Secondary Structure of the M2 Region of Influenza Virus and Its Potential Effect on Mutation and Evolution

Morgan Elrick and Michael Sierk
Bioinformatics Program, Interdisciplinary Science Department
Saint Vincent College, Latrobe, PA

Abstract

Influenza is a virus that infects about 45 million people and kills up to 62,000 people in the United States every year (Centers for Disease Control and Prevention, 2009). Since the virus mutates frequently it is difficult to create a vaccine for it, and therefore it is important to continue to seek ways to combat the virus. Influenza’s genome is a single stranded, negative strand RNA sequence. Previous work had shown that the loops within the predicted secondary structure of the genomic RNA sequence in the hemagglutinin (HA) gene segment correlated with mutation rates in specific regions of the protein. We applied a similar approach to the M2 proton channel protein and the NS1 gene of the influenza virus. We created a multiple sequence alignment of 400 different virus strains from the NCBI Virus Database and examined the regions of the sequence alignment that corresponded to the predicted loops in the M2 and NS1 regions of the genome. Only three nucleotide positions had a level of conservation below 93% in the M2 region, and none were below 80% in the NS2 segment. It appears that the correlation between loop regions and mutations in the protein only apply to the HA protein and not to other proteins in the influenza virus.

Background: RNA Secondary Structure

The influenza virus is a single, negative stranded, RNA sequence (Bouvier & Palese, 2008). The secondary structure of the RNA is all to be important for the replication and regulation of the virus. The secondary structure is formed by the RNA base-pairing with itself, creating various stem-loop structures as shown in Figure 3 (Fouchier, 2019).

Methods

• Gather sequences from the NCBI Influenza Virus Database
• BLAST Search
• Found 400+ M2 sequences using a 75-95% similarity parameter search

Results

The M2 region of the M2 protein is seen in Figure 5. A predicted model of the secondary structure of the genomic RNA was created using the MFold software. MFold created multiple structures, the structure shown is the thermodynamically most stable.

Discussion

With regard to the M2 gene, we found little to no correlation between RNA secondary structure and mutation rate. Nearly all positions of the nucleotides in loop regions had very high conservation rates, with only three specific nucleotides having low conservation. Thus the initial hypothesis that the genomic RNA secondary structure affects the M2 protein appears to be incorrect. However, further study might show more subtle interaction between genomic RNA structure and protein mutation.

It’s possible that changing the sequences used would provide different results. It may be that the 400 sequences used were too similar to each other, and that a different set of sequences would show more variation. The M2 gene is not as heavily studied as the HA and NA genes, and thus there are fewer sequences available for this region of the genome. It may also be that the M2 protein is just more conserved than the surface proteins, since it has less exposure to the immune system. Lastly, without further experimental study we cannot be sure that the predicted secondary structure is correct.

Future Research

For future research, other areas of the influenza genome could be studied to see if similar trends are followed. There has been research done regarding the HA and NA regions (as discussed in the introduction), but there are plenty of other areas for research to observe. Another topic for future research could look at the three specific mutation points of the M2 region that we found. These points could have significant impact on the structure and function of the M2 protein. These areas could lead to significant advancements in understanding influenza and should be studied thoroughly.

References


