Virus: Norovirus

<https://www.ncbi.nlm.nih.gov/protein/AKN44263>

Protein target: VPg

>YP\_009701442.1 VPg [Norovirus GI]

GKNKGKTKKGRGRKSNFNAFSRRGLSDEEYEEYKKIREEKSGNYSIQEYLEDRQRYEEELAEVQAGGDGG

IGETEAEIRHRVFYKSKSGMRKQRQEERRQLGLVSGSEIRKRKPIDWTPPKNDWSEDTRTVNYDEHISFE

For this discussion, I searched through the NCBI taxonomy browser for a virus containing a small protein of less than 125 residues in length. I found a VPg protein sequence from the Norovirus and I decided to run my analysis on it. Running the FASTA through PDB-Blast to see if there are any similar sequences in non-redundant databases and I found that the highest SeqID I found was 23%. Only the top match is shown below.

PDB-Blast SearchA screenshot of a social media post

Description automatically generated

My next step was to run my sequence in FFAS03 and Phyre2 for fold recognition.

Phyre2

**Top model**

A screenshot of a cell phone

Description automatically generated

**Secondary structure/disorder prediction**

A screenshot of a computer

Description automatically generated

**Domain Analysis**

A screenshot of a social media post

Description automatically generated

**All template/model information**

**A screenshot of a cell phone

Description automatically generated**

FFAS03

A screenshot of a cell phone

Description automatically generated

Phyre2 and FFAS03 have a consensus between the top three of both programs, albeit in a slightly different order. The top 3 consensus proteins are murine norovirus vpg protein, feline calicivirus vpg protein, and viral protein genome-linked. The consensus fold types are 2m4h, 2mxd, and 2m4g.

2m4h: Score = -27.3 (FFAS), 96.7% confidence (Phyre2), seq ID = 23% (FFAS), 33% (Phyre2), length of overlap: 3-67 (FFAS), 7-55 (Phyre2)

2mxd: Score = -25 (FFAS), 95.8% confidence (Phyre2), seq ID = 20% (FFAS), 29% (Phyre2), length of overlap:1-62 (FFAS), 25-55 (Phyre2)

2m4g: Score = -18.5 (FFAS), 100% confidence (Phyre2), seq ID = 50% (FFAS), 50% (Phyre2), length of overlap: 4-53 (FFAS), 18-101 (Phyre2)

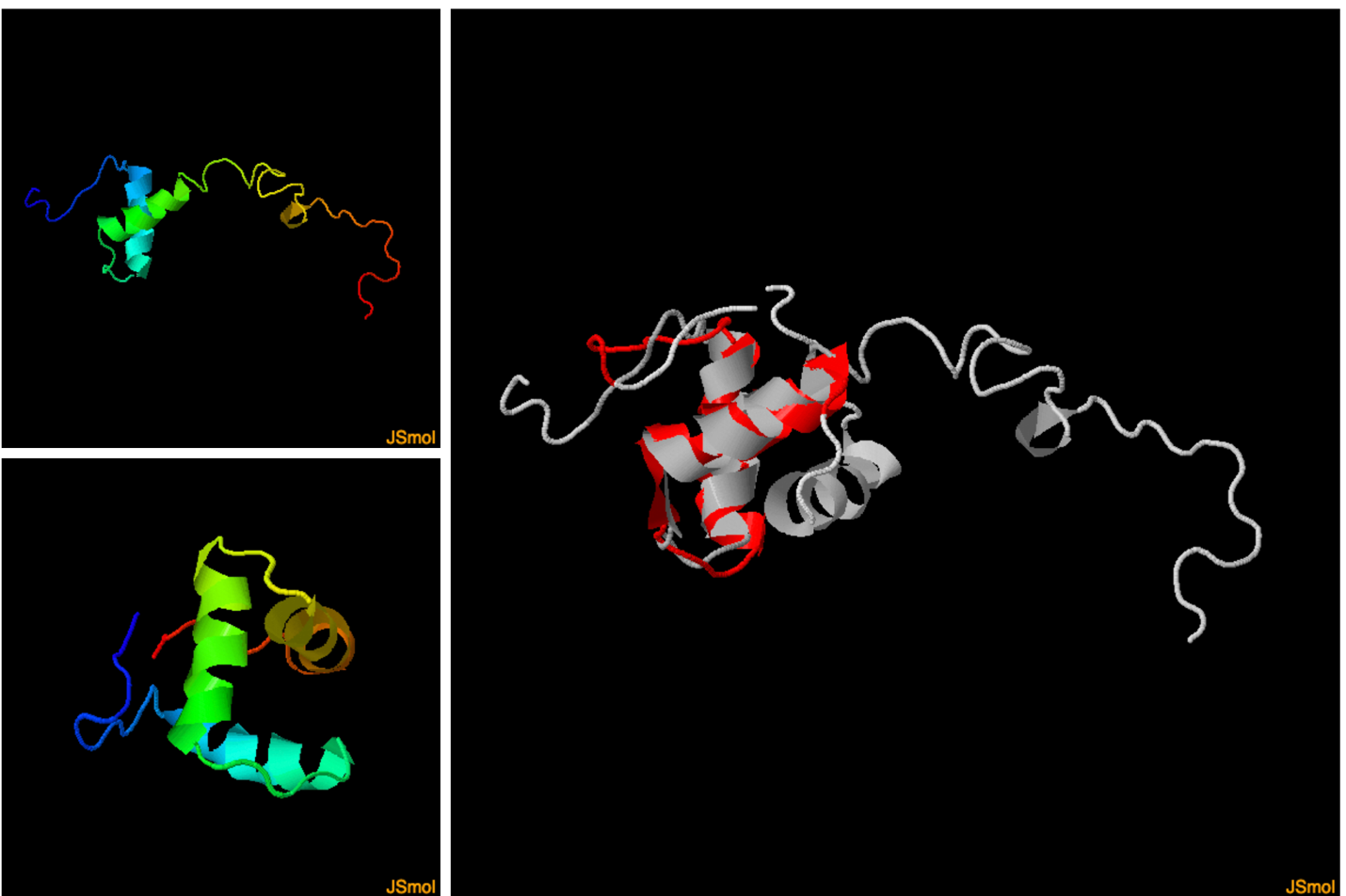
For I-TASSER

A screenshot of a cell phone

Description automatically generated

2m4h was missing from the top 10 hits from I-TASSER.

All three of the top consensus hits were missing from SCOP. From PDB: 2m4h is classified as viral protein, 2mxd is classified as a viral protein, and 2m4g is also classified as a viral protein.

For structural alignments, I chose to use 2m4g from I-TASSER and 2m4h from FFAS. I ran these through a flexible alignment using FATCAT and received the following output: RMSD = 2.66 and seq ID = 37.50% The top left image is 2m4hA, the bottom left is 2m4gA, and the right are the two models superimposed on each other.

Because this is a blind prediction, our alignments of these proteins give us a good starting point for further investigations. If I was interested in studying this particular protein, I would treat this information as a basis for further research on the properties and activities of the two related sequences above. The next step afterwards would be to gather empirical evidence using x-ray crystallography to determine the “actual” structure of this protein and continue investigations from there.