Computational Analysis of RNA-Mediated Host-Virus interactions

Reyansh Gangal, Wilson Lee, Zhipeng Lu

Dept of Pharmacy and Pharmaceutical Sciences, Bridge Institute, University of Southern California, Los Angeles, CA , USA

USC University of Southern California

Bridge UnderGrad Science (BUGS) *Summer Research Program*

Abstract

RNA-mediated host-viruses are viruses that happen from an RNA viral strand attaching to the host's exon strands. This leads to the viruses taking over the gene expression of the host. Many RNA viruses can cause death or severe symptoms. RNA viruses include Ebola, HIV, COVID-19, and many others. Studying these interactions between the RNA viral strand and the host's exons help us figure out which strands in the host's body to target in vaccines and other medicine.

There are many difficulties that occur when studying RNA viral interactions. First, handling viruses is extremely difficult. Many research groups would not get authorization to handle a virus. Another problem is the amount of time these processes take. Studying RNA interactions manually is extremely time consuming, therefore limiting the amount of strands that can be analyzed. Because of the amount of time, and the restrictions of materials, our group used a software called IntaRNA.

IntaRNA simulates RNA strand interactions and displays the energy used in the interactions. This energy number can help us see which RNA strands to target when trying to create a vaccine or cure for the disease. From our research, we found the interactions that took the least energy for a disease's interaction with a human gene. This can help us compare different RNA strands to see which is most suitable for analysis. This research contributes to the field of virology, informing targeted therapies, and vaccine design.

Objectives

The objective of this research project is to simulate the interaction between a viral RNA strand and a human exon strand. By simulating this using IntaRNA we are trying to compare the minimal energy values to see which RNA strand to target when using vaccines.

Steps:

1. Find a Viral RNA sequence, and a human exon sequence

↓

2. Run the sequences using IntaRNA, and store them in a CSV file

↓

3. Graph the minimal energy values in the data

↓

4. analyze and compare the minimal energy values using the graph

RNA Strands

Protein-coding transcript sequences	CHR	• Nucleotide sequences of coding transcripts on the reference chromosomes • Transcript biotypes: protein_coding, nonsense_mediated_decay, non_stop_decay, IG_*_gene, TR_*_gene, polymorphic_pseudogene, protein_coding_LoF	Fasta

Acidianus filamentous virus 2, complete genome

NCBI Reference Sequence: NC_009884.1

GenBank   Graphics

```
>NC_009884.1 Acidianus filamentous virus 2, complete genome
GTATATATACGGTAAAATACCCAGACGATATTCTCCGTCAGTTTGAAGAGCAAGTAAAGGAACTGTTAAC
AGTTCCTTTACTTGACGAAAAGGAGTATTTCGCATTCCAAGAGATTAATAGATTTTTTTCTCTTTCACCC
TCAATTAAGAGGGTGAAAGATGAATGTAAGATCTATAGGAGGGATCGCTGGGGATTTTTTTACCCTAACG
GTTCATGCTAGTCTTTTTTATGCATCATTTTTTTCATATTCGTTATTATATTTCTATTTCATATTTATT
TTTAATTACCTTTTTTTATCAATTATTTTTCTTATTTTATTAATAAATTTTCTGCTATATATTTTGCATTTGG
CATATTATTTCTATTTTAACTCTTTTTCTACTAGAAATTTCTAGTAGAAAAAGGGGAGTTAGATAAGAGT
TGATGAGCGAAAGATTTTTATATGGGGAGAGAGAAGTATAATATAGAGGTGAAAAAAGTGGGAGAAGAAG
AAATTAAACAAGAGGAAAATGAAAAGAAAAAAATAGTAAAAAAATTAAATGAAAGTGATGTAAAAAGCAT
AATTGAGGATGTATTAAAAAACAAAGAAAAGTACGGGCTTGAAGCAAGGTGGTTGCTAGGGATGGATGAT
AACAAAAATGATAGTATTGAAGTTTTGTTGGGTAGAGTGAATTTTATTGAAGTGGAAAGAGGTTTTATTG
AAATAGTCCCGATGACACGTATGGTTGTGTTGCTGAAGAGAGTTAACTATTACACGTCTAACTATACATC
GTACAATGATGAAATTATATGTTTTCTTCTATTCAGTCGGCTGGTGAAATTTTTCTTAGCCTATTTTTTA
TTTTATTATTTTTTTTATTTTCGTTTTTATGTTCTATTTGTTAGAATGCCTATAAATCCCGCTTTTTTC
TTCTTCTAAAAAAATATTTTAATAAAAATAGTAGTTAATATTCAAAAACTTTTGTTCTATCATCTAAATC
GTACATAAATTTATAAACAGCACTAATTTTCCAATATAGCATTAAATTTGAAATTGACACACACCACAGA
TTATTTTTGTAATCTATACCGCATAATGTAGTAGCAAGTACTTCTTTTTCTGTCGTTACGATAGGATCTC
TCGTATCCACTGCAATTAGTTTGAATACTATCTGATTTACGCCGTCATAGTAATAATCTGTAATTATAAA
ATTGTCAATATAGCTAATTCCTTGTAAATTAACTCTATCTCTATTAAATCTGTAAATTGGCATTGTAAAA
TATTTCATATACTTTATATCTCTGAAATTCATTTTTTTATACAGCAATTTTTCTAATTTGAGTAATGGCA
TGATTCTGAATTCTCCTTGTTCAAGTATTCCTTTCCATCTAATTTTTGCAAACACATGTGTAATTTTT
TCTATATTTCAACATCATTTTTGCATATATTGGTATTCTCAATATTCTTCTCATAAAGTATTTTTTTCCT
TTGTTTCCAACTTTTTTGATGACTAGAGTTTTTCCATCTAATTTATTTAGTTTAGTTTGCAATTGCATTTT
TTATTTCATTTTCATTTTCCGTTAATTTCAATGAAATTTTAAATTCTTTAGGTCTTATTGGAATATAATA
TGTTACGTCATTATTTACAAATTTCTATTTCTATAGTATTTTGGTAAAAACGCATAACAGGATAAAATACC
TGGTCTAATATTTTCATTAGATCCCAGCACTGTATTTTTTTTAGCATCGTCTAAAGTTATTTCTATTTCAT
```

The top strand is a human exon sequence, while the bottom strand is a viral sequence. We simulated the interaction between these strands using IntaRNA.

CSV Results

Name of Sequence 1	Start position of Strand	End position of Strand	Name of Sequence 2	Start position of Strand	End position of Strand	Result of Interaction in Nucleotide format	Result of Interaction in hybrid format	Minimal Interaction Energy Required
id1	start1	end1	id2	start2	end2	subseqDP	hybridDP	E
ENST00000641!	401	764	lcl|NC_009884.1	1	357	AGAUAUUUCU	((((-((((-((-..(-(..	-132.9
ENST00000641!	382	1334	lcl|NC_009884.1	27	1032	UUCAAGGGCU	((((((((((...((((((-(.	-260.9
ENST00000641!	503	917	lcl|NC_009884.1	3	352	CAAUUAUGUG	((((-(((((.........(	-135.1
ENST00000641!	90	444	lcl|NC_009884.1	3	298	CUGGAAUGAA	((((.((......(((...	-84
ENST00000641!	399	768	lcl|NC_009884.1	2	315	UCAGAUAUUU	((((........((.((....((.(	-123.3
ENST00000641!	418	605	lcl|NC_009884.1	3	176	UUCUUUGGUG	(((((((.....(((((.(((	-74.3
ENST00000641!	507	764	lcl|NC_009884.1	1	249	UAUGUGUGGC	((..(((((.((......((((.	-89.8
ENST00000641!	395	735	lcl|NC_009884.1	4	315	UUGUUCAGAU	(((((((.....(.((-((	-118.3
ENST00000641!	412	882	lcl|NC_009884.1	25	423	CUUCACUUCU	((((((((.((((.(((...	-139.8
ENST00000641!	301	901	lcl|NC_009884.1	1	537	UCACUCAUUG	(((((((((.((((((.	-200.3
ENST00000641!	421	700	lcl|NC_009884.1	6	312	UUUGGUGGGA	((((((((((-(((((.(.((	-122.3
ENST00000641!	92	467	lcl|NC_009884.1	1	363	GGAAUGAAUC	(((((-(((((.(((.((((((.	-142.3
ENST00000641!	356	487	lcl|NC_009884.1	10	143	UUUUUCAGCCA	((((((((-(((-(.-((((((	-57.9
ENST00000641!	424	971	lcl|NC_009884.1	10	531	GGUGGGAGUC	(((((((.(((((.........(((	-197.5

These are the results of running IntaRNA into a CSV file. These columns represent which parts of each strand are interacting and what their minimal energy value is.

Minimal Energy Graph

Minimal Energy Values
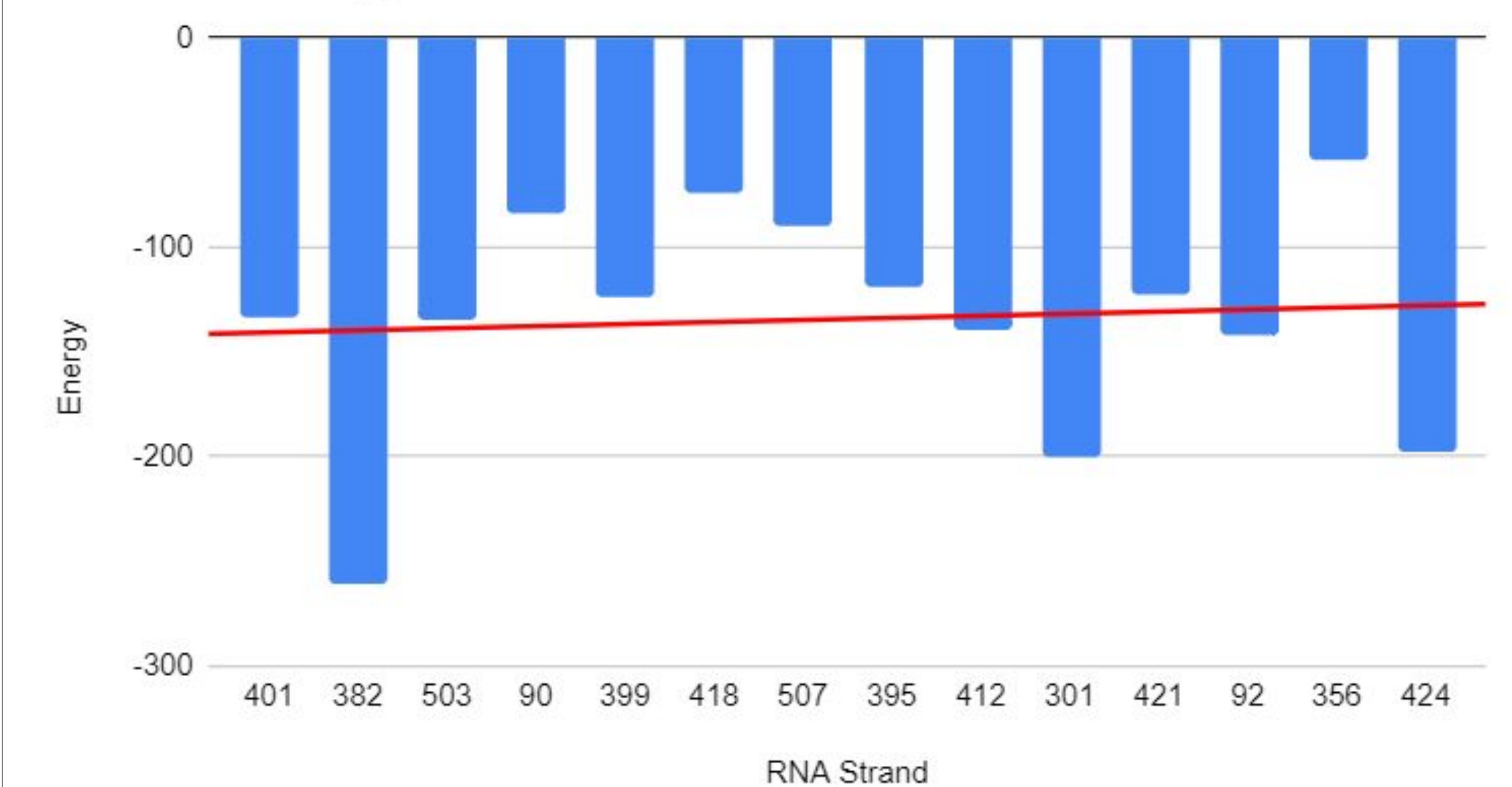
This is the graph with the RNA sequence start values at the bottom, and the Minimal Energy values on the side. This can help us see which RNA strands to target when studying the virus.

Summary

Following these steps for multiple viruses can help us figure out which RNA strands to target when creating medicine.

In this case with the interaction between the human exon strand and the Acidianus filamentous virus 2 strand, the RNA strands close to nucleotide 382 have the lowest minimal energy. This means it would be easiest to target that strand when making medicine.

References

- Martin Mann, Patrick R. Wright, and Rolf Backofen
IntaRNA 2.0: enhanced and customizable prediction of RNA–RNA interactions
Nucleic Acids Research, 2017, 45 (W1), W435–W439.
- Patrick R. Wright, Jens Georg, Martin Mann, Dragos A. Sorescu, Andreas S. Richter, Steffen Lott, Robert Kleinkauf, Wolfgang R. Hess, and Rolf Backofen
CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains
Nucleic Acids Research, 2014, 42 (W1), W119-W123.
- Anke Busch, Andreas S. Richter, and Rolf Backofen
IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions
Bioinformatics, 2008, 24 (24), 2849-56.
- Martin Raden, Syed M Ali, Omer S Alkhnbashi, Anke Busch, Fabrizio Costa, Jason A Davis, Florian Eggenhofer, Rick Gelhausen, Jens Georg, Steffen Heyne, Michael Hiller, Kousik Kundu, Robert Kleinkauf, Steffen C Lott, Mostafa M Mohamed, Alexander Mattheis, Milad Miladi, Andreas S Richter, Sebastian Will, Joachim Wolff, Patrick R Wright, and Rolf Backofen
Freiburg RNA tools: a central online resource for RNA-focused research and teaching
Nucleic Acids Research, 46(W1), W25-W29, 2018.

CONTACT US

bridge.usc.edu/bugs