A Novel Method for Designing Vaccines Against Constantly Mutating Pathogens

Eduardo A. Padlan

Marine Science Institute, College of Science
University of the Philippines, Diliman, Quezon City, Philippines

A method is described that could be used to design molecules suitable vaccines for immunization as or for the generation of therapeutic antibodies against constantly mutating pathogens. The method consists of (1) locating the putative immunodominant epitopes of a protein antigen, and identifying the residues that contribute to the high antigenicity of the epitopes and (2) replacing those residues judiciously to reduce the antigenicity of the immunodominant epitopes, while preserving structure. The antibody response to the modified antigen would then be expected to be directed against more parts of the molecule, and not mainly against the formerly immunodominant epitopes. The method is also useful in the design of molecules suitable for immunization against pathogens that had been intentionally mutated so as to render those pathogens able to infect previously immune individuals.

Key Words: immunodominant epitopes, antigenicity reduction, influenza virus

INTRODUCTION

Many pathogens are able to evade the immune system by constantly mutating their surface molecules. In so doing, the antibodies, which had been generated against previous strains of the pathogen, are no longer totally protective. In many instances, the pathogens are able to make the evasive move by mutating a few amino acids in a surface molecule that may contain several hundred residues.

Among the pathogens that are constantly mutating to achieve immune evasion are the influenza virus, the cold virus, and the virus that causes AIDS. These and other viruses and other pathogens have molecules on their surface, some of which are used to recognize and bind to the cells that the pathogens infect. Those surface molecules are the targets of the antibodies that our immune system produces to combat the pathogens. The external regions of those molecules can be used as vaccines. But a vaccine that utilizes a surface molecule, or part thereof, of a particular pathogen will be effective against mainly that pathogen, and may not offer protection against new or other strains of the organism.

It is generally accepted that every accessible region of an antigen could elicit an antibody response. Nevertheless, it is recognized that certain regions are substantially more antigenic than others; these are the immunodominant epitopes. Intuitively, it would seem obvious that constantly mutating pathogens could continually achieve immune evasion by having immunodominant epitopes on their surface molecules, and localizing the mutations in those epitopes.

If we are able to identify the immunodominant epitopes and reduce the antigenicity of those epitopes, we might be able to divert the antibody response to more parts of the antigen and not mainly to its formerly immunodominant epitopes. Indeed, there were hints from the early findings of Fazekas de St. Groth (1977) that there is a possibility of modulating the antibody response by engineering mutations in epitopes and the notion has been explored by others (e.g., Temoltzin-Palacios et al. 1994; Martinet et al. 1998).

*Corresponding author: edpadlan@aol.com
edpadlan@yahoo.com
The definitive identification of the residues that constitute an antigenic epitope can only be accomplished by a three-dimensional structural analysis of the antibody-antigen complex. Nonetheless, starting from the seminal work of Hopp & Woods (1981), various procedures have been proposed to locate the linear and the conformational epitopes of proteins by computation. Further, some residues in antigenic epitopes could be identified by the analysis of natural or laboratory-induced mutants (e.g., Gulati et al. 2002). Moreover, in some cases, the identity of antigenic residues could be inferred from a comparison of the amino-acid sequences of closely-related strains of the pathogen (e.g., Lee & Chen 2004).

A purely computational method is described here, which (1) locates the putative immunodominant epitopes of a protein antigen (putative because it is impossible to identify and delineate all the immunodominant epitopes of any antigen) and identifies the residues within those epitopes, and (2) suggests how those residues should be replaced to reduce antigenicity, a process hereinafter referred to as de-Antigenization, while preserving structure. The objective is to make all the accessible regions of the molecule essentially antigenically equivalent, so that the antibody response will be directed against more parts of the molecule and not mainly against the formerly immunodominant epitopes. De-Antigenized molecules are therefore useful for immunization, e.g., as effective vaccines or for the generation of therapeutic antibodies against more and potentially all strains of the pathogen. The method can also be used to design molecules useful against a pathogen that had been intentionally mutated so as to render that pathogen able to infect previously immune individuals.

The steps taken in designing vaccines against constantly mutating pathogens through de-Antigenization process is described in detail in this paper. The procedure is illustrated by the de-Antigenization of 2 surface antigens, 1 from an influenza virus that is currently circulating in the Philippines, and the other from a virus that is included in the current (2006-2007) flu vaccine (US Centers for Disease Control and Prevention, available in http://www.cdc.gov/flu). The development and testing of candidate vaccines generated using the method will be the subject of subsequent publications.

Influenza virus infects humans and a variety of animals including domestic and wild birds and lower mammals like cats, dogs, pigs, seals, and whales. The virus comes in 3 types: A, B, and C, and the antigens mainly targeted by the antibody response are the 2 surface glycoprotein molecules on the virus: hemagglutinin and neuraminidase. For type A influenza virus, there are 16 hemagglutinin subtypes and 9 neuraminidase subtypes currently known, and a particular influenza virus A is identified by its hemagglutinin and neuraminidase subtypes. Thus, an H3N2 influenza A virus has a subtype 3 hemagglutinin and a subtype 2 neuraminidase. The virus uses hemagglutinin and neuraminidase to infect cells, and infection is initiated by the activation of the hemagglutinin by enzymatic cleavage into 2 fragments, HA1 and HA2. The cleavage results in a significant structural transformation of the hemagglutinin (Wilson & Cox 1990, Skehel & Wiley 2000).

MATERIALS AND METHODS

General description of the method of designing vaccines by de-Antigenization:
The design of vaccines by the de-Antigenization of the immunodominant epitopes is achieved through the following steps:

**Step 1** Choose a protein molecule that has been identified as a major antigen in a particular pathogen.

**Step 2** Calculate the antigenicity of the various regions of the antigen using three-dimensional structural information about the molecule, and the known physicochemical properties of the amino-acid residues. Locate the regions with high antigenicities, i.e. the putative immunodominant epitopes.

**Step 3** Identify the amino-acid residues comprising the putative immunodominant epitopes particularly those residues which, by virtue of their physicochemical properties and accessibility can contribute significantly to tight binding by antibody. Replace those residues with amino acids that would be expected to contribute less to the binding by antibody, while ensuring that the replacements will not significantly alter the structure of the antigen. At least 1 T-cell epitope should be preserved.

**Step 4** Using the new structure (the structure with the replacements), repeat Steps 2 and 3 as needed until the putative immunodominant epitopes have significantly lower antigenicities.

**Step 5** The amino-acid sequences, which result in significantly lower antigenicities for the putative immunodominant epitopes provide the basis for molecules useful for immunization, for example, as vaccines or for the generation of therapeutic antibodies against the pathogen.
Structural and sequence data
Amino-acid sequence data for molecules of influenza viruses can be obtained from the Influenza Sequence Database (ISD) maintained at the Los Alamos National Laboratory (Macken et al 2001) (http://www.flu.lanl.gov), or from the sequence database maintained at the National Center for Biotechnology Information at the National Library of Medicine (NCBI/NLM) (http://www.ncbi.nlm.nih.gov). Three-dimensional structure data for some of the hemagglutinin and neuraminidase subtypes are available at the Protein Data Bank (Berman et al 2000) (http://www.rcsb.org/pdb). It is found that the hemagglutinin structures are quite similar to each other, and so are the neuraminidase structures, so that models for the other subtypes for which no experimentally-determined structures are available could be built with confidence.

Solvent accessibilities
The accessibility of an amino acid is a proper weight to use in the calculation of the residue's contribution to the total binding interaction. If the three-dimensional structure of the antigen is known, the solvent accessibility of the individual residues can be calculated using the computer algorithms developed by Connolly (1983); those algorithms are incorporated in the program DSSP (Kabsch & Sander 1983) (accessible for example in http://bioweb.pasteur.fr/seqanal/interfaces/dssp-simple.html). Fractional solvent accessibilities can be used to describe the degree of exposure of individual residues, e.g., as being completely buried, partly buried/partly exposed, etc. (Padlan, 1990); those can be estimated by simply dividing the solvent accessibility obtained from DSSP by the total surface area of the amino acid (obtained for example from http://prowl.rockefeller.edu/aainfo/volume.htm).

Calculation of antigenicities and identification of the putative immunodominant epitopes
The method, which had been proposed earlier for quantifying the antigenicity of a given region in a protein molecule using the physicochemical attributes of the amino-acid residues in the region (Padlan 1985), is particularly suitable for locating the putative immunodominant epitopes and is followed here. Structural parameters describing the physicochemical attributes of the various amino acids have been computed by various authors, starting with the seminal work of Sneath (1966). Physicochemical parameters have been found useful in quantifying the reactivities and interaction of antibodies and antigens (e.g., Padlan 1990, Novotny 1991, David et al 2007) and other protein-protein interactions (e.g., De Genst et al 2002). The physicochemical parameters compiled by Grantham (1974) and by Sandberg et al (1998) are particularly useful in computing antigenicities and are presented in Figure 1.

The antigenicity of a region in the molecule is computed by taking the sum of the structural parameters, weighted or unweighted, corresponding to all the residues within that region. In the examples that follow, the solvent accessibilities are used as weighting factors in the calculation of the antigenicities. The use of solvent accessibilities as weighting factors de-emphasizes the contribution of residues that are not too accessible and probably do not contribute much to the interaction with antibody. Antigenicity computed in this manner is directly correlated with the ability of a particular region to engage in tight binding to antibody. The regions displaying highest antigenicities are identified as the putative immunodominant epitopes. In the examples below, regions with antigenicity values at least 2 root-mean-square (r.m.s.) deviations above the average are identified as the putative immunodominant epitopes. The value of 2 r.m.s. deviations is a personal choice; preliminary calculations had shown that 3 r.m.s. deviations above the average yielded too few putative immunodominant epitopes and sometimes none at all, while 1 r.m.s. deviation yielded too many.

De-Antigenization of the putative immunodominant epitopes
The de-Antigenization of the putative immunodominant epitopes is achieved by the judicious replacement of the residues in those epitopes with amino acids that would contribute less to the total antigenicity values, while preserving the structure of the molecule. The replacement rules used here (included in Figure 1) were developed through a careful comparison of the propensities of the various amino acids for helix formation (e.g., Pace & Scholtz 1998), beta-sheet formation (e.g., Street & Mayo 1999), turn formation (e.g., Hutchinson & Thornton 1994), and the relative probabilities of their occurrence in coils (e.g., Linding et al 2003). Residues, which are usually structurally critical, e.g., cysteines, prolines, and glycines are not replaced.

The residues to be replaced would be the ones located within a certain radius of the epitope centers (the alpha-carbon positions in the examples below). A suitable value for the radius could be determined by examining known antibody-antigen complexes. In the examples below, the residues to be replaced were chosen on the basis of their solvent accessibility and relative contribution to the overall antigenicity of the epitope. The residues that are probably critical to the structure (secondary, tertiary, quaternary)
### The amino acid parameters used in the calculation of antigenicities and the replacement suggestions

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>zz1</th>
<th>zz2</th>
<th>zz3</th>
<th>zz4</th>
<th>zz5</th>
<th>SDGly</th>
<th>Helix</th>
<th>Sheet</th>
<th>Coil Propensities</th>
<th>Turn</th>
<th>If in Helix</th>
<th>Sheet</th>
<th>Coil</th>
<th>Turn</th>
<th>Change to:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>0.24</td>
<td>-2.32</td>
<td>0.60</td>
<td>-0.14</td>
<td>1.30</td>
<td>60.0</td>
<td>0.00</td>
<td>0.47</td>
<td>-0.26154</td>
<td>0.83</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>3.52</td>
<td>2.50</td>
<td>-3.60</td>
<td>1.99</td>
<td>-0.17</td>
<td>125.0</td>
<td>0.21</td>
<td>0.35</td>
<td>-0.17659</td>
<td>0.82</td>
<td>Ala</td>
<td>Thr</td>
<td>Ala</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>3.05</td>
<td>1.62</td>
<td>1.04</td>
<td>-1.15</td>
<td>1.61</td>
<td>80.0</td>
<td>0.65</td>
<td>0.40</td>
<td>0.22989</td>
<td>1.44</td>
<td>Ala</td>
<td>Thr</td>
<td>Ser</td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>3.98</td>
<td>0.93</td>
<td>1.93</td>
<td>-2.46</td>
<td>0.75</td>
<td>94.0</td>
<td>0.69</td>
<td>0.72</td>
<td>0.22763</td>
<td>1.41</td>
<td>Ala</td>
<td>Thr</td>
<td>Ser</td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td>0.84</td>
<td>-1.67</td>
<td>3.71</td>
<td>0.18</td>
<td>-2.65</td>
<td>189.0</td>
<td>0.68</td>
<td>0.25</td>
<td>-0.015182</td>
<td>1.08</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>1.75</td>
<td>0.50</td>
<td>-1.44</td>
<td>-1.34</td>
<td>0.66</td>
<td>97.0</td>
<td>0.39</td>
<td>0.34</td>
<td>-0.187677</td>
<td>0.94</td>
<td>Ala</td>
<td>Thr</td>
<td>Ala</td>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>3.11</td>
<td>0.26</td>
<td>-0.11</td>
<td>-3.04</td>
<td>-0.25</td>
<td>98.0</td>
<td>0.40</td>
<td>0.35</td>
<td>-0.20469</td>
<td>1.01</td>
<td>Ala</td>
<td>Thr</td>
<td>Ala</td>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>2.05</td>
<td>-4.06</td>
<td>0.36</td>
<td>-0.82</td>
<td>-0.38</td>
<td>0.0</td>
<td>1.00</td>
<td>--</td>
<td>0.43323</td>
<td>1.48</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>2.47</td>
<td>1.95</td>
<td>0.26</td>
<td>3.90</td>
<td>0.09</td>
<td>98.0</td>
<td>0.56</td>
<td>0.37</td>
<td>-0.0012174</td>
<td>1.07</td>
<td>Ala</td>
<td>Thr</td>
<td>Thr</td>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>-3.89</td>
<td>-1.73</td>
<td>-1.71</td>
<td>-0.64</td>
<td>0.26</td>
<td>135.0</td>
<td>0.41</td>
<td>0.10</td>
<td>-0.42224</td>
<td>0.59</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>-4.28</td>
<td>-1.30</td>
<td>-1.49</td>
<td>-0.72</td>
<td>0.84</td>
<td>138.0</td>
<td>0.21</td>
<td>0.32</td>
<td>-0.33793</td>
<td>0.66</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>2.29</td>
<td>0.89</td>
<td>-2.49</td>
<td>1.49</td>
<td>0.31</td>
<td>127.0</td>
<td>0.26</td>
<td>0.34</td>
<td>-0.100092</td>
<td>1.01</td>
<td>Ala</td>
<td>Thr</td>
<td>Thr</td>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td>-2.85</td>
<td>-0.22</td>
<td>0.47</td>
<td>1.94</td>
<td>-0.98</td>
<td>127.0</td>
<td>0.24</td>
<td>0.26</td>
<td>-0.22590</td>
<td>0.57</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>-4.22</td>
<td>1.94</td>
<td>1.06</td>
<td>0.54</td>
<td>-0.62</td>
<td>163.0</td>
<td>0.54</td>
<td>0.13</td>
<td>-0.22557</td>
<td>0.89</td>
<td>Ala</td>
<td>Thr</td>
<td>Ala</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>-1.66</td>
<td>0.27</td>
<td>1.84</td>
<td>0.70</td>
<td>2.00</td>
<td>42.0</td>
<td>3.01</td>
<td>--</td>
<td>0.55232</td>
<td>1.38</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>2.39</td>
<td>-1.07</td>
<td>1.15</td>
<td>-1.39</td>
<td>0.67</td>
<td>56.0</td>
<td>0.50</td>
<td>0.30</td>
<td>0.14288</td>
<td>1.15</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>0.75</td>
<td>-2.18</td>
<td>-1.12</td>
<td>-1.46</td>
<td>-0.40</td>
<td>59.0</td>
<td>0.66</td>
<td>0.06</td>
<td>0.0088780</td>
<td>1.00</td>
<td>Ala</td>
<td>Thr</td>
<td>Ala</td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>Trp</td>
<td>-4.36</td>
<td>3.94</td>
<td>0.59</td>
<td>3.44</td>
<td>-1.59</td>
<td>164.0</td>
<td>0.49</td>
<td>0.24</td>
<td>-0.248377</td>
<td>0.70</td>
<td>Ala</td>
<td>Thr</td>
<td>Ala</td>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>-2.54</td>
<td>2.44</td>
<td>0.43</td>
<td>0.04</td>
<td>-1.47</td>
<td>147.0</td>
<td>0.53</td>
<td>0.11</td>
<td>-0.20751</td>
<td>0.92</td>
<td>Ala</td>
<td>Thr</td>
<td>Ala</td>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>-2.59</td>
<td>-2.64</td>
<td>-1.54</td>
<td>-0.85</td>
<td>-0.02</td>
<td>109.0</td>
<td>0.61</td>
<td>0.13</td>
<td>-0.38618</td>
<td>0.70</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1.** Structural characteristics of the different naturally-occurring amino acids and the amino-acid replacements designed to reduce the antigenicity of protein epitopes. The zz values are from Sandberg et al (1998). The SDGly values are from Grantham (1974) and represent the structural dissimilarities of the various amino acids relative to glycine. The helix propensities are from Pace & Scholtz (1998), the beta-sheet propensities are from Street & Mayo (1999), the coil propensities are from Linding et al (2003), and the turn propensities are from Hutchinson & Thornton (1994). A dash in the replacement suggestions signifies that no change is recommended.
of the antigen are preserved. Also preserved is at least 1 putative T-cell epitope, which may be available from the literature or obtained using predictors, e.g. SYFPEITHI (Rammensee et al 1999) (http://www.syfpeithi.de).

In the examples below, residues within each putative immunodominant epitope, whose fractional solvent accessibility is at least 40% and whose contribution to the antigenicity value of the epitope is at least 3% were replaced according to the rules proposed (Figure 1). The cut-off of 40% in the fractional accessibility value preserves residues that are buried or mostly buried (Padlan 1990). A minimum contribution of 3% to the total antigenicity was chosen, so that the number of residues contributing to the putative epitopes located by the procedure described here are comparable to those in epitopes of known structure (e.g., Padlan 1996).

After the replacements, a new model based on the revised amino-acid sequence is generated and antigenicity calculations are performed on this new model.

**Example 1 - Design of possible vaccines against H3 influenza A viruses based on the hemagglutinin molecule**

**Structural and sequence data**

In view of the significant structural transformation that accompanies the activation of hemagglutinin, the uncleaved hemagglutinin molecule is the proper basis for vaccine development. The three-dimensional structure of the uncleaved, bromelain-released H3 hemagglutinin from the influenza A virus, A/AICHI/68 (with the arginine at the cleavage site replaced by glutamine), has been provided by X-ray crystallography (Chen et al. 2002) (Protein Data Bank entry 1HA0). Hereinafter, this H3 hemagglutinin is referred to simply as 1HA0.

To illustrate the design of a possible vaccine against H3 influenza A viruses based on the hemagglutinin molecule, we use the hemagglutinin from the influenza A virus, A/Wisconsin/67e5/2005, a 2006-2007 vaccine strain whose amino-acid sequence is available as entry ISDN138724 in the ISD database. The molecule, which corresponds to the bromelain-released fragment of the mature hemagglutinin molecule includes residues 1-494 (sequential numbering) and is shown in Figure 2. In the sequence shown in Figure 2, the segment bounded by positions 298 and 310 was replaced by PKYVKQNTLKLAT, which is an immunodominant T-cell epitope in humans (Hennecke et al. 2000). Hereinafter, this composite H3 hemagglutinin, with the T-cell epitope imposed, is referred to simply as ISDN138724.

A three-dimensional model for ISDN138724 was generated with the help of the protein modeling server, SWISS-MODEL (Schwede et al 2003) (implemented in http://swissmodel.expasy.org), using the 1HA0 structure as template. The biological, i.e., natural, aggregation state of influenza hemagglutinin is a trimer; in generating the ISDN138724 trimer, the 1HA0 trimer structure was used as template. All subsequent calculations were based on the ISDN138724 trimer.

**Solvent accessibilities**

The solvent accessibilities of the individual residues of ISDN138724 were obtained using the program DSSP and fractional accessibilities were estimated as described above.

**Calculation of antigenicities and identification of the putative immunodominant epitopes**

Antigenicity calculations were performed on the modeled ISDN138724 molecule. The zz1-3 scales of Sandberg et al (1998) (reproduced in Figure 1), which relate the various amino acids on the basis of hydrophilicity, size and polarizability, and electronic properties, were the structural parameters used in the calculation of antigenicities. The antigenicity of a region centered at each alpha-carbon position was computed by taking the sum of the structural parameters corresponding to all the residues within 16 Angstroms of the alpha-carbon. The solvent accessibilities of the individual ISDN138724 residues were used as weighting factors in the calculation of the antigenicities. The average antigenicity value for the molecule and the root-mean-square deviation from the average were computed. Regions with antigenicity values at least two root-mean-square (r.m.s.) deviations above the average were identified as the putative immunodominant epitopes.

**De-Antigenization of the putative immunodominant epitopes**

Three rounds of de-Antigenization, i.e. modeling, followed by antigenicity calculations, followed by amino-acid replacements were computed. The first model of ISDN138724 had an average antigenicity value of 24.7 (r.m.s. deviation = 16.3) (arbitrary units). A total of 22 amino-acid replacements were performed. This resulted in an average antigenicity value of 11.4 (r.m.s. deviation = 13.6); 13 more changes were suggested. The additional replacements resulted in an average antigenicity value of 4.3 (r.m.s. deviation = 12.4); 3 more changes were suggested. This resulted in an average antigenicity value of 2.6 (r.m.s. deviation = 11.5). At this point, regions that had been identified as immunodominant and which had already been de-Antigenized were again indicated as
Figure 2. Amino-acid sequence of the H3 hemagglutinin, ISDN138724, before and after the three rounds of de-Antigenization.
being significantly above average. This signaled that the process had converged. No further de-Antigenization of ISDN138724 was performed after the third round.

**Example 2 - Design of possible vaccines against N2 influenza A viruses based on the neuraminidase molecule**

**Structural and sequence data**

A crystallographically-determined structure for the pronase-released fragment of an N2 neuraminidase from influenza A virus is available from the Protein Data Bank. The neuraminidase is from the influenza virus, A/Tokyo/3/67 (Varghese et al 1991) (Entry 2BAT), and hereinafter is referred to simply as 2BAT.

To illustrate the design of a possible vaccine against N2 influenza A virus based on the neuraminidase molecule, we use the neuraminidase from the influenza A virus, A/California/7/2004, available as entry ISDN117767 in the ISD database. This virus is one of those currently circulating in the Philippines. Hereinafter, this N2 neuraminidase is referred to simply as ISDN117767.

The segment of ISDN117767, which corresponds to the pronase-released fragment of the mature neuraminidase molecule includes residues 1-388 (sequential numbering) and is shown in Figure 3. A three-dimensional model for ISDN117767 was generated with the help of the protein modeling server, SWISS-MODEL, using the 2BAT structure as template. The biological aggregation state of influenza neuraminidase is a tetramer; in generating the ISDN117767 tetramer, the 2BAT tetramer structure was used as template. All subsequent calculations were based on the ISDN117767 tetramer.

**Solvent accessibilities**

The solvent accessibilities of the individual residues of ISDN117767 and the fractional accessibilities were estimated as in Example 1.

Calculation of antigenicities and identification of the putative immunodominant epitopes

Antigenicity calculations were performed on the modeled ISDN117767 molecule using the procedure described in Example 1.

**De-Antigenization of the putative immunodominant epitopes**

The de-Antigenization ISDN117767 was done following the procedure used in Example 1. No residues within the segment bounded by positions 131 and 146 (sequential numbering) were replaced; the segment, 131-146, is identified as a possible T-cell epitope by the program SYFPEITHI.

**Figure 3. Amino-acid sequence of the N2 neuraminidase, ISDN117767, before and after one round of de-Antigenization.**

<table>
<thead>
<tr>
<th>De-An 1</th>
</tr>
</thead>
</table>
| ISDN117767 before and after one round of de-Antigenization  
| AYRNWSKCPQCDITGFAPFKSDNIRLSAGDINWTPRYVSCDPDKCYQFALGGTTLN | 60  
| AYRNWSKCPQCDITGFAPFKSDNIRLSAGDINWTPRYVSCDPDKCYQFALGGTTLN |  
| NVHSNDDTVHTDPYRRTLLMMNELGVPFHGTKQVCIAWSSSSHDGKAWLHVCVTGSDKNA | 120  
| NVHSNDDTVHTDPYRRTLLMMNELGVPFHGTKQVCIAWSSSSHDGKAWLHVCVTGSDKNA |  
| TASFYYNGRLVDSIVSWKSEILRTQESECVICNGTCVMTDGSASGDKAFTKILFIEEGK | 180  
| TASFYYNGRLVDSIVSWKSEILRTQESECVICNGTCVMTDGSASGDKAFTKILFIEEGK |  
| IVHTSTLSGSAQHVEGECYPRTVGRCRNWKGSNRPIVDINKYSIVSYVCSSL | 240  
| IVHTSTLSGSAQHVEGECYPRTVGRCRNWKGSNRPIVDINKYSIVSYVCSSL |  
| VGDTPRKNDSSSSSHCLUDPNNEEGGHGVKWAFDDGDNWVMGRTISEKLRSGYETIKVIE | 300  
| VGDTPRKNDSSSSSSTCLGNSEEGGTGVKWAFDDGDNWVMGRTISEKLRSGYETIKVIE |  
| GWSNPNSKQLQINRQIVVDGRNRSYGSGIFSVEKSCINRCFYVLEIRGKRKETEVLWTSN | 360  
| GWSNPNSKQLQINRQIVVDGRNRSYGSGIFSVEKSCINRCFYVLEIRGKRKETEVLWTSN |  
| SIVVFCGTSTGYGTSYGTGSWPDDILMI | 388  
| SIVVFCGTSTGYGTSYGTGSWPDDILMI |  

45
Only 1 round of de-Antigenization was required to reduce the antigenicity of the putative immunodominant epitopes of ISDN117767 below the chosen level of significance (2 r.m.s. deviations above the average). The first model of ISDN117767 had an average antigenicity value of 21.0 (r.m.s. deviation = 11.5) (arbitrary units). A total of 25 amino-acid replacements were performed. This resulted in an average antigenicity value of 1.2 (r.m.s. deviation = 12.0). At this point, regions that had already been de-Antigenized were re-identified as immunodominant, signaling that the de-Antigenization had converged. The procedure was halted at this point.

RESULTS

Possible vaccines against an H3N2 influenza A virus that was predicted to circulate in 2006-2007 and against one that is currently circulating in the Philippines

The plots of antigenicities computed for the ISDN138724 hemagglutinin before and after the 3 rounds of de-Antigenization, and for the ISDN117767 neuraminidase before and after 1 round of de-Antigenization, are presented in Figures 4 and 5, respectively. Note the significant lowering of the average antigenicity in both cases after de-Antigenization. Further, the peaks in the antigenicity values, which existed before de-Antigenization are practically gone and a flatter distribution is obtained.

The amino-acid sequence and the suggested replacements based on the rounds of de-antigenization are included in Figure 2 for ISDN138724 and in Figure 3 for ISDN117767. Since every round of de-Antigenization resulted in a significant decrease in the antigenicity of the putative immunodominant epitopes, any of the de-Antigenized sequences (De-An 1 through 3 in Figure 2 for ISDN138724, De-An 1 in Figure 3 for ISDN117767) could be the basis for a molecule useful for immunization, for example as a vaccine, or for the generation of therapeutic antibodies against currently-circulating H3N2 influenza A viruses.

DISCUSSION

One major difference in the calculation of antigenicities between the method proposed earlier (Padlan 1985) and the method described here is the size of the presumed antigenic epitope. In the earlier method, a radius of 8.5 Angstroms was used as the extent of the epitope. The value of 8.5 Angstroms was a guess, in view of the fact that no structure was available at that time for any antibody-antigen complex. The structures of many antibody-antigen complexes are now available in the Protein Data Bank. In the method described here, a more appropriate value of the radius was calculated from the known structures of the complexes of antibodies with the antigens of interest. A survey of the structures of the 7 unique antibody-hemagglutinin and antibody-neuraminidase complexes currently available in the Protein Data Bank (Entries: 1KEN, 1EO8, 1QFU, 2VIR, 2AEP, 1NMB, 1NCD) revealed that the epitopes may extend up to 24 Angstroms from the alpha-carbon nearest the epitope center, but that 93.3% of the residues in those epitopes are within 16 Angstroms of that alpha-carbon. Thus, a 16-Angstrom radius was used in the examples. Another major difference is the use here of solvent accessibilities as weights in the computation of antigenicities.

It is vitally important that the structure of the antigen be for the most part preserved during de-Antigenization. While structure preservation can only be ascertained by actual three-dimensional structure analysis, precautions can be taken to ensure that no major departures from the original structure will result from the amino-acid replacements. The replacement rules used here (included in Figure 1) were carefully designed to ensure that the original and replacing residues have very similar secondary-structure propensities, and that residues that are critical for maintaining three-dimensional structure, e.g. cysteines, prolines, and glycines as well as inward-pointing residues and those involved in quaternary contacts are preserved.

In the 2 cases described here (and in other cases analyzed by the author so far), 1 round of de-Antigenization already reduces the antigenicity of the putative immunodominant epitopes below the chosen significance level (2 r.m.s. deviations above the mean in these examples). Second-tier epitopes may be revealed after reduction of the antigenicity of the primary immunodominant epitopes, and de-Antigenization of those minor epitopes can be done by additional rounds of de-Antigenization if desired. This should be weighed against the possibility that significant structural changes might result from too many amino-acid replacements.

So far, no method exists that can pinpoint the precise location of immunodominant epitopes in a molecule. While three-dimensional structure analysis of antibody-antigen complexes can locate antigenic epitopes, one cannot be certain that those epitopes are immunodominant. We can guess.
Figure 4. Plots of antigenicity vs residue position for the H3 hemagglutinin, ISDN138724, before (top) and after three rounds of de-Antigenization (bottom).
Figure 5. Plots of antigenicity vs residue position for the N2 neuraminidase, ISDN117767, before (top) and after one round of de-Antigenization (bottom).
For hints, we could look at the other molecule that is involved in the interaction, namely, the antibody. Antibodies are generalized ligand binders capable of binding all sorts of structures. An analysis of antibody combining sites (Padlan 1990) revealed the preponderance in those sites of amino acids whose physicochemical properties allow the residues to contribute significantly to the binding. Indeed, there is a dearth of alanines, valines, isoleucines, and leucines, i.e. small, alipathic, apolar residues in antibody combining sites. What is found instead is an unusually large proportion of histidines, asparagines, tyrosines, and even solvent-exposed tryptophans in those sites (Padlan 1990). We could probably expect the same for immunodominant epitopes and that is the basic assumption here, i.e. the residues that constitute the immunodominant epitopes are mostly those whose physicochemical properties make them most suitable for participation in tight binding. Thus, the use of structural parameters that reflect the different physicochemical properties of the amino acids in the calculation of antigenicities is expected to reveal the regions that have high binding capacity; in other words, the immunodominant epitopes.

However the available three-dimensional data reveal that the shape and extent of antigenic epitopes are quite variable (see, for example, Padlan 1996) and cannot be predicted. Thus, without an actual three-dimensional analysis, it would be very difficult if not impossible to identify the residues that constitute a particular epitope. Here, again, we have to guess.

Various strategies have been used in the attempt to identify antigenic residues, including the comparison of the sequences of different strains, the identification of the mutations in variants selected using monoclonal antibodies (escape mutants), and the search for the existence of positively-selected or highly-mutable codons. Lee & Chen (2004) have collected the results of such analyses for the HA1 fragment of H3N2 viruses, while Gulati et al (2002) have identified antigenic residues in the neuraminidase of a recent H3N2 virus. Comparing their findings with ours shows that only 22 of the 42 positions, which had been identified as antigenic in H3 hemagglutinin by Lee & Chen, are predicted to be part of the immunodominant epitopes of ISDN138724 by the method presented here. In contrast, all of the 27 residues (the 27 positions in common with the ISDN117767 structure), which had been identified as antigenic by Gulati et al. are predicted to be part of immunodominant epitopes by the present method. The disparity emphasizes the difficulty in identifying the antigenic residues with precision.

CONCLUSION

The method that is presented here could be used to design molecules for possible use as vaccines or for the generation of therapeutic antibodies against antigenically unstable pathogens. The method was illustrated using as examples antigens from pathogens that are naturally continually mutating. The method can also be used to generate molecules useful against pathogens that have been intentionally engineered to escape immune system responses.

The possible use of the method in designing effective vaccines will have to be verified by actual testing, first in animal studies and ultimately in human clinical trials. Candidate vaccines, either in the form of polypeptides obtained by recombinant technology or in the form of DNA, could be used. In this paper, only the possible effects of sequence modifications, i.e. the intrinsic factors concerning immunogenicity are considered; naturally, other factors, e.g. mode of administration, dosage, etc., in other words, the extrinsic factors will have to be considered during testing.


ACKNOWLEDGEMENTS

I thank Dr. Remigio M. Olveda of the Research Institute for Tropical Medicine, Department of Health, for the information on strains of influenza virus currently circulating in the Philippines, and Dr. Gisela P. Concepcion of the Marine Science Institute, University of the Philippines Diliman, for comments and discussions. This paper is dedicated to my parents, Dr. Feliciano Macaraeg Padlan and the late Dr. Aida Almeda Agustin-Padlan.

REFERENCES


STREET AG and MAYO SL. 1999. Intrinsic beta-sheet propensities result from van der Waals interactions between side chains and the local backbone. Proceedings of the National Academy of Science USA 96:9074-9076.

