Identification and Modeling of Conserved Secondary Structures of Influenza A Virus Hemagglutinin Subtypes H1, H2, H3, H5

Bradley Brooks B.S.
Faculty Advisor: Audrey Shor, Ph.D., M.P.H.
Saint Leo University Department of Mathematics and Science, 33701 State Road 52, Saint Leo, FL 33574

Introduction

Influenza viruses cause lethal epidemics and pandemics that result in millions of cases of severe illness and hundreds of thousands of deaths annually. Hemagglutinin is a surface glycoprotein that plays a critical role in influenza A virus host recognition, attachment and entry. The HA1 peptide of hemagglutinin originates at the base of the structure and forms a globular region at the top of the protein, which contains the sialic acid receptor binding site needed for attachment to host cells. The globular head region of the protein is highly exposed and a major target of an immune response.

Objective

This study investigates both highly variable and invariable regions in the HA1 peptide of hemagglutinin (subtypes H1, H2, H3, H5) and aims to elucidate the significance of these regions in terms of overall structure and function.

Methods

- Sequences and crystallized structures corresponding to hemagglutinin subtypes H1, H2, H3, and H5 were acquired from the RCSB Protein Data Bank database (www.pdb.org).
- H2 was chosen to be the base sequence of comparison as it is no longer in circulation.
- A multiple sequence alignment of the four proteins was completed using UniProt, which utilizes Clustal Omega alignment program to generate alignment profiles.
- The crystallized structures obtained from the RCSB protein data bank database were used to model variations in Jmol and to overlay paired hemagglutinin subtypes in Magics. H3-H2 were further modeled into 3D structures highlighting conserved secondary structures and amino acids.
- A BLAST was performed to confirm presence of the conserved amino acids in other hemagglutinin protein sequences.

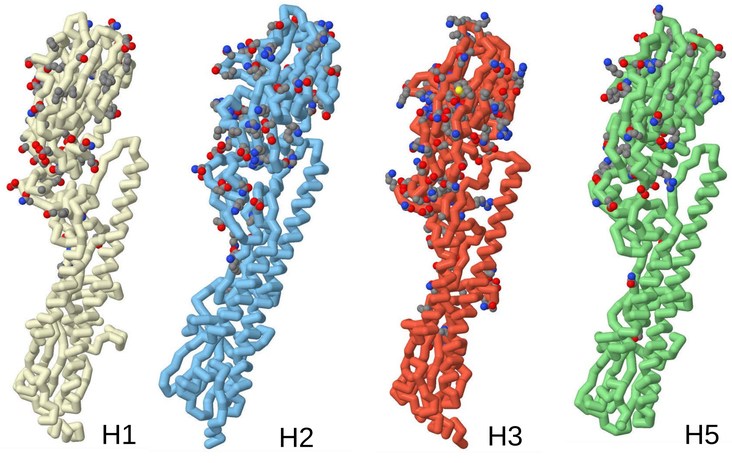

Fig 1. Crystallized structures used in the overlay. From left to right H1 (3ztn.pdb), H2 (2wrc.pdb), H3 (2yp7.pdb) and H5 (2ibx.pdb).

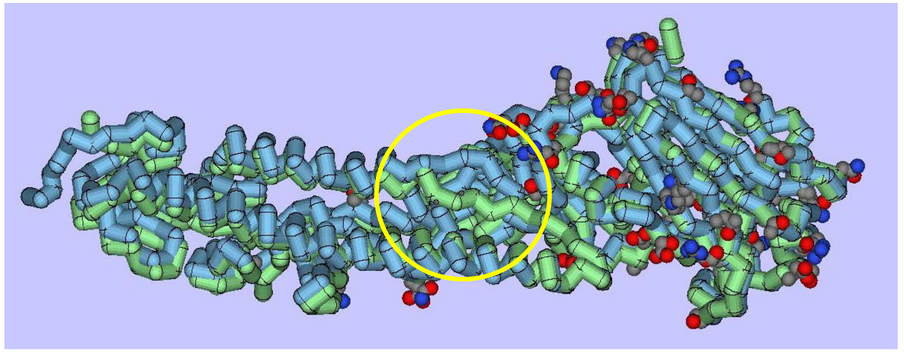

Fig 2. Overlay of H2 and H5 revealing a non symmetrical central loop (yellow circle).

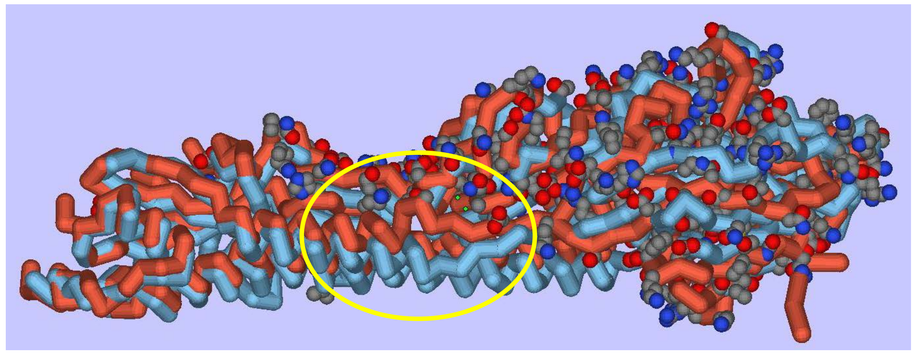

Fig 3. Overlay of H2 and H3 revealing torsional twist starting at prominent alpha helix (yellow circle).

Corresponding conserved regions contribute to an alpha helix, sialic acid receptor binding site, and to another alpha helix

```
60   NGIPPLELGDCSIAGWLLGNPECDRLLSVPEWSYIMEKENPRDGLCYPGSFNDYEELKHL 119 H2
55   PH_QILDGENCTLIDALLGDPQCDGFQN_KKWDLFVERSK_AYSNCYPYDVPDYASLRSL 111 H3
45   RGVAPLHLGKCNIAGWILGNPECESLSTASSWSYIVETPSSDNGTCYPGDFIDYEELREQ 104 H1
61   DGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGDFNDYEELKHL 120 H5
```

Corresponding conserved regions contribute to a beta sheet and the 190 alpha helix

```
178  YNNTSGEQMLIIWGVHHPNDETEQRTLYQNVGTYVSVGTSTLNKRSTPDIATRPKVNGLG 237 H2
168  MPNNEKFDKLYIWGVHHPGTDNDQIFLYAQASGRITVSTKRSQQTVIPNIGSRPRVRNIP 227 H3
165  YINDKGKEVLVLWGIHHPSTSADQQSLYQNADTYVFVGSSRYSKKFKPEIAIRPKVRDQE 224 H1
180  YNNTNQEDLLVLWGIHHPNDAAEQTKLYQNPTTYISVGTSTLNQRLVPRIATRSKVNGQS 239 H5
```

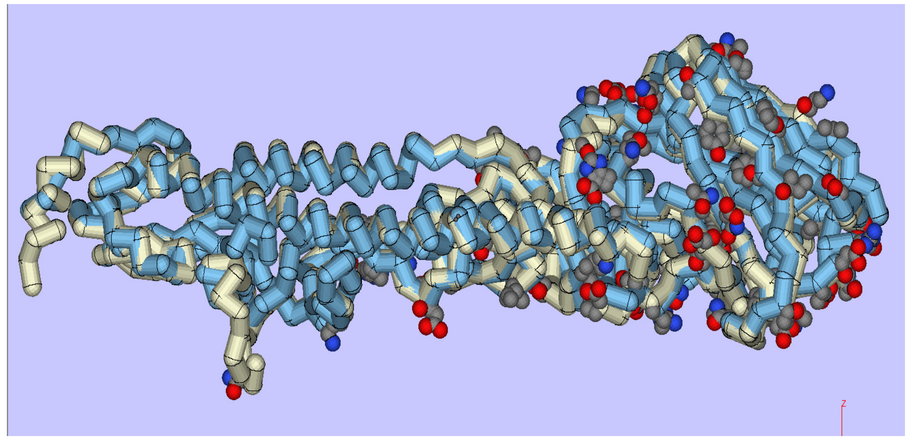

Fig 4. Overlay of H2 and H1 displaying high symmetry.

The CREST Program is funded by grant #1022793 from NSF-CCLI.

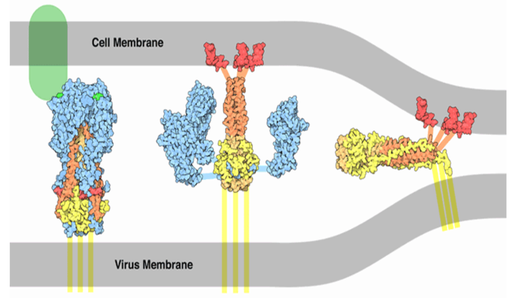

Fig 5. Hemagglutinin attachment and entry into host. (http://www.rcsb.org/pdb/101/motm.do?momID=76)

Results

Pair	Total AA (HA1)	Similar AA	Identical AA	% Similarity
H1-H2	327-324	99	176	53.8
H3-H2	329-324	110	112	33.4
H5-H2	324-324	61	222	68.3

- The structural analysis demonstrated high symmetry among both H1 and H2.
- High symmetry among H5 and H2 with the exception of one loop near the center of the peptide.
- Low symmetry between H3 and H2. The structural overlay of the H3 and H2 structures showed a torsional twist between the two proteins.

Conclusions

- These findings indicate that both H1 and H5 are more closely related to H2 than H3.
- The following amino acid sequences all corresponded to highly conserved secondary structures amongst all four subtypes: 11-18, 66-79, 105-115, 176-185, 188-195, 229-237, 281-288, 367-384, and 405-455 (H3 numbering).
- The following amino acids did not necessarily correspond to similar secondary structures but were all highly conserved amongst all four subtypes: 34-39, 97-102, 136, 153, 183, 245-254, 293-298, 300-310, 314-324 (H3 numbering).
- Structured based approaches to drug discovery can utilize the 3D structures identified in this research for investigation and predication of anti-viral therapy.

References

Leepakshi S, Tempczyk-Russell A, Agarwal, R. (2010). *PLoS One*. 5 (2), e9268.

Lin YP, Xiong X, Wharton SA, et al. (2013). *Proc Natl Acad Sci USA.* Feb 12; 110(7):2677.

Liu J, Stevens DJ, Haire LF, et al. (2009). *Proc Natl Acad Sci USA*. Oct 6; 106(40):17175-80.

Neumann G, Noda T, Kawaoka Y. (2009). *Nature*. 459 (18), 931-939.
Russell RJ, Gamblin SJ, Haire LF, Stevens DJ, Xiao B, Ha, Y, Skehel JJ. (2004). *Virology*. 325, 287-296.