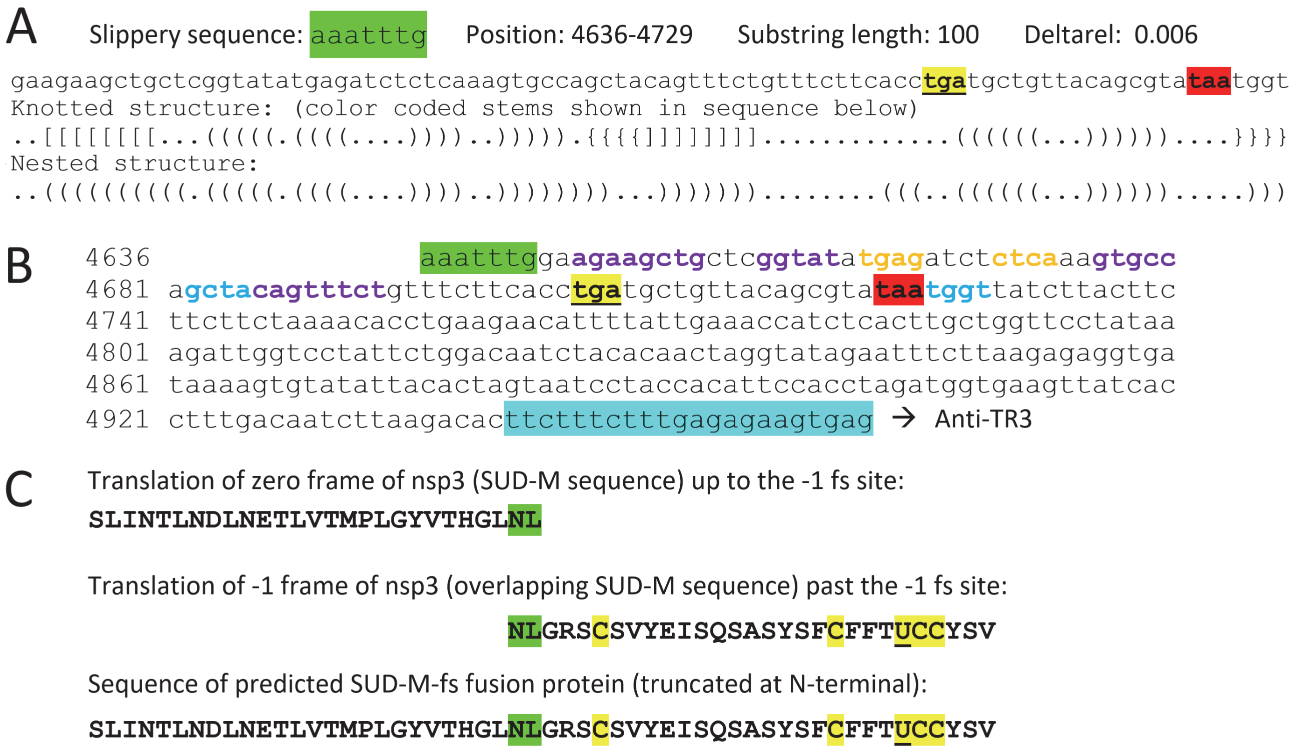
Supplementary Material

RNA viruses vs. DNA synthesis: a general viral strategy
that may contribute to the protective antiviral effects of selenium

Ethan Will Taylor



A Slippery sequence: aaatttg Position: 4636-4729 Substring length: 100 Deltarel: 0.006

```
gaagaagctgctcggtatatgagatctctcaaagtgccagctacagtttctgtttcttcacctgatgctgttacagcgtataatggt
Knotted structure: (color coded stems shown in sequence below)
..[[[[[[[[...(((((.((((....))))..))))).{{{{]]]]]]]].............((((((...))))))....}}}}
Nested structure:
..((((((((((.(((((.((((....))))..))))))))...)))))))........(((..((((((...))))))).....)))
```

B

```
4636                        aaatttggaagaagctgctcggtatatgagatctctcaaagtgcc
4681 agctacagtttctgtttcttcacctgatgctgttacagcgtataatggttatcttacttc
4741 ttcttctaaaacacctgaagaacattttattgaaaccatctcacttgctggttcctataa
4801 agattggtcctattctggacaatctacacaactaggtatagaatttcttaagagaggtga
4861 taaaagtgtatattacactagtaatcctaccacattccacctagatggtgaagttatcac
4921 ctttgacaatcttaagacacttctttctttgagagaagtgag → Anti-TR3
```

C Translation of zero frame of nsp3 (SUD-M sequence) up to the -1 fs site:

SLINTLNDLNETLVTMPLGYVTHGLNL

Translation of -1 frame of nsp3 (overlapping SUD-M sequence) past the -1 fs site:

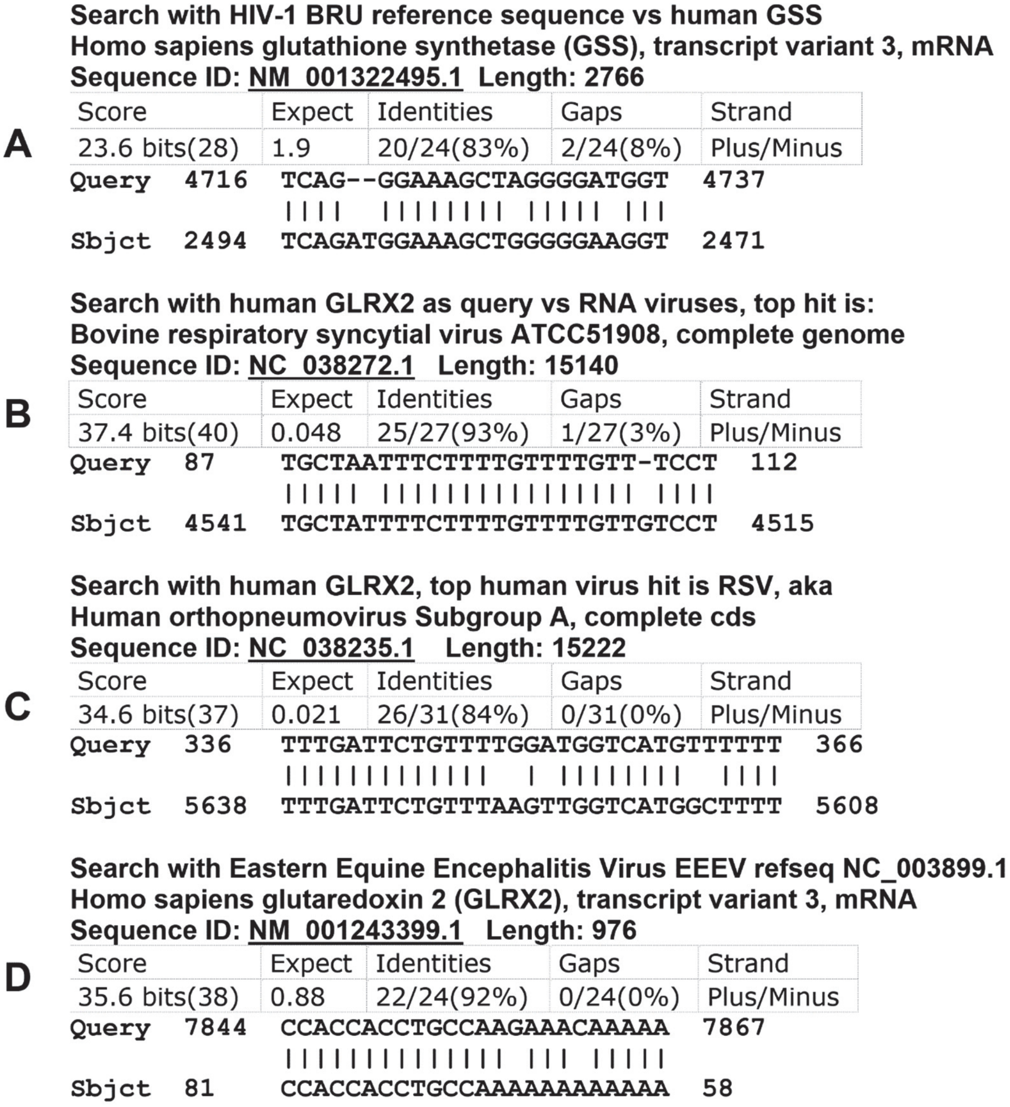
NLGRSCSVYEISQSASYSFCFFTUCCYSV

Sequence of predicted SUD-M-fs fusion protein (truncated at N-terminal):

SLINTLNDLNETLVTMPLGYVTHGLNLGRSCSVYEISQSASYSFCFFTUCCYSV

Supplementary Figure 1. A predicted -1 ribosomal frameshift site < 300 nucleotides upstream from the anti-TR3 site in SCoV2 shown as Figure 2A. (A) Elements of the -1 frameshift site as identified by the KnotInFrame* computer program, which include an ideal XXXYYYZ type "slippery sequence", aaatttg, and downstream RNA structures shown as knotted or nested structure. A single in-frame TGA codon (UGA in RNA) in the overlapping -1 reading frame is highlighted, as well as a TAA stop codon about 7 codons downstream. **(B)** A numbered portion of the SCoV2 genomic sequence (as DNA from the Genbank file NC_045512 for the SCoV2 reference sequence) is highlighted to show the frameshift site features, including color coded RNA complementary stem structures, and the downstream region that is an antisense match to the 3'-UTR of the TR3 mRNA (shown as main Figure **2A**). The complementary site in TR3 is only 150 bases from its SECIS element, which enables recoding of a UGA stop codon as selenocysteine. **(C)** Translations in single letter protein code of the corresponding regions of the nsp3 SUD-M coding region immediately upstream of the frameshift site, ending at the location of the slippery sequence, which encodes NL, then the sequence in the -1 reading frame following the frameshift, and the sequence of the predicted fusion protein that would be formed when frameshifting occurs (typically in only a few percent of translation events). The potential selenocysteine (U) and several Cys residues (C) are highlighted. The translatable overlapping sequence terminates after the valine (V) at the TAA stop codon highlighted in red in **(A)** and **(B)**.

* https://bibiserv.cebitec.uni-bielefeld.de/knotinframe, Stefan Janssen and Robert Giegerich, *The RNA shapes studio*, Bioinformatics, 2015, http://dx.doi.org/10.1093/bioinformatics/btu649



Search with HIV-1 BRU reference sequence vs human GSS
Homo sapiens glutathione synthetase (GSS), transcript variant 3, mRNA
Sequence ID: NM_001322495.1 Length: 2766

Score	Expect	Identities	Gaps	Strand
23.6 bits(28)	1.9	20/24(83%)	2/24(8%)	Plus/Minus

```
Query  4716  TCAG--GGAAAGCTAGGGGATGGT  4737
             ||||  |||||||| ||||| |||
Sbjct  2494  TCAGATGGAAAGCTGGGGGAAGGT  2471
```

Search with human GLRX2 as query vs RNA viruses, top hit is:
Bovine respiratory syncytial virus ATCC51908, complete genome
Sequence ID: NC_038272.1 Length: 15140

Score	Expect	Identities	Gaps	Strand
37.4 bits(40)	0.048	25/27(93%)	1/27(3%)	Plus/Minus

```
Query  87    TGCTAATTTCTTTTGTTTTGTT-TCCT  112
             ||||| |||||||||||||||| ||||
Sbjct  4541  TGCTATTTTCTTTTGTTTTGTTGTCCT  4515
```

Search with human GLRX2, top human virus hit is RSV, aka
Human orthopneumovirus Subgroup A, complete cds
Sequence ID: NC_038235.1 Length: 15222

Score	Expect	Identities	Gaps	Strand
34.6 bits(37)	0.021	26/31(84%)	0/31(0%)	Plus/Minus

```
Query  336   TTTGATTCTGTTTTGGATGGTCATGTTTTTT  366
             |||||||||||||  | ||||||||  ||||
Sbjct  5638  TTTGATTCTGTTTAAGTTGGTCATGGCTTTT  5608
```

Search with Eastern Equine Encephalitis Virus EEEV refseq NC_003899.1
Homo sapiens glutaredoxin 2 (GLRX2), transcript variant 3, mRNA
Sequence ID: NM_001243399.1 Length: 976

Score	Expect	Identities	Gaps	Strand
35.6 bits(38)	0.88	22/24(92%)	0/24(0%)	Plus/Minus

```
Query  7844  CCACCACCTGCCAAGAAACAAAAA  7867
             |||||||||||||| ||| |||||
Sbjct  81    CCACCACCTGCCAAAAAAAAAAAA  58
```

Supplementary Figure 2. Raw BLAST* search results showing sites of potential antisense interactions between RNA viruses and components of the glutaredoxin system. These sites have not been subject to any further analysis for potential RNA:RNA hybridization, internal folding energy etc. and thus are unvalidated. These search results show only identities between a plus and a minus strand, so the actual antisense RNA:RNA match would involve the complement of the bottom sequence of each pair. The method also fails to account for GU base pairs in RNA stem structures, so apparent mismatches such as those involving G and A near the middle of **(C)** and **(D)** would actually be GU base pairs in RNA. Individual sequence pairs **A-D** are referred to and discussed in the main text.

* https://blast.ncbi.nlm.nih.gov/Blast.cgi