more (which could correspond to Shine Dalgarno sequences) located between 4 and 20 nucleotides upstream of the start codon are indicated in blue.
Column P = Reference for primary identification.
Column Q = PMID for primary identification.
Column R = Other relevant references.
Column S = PMID for other relevant references.

Table S3. Compilation of all sORFs for which expression has been tested by for which synthesis has not yet been detected as of 2019.

Column A = Name.
Column B = Number of predicted amino acids in protein.
Column C = Left coordinate for small protein gene with respect to sense (clockwise or Watson strand) based on *E. coli* MG1655 genome version NC_000913.3.
Column D = Right coordinate for small protein gene with respect to sense (clockwise or Watson strand) based on *E. coli* MG1655 genome version NC_000913.3.
Column E = Orientation of gene with respect to sense (clockwise or Watson strand) based on *E. coli* MG1655 genome version NC_000913.3.
Column F = Adjacent genes.
Column G = Orientation of adjacent genes.). For the orientation, ">" corresponds to the sense (clockwise or Watson) strand and "<" corresponds to the antisense strand.
Column H = Detection method attempted.
Column I = Predicted transmembrane helix.
Column J = Amino acid sequence. "*" corresponds to stop codon.
Column K = Nucleotide sequence.
Column L = Sequence of start codon (red) and 30 nucleotides upstream. Stretches of A and G residues of 4 or more (which could correspond to Shine Dalgarno sequences) located between 4 and 20 nucleotides upstream of the start codon are indicated in blue.
Column M = Original citation.
Column N = PMID for original citation.
Column O = Notes.