Influenza viruses evolve through competing forces of immune pressure and protein functional constraints [1]. Immune pressure is exerted on the viruses to select for naturally occurring sequence variants that alter protein regions recognized by both the humoral and cellular arms of the adaptive immune system. However, the ability of influenza virus to evade host immunity through mutational selection is constrained by the requirement to retain essential protein functions necessary for effective virus replication and spread. This results in a constant mutation/selection tug-of-war that selects for sequence alterations in protein regions recognized by the host immune system (immune epitopes) while selecting against alterations in sequence features that impair protein function.

For vaccine development, the ideal immunogen would focus the vaccine immune response on protein regions that are both strong immune epitopes and functionally constrained.

Materials and Methods

We utilized data from the Influenza Research Database (IRD) [4] (www.fludb.org), which integrates information from primary sources such as:
- NCBI's Influenza Virus Database (IVD) [5]
- Immune Epitope Database (IEDB) (www.immuneepitope.org) [2]
- RCSB Protein DataBank (PDB)

Results

SNP/Consensus sequence generated through an internal IRD pipeline (adapted from Crooks et al. [6]) was used as well as custom script to quantify epitope coverage across each protein.

Introduction

To investigate the relationships between protein functional regions and immune epitopes, we compared the location of amino acids included in immune epitopes with a surrogate of functional constraint – sequence conservation. Figure 1B shows an overlap of graphs for epitope coverage and sequence variation for several influenza viruses.

In some cases:

1. Curves closely parallel each other (e.g. the region between amino acids 199 - 232 in HA (left))
2. Regions showing sequence variation are not found in validated immune epitopes.
3. Regions with limited sequence variation include amino acid positions found in multiple immune epitopes (e.g. the region around aa441 in HA; the region between aa406 – 421 in PB1). These amino acids appear to be highly conserved in spite of immune recognition: we refer to these as Conserved Epitope Regions (CER).

Results

Five CERs can be found in the HA protein (Figure 1B, Table 2). These regions are not only well conserved, but also include both B cell and T cell epitopes.

For the most part, the HA CER are found in the stalk region of the protein, with only one of the five CER located at the base of the globular head. Similar CER are found in each of the influenza proteins with significant representation in the IEDB data (Figure 4B, Table 2).

Table 1. Conserved Epitope Regions.

<table>
<thead>
<tr>
<th>Protein Regions</th>
<th>Location (aa)</th>
<th>CER coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>HA CER1</td>
<td>50 – 55</td>
<td>HA-CER1</td>
</tr>
<tr>
<td>HA CER2</td>
<td>115 – 130</td>
<td>HA-CER2</td>
</tr>
<tr>
<td>HA CER3</td>
<td>130 – 140</td>
<td>HA-CER3</td>
</tr>
<tr>
<td>NA CER1</td>
<td>395 – 410</td>
<td>NA-CER1</td>
</tr>
<tr>
<td>NA CER2</td>
<td>415 – 435</td>
<td>NA-CER2</td>
</tr>
</tbody>
</table>

Discussion

The current strategy for seasonal influenza vaccine development relies on the selection of two type A strains, an H1N1 and an H3N2, and one type B strain from recently circulating viruses predicted to be similar to the strains that will emerge in the coming season. This approach is problematic in that it:

- Requires the strain selection committee to accurately predict the basic structures of future viruses and also
- Requires vaccine manufacturers to develop new vaccine formulations each year.

Through the integration of data about the location of experimentally validated immune epitopes from the IEDB with an analysis of amino acid sequence conservation from the IRD, we have identified Conserved Epitope Regions (CER) within each of the influenza proteins that are targeted by the host immune system and that are conserved within type A influenza viruses.

We propose that these CER, especially those within the HA protein, could be excellent candidates for epitopes that could generate broad cross-reactive antibodies and therefore cross-protective immunity.

Recent experimental evidence suggests that the computational strategy described here does indeed generate experimentally verifiable results. Recently, investigators have isolated monomeric antibodies that demonstrate heterosubtypic cross-reactivity [7-9].

A comparison of the regions identified by the computational methods described here with the regions identified by these experimental methods shows a dramatic correlation (Table 2) with each of the antibody binding regions covered by one or more CER.

Table 2. Overlap between experimentally determined cross-reactive epitopes and computationally determined epitope regions.

<table>
<thead>
<tr>
<th>Region*</th>
<th>Location*</th>
<th>CER coordinates</th>
<th>coverage*</th>
</tr>
</thead>
<tbody>
<tr>
<td>HA-CER1</td>
<td>50 – 55</td>
<td>HA-CER1</td>
<td></td>
</tr>
<tr>
<td>HA-CER2</td>
<td>115 – 130</td>
<td>HA-CER2</td>
<td></td>
</tr>
<tr>
<td>HA-CER3</td>
<td>130 – 140</td>
<td>HA-CER3</td>
<td></td>
</tr>
<tr>
<td>NA-CER1</td>
<td>395 – 410</td>
<td>NA-CER1</td>
<td></td>
</tr>
<tr>
<td>NA-CER2</td>
<td>415 – 435</td>
<td>NA-CER2</td>
<td></td>
</tr>
</tbody>
</table>

Discussion

The CERs for HA:
- Include all four highly cross-reactive epitopes predicted by Duvvuri et al. [10]
- Are conserved in both seasonal H1N1 and the pandemic H1N1 2009 virus.
- Overlap with 6 other epitopes found by Duvvuri et al. for 10 of 18 (55.6%) epitopes in total.

In conclusion, the CERs described here provide excellent candidates for vaccine development that are known targets of the adaptive immune system. Epitopes from these CER offer broad cross-reactive immune response while being conserved in both seasonal H1N1 and the pandemic H1N1 2009 viruses.