JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Structurally Conserved Binding Sites of Hemagglutinin as Targets for Influenza Drug and Vaccine Development

Muhammad Yusuf,[†,#] Janez Konc,[‡,#] Choi Sy Bing,[†] Joanna Trykowska Konc,[‡] Nurul Bahiyah Ahmad Khairudin,[§] Dusanka Janezic,[*,‡,∥] and Habibah A. Wahab[†,⊥]

[†]Pharmaceutical Design and Simulation (PhDS) Laboratory, School of Pharmaceutical Sciences, Universiti Sains Malaysia, 11800 Minden, Pulau Pinang, Malaysia

[‡]National Institute of Chemistry, Ljubljana, Slovenia

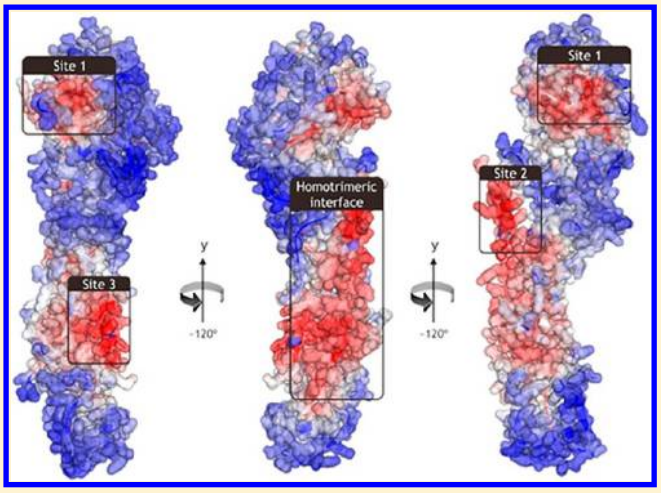
[§]Malaysian-Japan International Institute of Technology, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia

[∥]Faculty for Mathematics, Natural Sciences and Informatics, University of Primorska, Koper, Slovenia

[⊥]Malaysian Institute of Pharmaceuticals and Nutraceuticals, Ministry of Science, Technology and Innovation, Halaman Bukit Gambir, 11900, Bayan Lepas, Pulau Pinang, Malaysia

Supporting Information

ABSTRACT: ProBiS is a new method to identify the binding site of protein through local structural alignment against the nonredundant Protein Data Bank (PDB), which may result in unique findings compared to the energy-based, geometry-based, and sequence-based predictors. In this work, binding sites of Hemagglutinin (HA), which is an important target for drugs and vaccines in influenza treatment, have been revisited by ProBiS. For the first time, the identification of conserved binding sites by local structural alignment across all subtypes and strains of HA available in PDB is presented. ProBiS finds three distinctive conserved sites on HA's structure (named Site 1, Site 2, and Site 3). Compared to other predictors, ProBiS is the only one that accurately defines the receptor binding site (Site 1). Apart from that, Site 2, which is located slightly above the TBHQ binding site, is proposed as a potential novel conserved target for membrane fusion inhibitor. Lastly, Site 3, located around Helix A at the stem domain and recently targeted by cross-reactive antibodies, is predicted to be conserved in the latest H7N9 China 2013 strain as well. The further exploration of these three sites provides valuable insight in optimizing the influenza drug and vaccine development.



INTRODUCTION

Many concepts are used in binding site prediction such as genomic, geometric, energy-based, probe-mapping, docking, and evaluation of binding site similarities using a related database.[1] The exploration of locally similar 3D patterns of physicochemical properties of amino acid on the protein surface could also be used in defining binding sites that may lack sequence and global structural conservation. ProBiS,[2] which is a recent Web-based binding site predictor, locates binding sites based on the local structural comparison of the entire protein surface against nonredundant PDB structures.[3] It thus enables one to define the structurally conserved sites in a protein, which are regarded as the functional sites.[4] ProBiS prediction and analysis have been successfully applied in many studies, including elucidation of biochemical function of hypothetical proteins, old (existing) drug repositioning, off-target analysis, mechanism of action of ligand binding, and many more.[3,5−14] In this study, we revisited the binding sites of hemagglutinin (HA) through the perspective of local structural conservation across its various strains and subtypes using the ProBiS-2012 Web server (http://probis.cmm.ki.si/).

Hemagglutinin (HA) is the receptor-binding and membrane fusion glycoprotein of the influenza virus.[15] The influenza virus still poses a great threat to humanity. Pandemic influenza has a terrifying record, of which the worst killed more than 20 million people.[16] The appearance of unexpected and novel pandemic influenza viruses in human populations over the centuries[17] and their potential financial, societal, and political impacts have laid the foundation for the need of the development of influenza vaccines.[18] Most recently, as of May 30, 2013, a new influenza virus, which is believed to be H7N9, was found in 132 patients resulting in 37 deaths in China.[19] This new virus is different

Received: July 18, 2013
Published: August 27, 2013

from other H7N9 viruses previously found in birds,[20] and it infected humans for the first time in history.[21]

The influenza virus undergoes frequent and unpredictable changes in the surface glycoproteins. Some of these changes may result in a highly virulent strain with the potential for causing a pandemic.[22] Furthermore, these antigenic variations enable the virus to escape the immune system, thus reducing the effectiveness of vaccines. The currently available vaccines include whole virus vaccines, subunit vaccines, and live-attenuated influenza virus vaccines. They are all strain-specific. Their efficacy is limited against seasonal and closely related antigenic viral strains, but these vaccines do not protect against antibody-escape variants of seasonal or novel influenza A viruses. Hence, there is an urgent call to develop more effective vaccines, and novel approaches are being considered, including the use of conserved epitopes of influenza A virus proteins, which induce cross-protective immune response.[23−26]

During influenza infection, the influenza virus attaches to cell-surface receptors containing sialic acids (such as SA-α-2,6-Gal and SA-α-2,3-Gal terminal glycan that are on human and avian receptors, respectively) by HA that mediates both the initial attachment of the virus to target cells and the subsequent fusion of the viral and cell membranes (as reviewed in ref 27). Hemagglutinin has millimolar affinity toward the sialic acids, which indicates that the receptor binding site has low affinity properties.[15] The receptor-binding site of HA is a shallow depression at the globular domain (top in Figure 1) of this protein and is composed of amino acid residues Y98, H183, E190, W153, and L194.[28] These residues are generally conserved in many strains of HA.[27] It is suggested that tight binding of the influenza virus to the host cell is a consequence of simultaneous interaction of many HA surfaces.[15] Hence, unlike the current neuraminidase (NA) inhibitor, a single molecule is expected to be ineffective against HA. Multivalent molecules are preferred in the effort of designing molecules that inhibit the receptor binding site of HA.[29,30] For this reason, HA is a more attractive target for vaccine development. Nevertheless, the effort in designing both drugs and vaccines shall be improved by the deeper understanding of their protein target, e.g., structurally conserved binding sites.

Hemagglutinin is synthesized as a single polypeptide precursor HA0 that forms a trimer and is transferred through the Golgi apparatus to the cell surface. Cleavage of HA0 generates two subunits HA1 and HA2 and is required for membrane fusion activity and infectivity (reviewed in ref 15). A part of the monomer, HA1 makes up a disulfide-linked globular head, while the part consisting of HA2 makes up the stem domain (Figure 1). The globular head HA1 contains the receptor-binding pocket surrounded by variable antigenic sites, or epitopes. Anti-HA antibodies neutralize virus infectivity[27] and, hence, halt infection by binding to specific regions of HA called (functional) epitopes, which are located within presumed topologically distinct antigenic sites.[31] Five antigenic sites (A−E) have been identified on the HA1 region of H3 subtype.[27,28,32] These antigenic sites cover much of the surface of the globular head of HA, including residues around the receptor-binding site. The accumulation of amino acid mutations in the antigenic sites of HA causes antigenic drift to escape immune suppression.[27] Although HA is known to undergo extensive sequence and antigenic variation in its receptor-binding subunit, it also possesses conserved structural features in the HA2 segment involved in anchoring to the viral membrane. Conservation of HA2 makes this domain suitable as a potential target for inducing broadly cross-reactive immunity.[33]

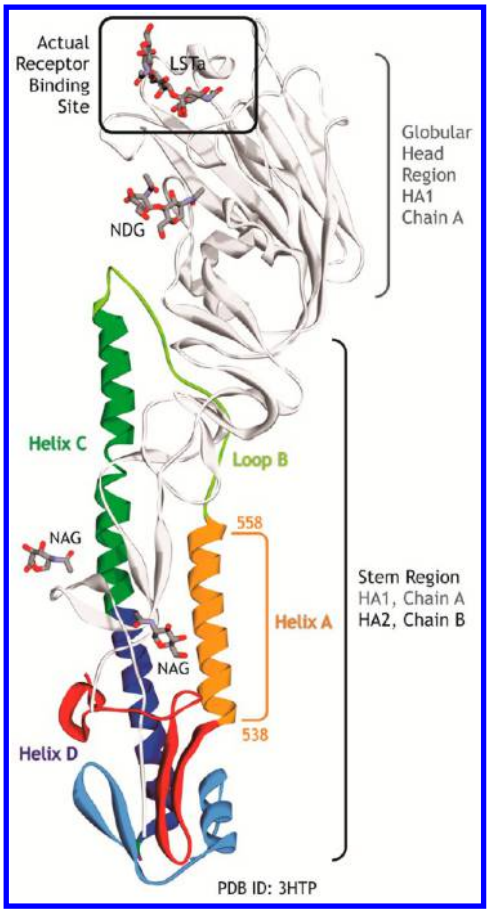


Figure 1. Crystal structure of 2005 H1N1 HA (PDB ID 3HTP) complexed with LSTa.[40] The receptor binding site (Site 1) is inside the box at the globular head region (HA1). The structure was generated using Accelrys Discovery Studio Visualizer 3.5.[41]

The choice of HA as a test protein in this study is not only due to the potential application of this protein in influenza drug and vaccine development but also due to its shallow cavities that make up the receptor binding sites that are typically difficult to detect using methods that rely on structural geometry alone.[34] Because the receptor binding site of HA is already well characterized, the accuracy of ProBiS prediction in this work could be measured.

Prior to the prediction of conserved binding sites, we tested the capability of ProBiS to define the actual receptor binding sites on the protein. The results from ProBiS were then compared with those by other binding site predictors, such as Q-Site Finder,[35] Pocket Finder,[35,36] Binding Site tools in Accelrys Discovery Studio 2.5.5,[37] and ConSurf.[38] These methods use different concepts of binding site calculation, i.e., energy-based, geometric-based (Ligsite algorithm[36]), geometric-based (eraser algorithm[39]), and evolutionary sequence conservation, respectively. Our results showed that only ProBiS could accurately define and differentiate the actual receptor binding site (here noted as Site 1) from the rest of residues. Interestingly, ConSurf at the highest conservation grade unselectively identified the actual receptor-binding site together with many more false binding-sites on HA.

Moreover, ProBiS located additional two sites (Site 2 and Site 3) at the highest structural conservation degree around

Table 1. Representative PDB Structures of Each HA Subtype Used in Conserved Residues Prediction Using ProBiS

subtype	PDB and Chain ID	strain	complexed with	ref
H1	3AL4_KL	A/California/04/2009(H1N1)	–	42
	3M6S_GH	A/Darwin/2001/2009(H1N1)	–	43
	2WRG_HI	A/Brevigmission/1/1918(H1N1)	–	44
	2WRH_HI	A/Mallard/Alberta/35/1976(H1N1)	–	44
	3HTO_AB	A/Wdk/Jx/12416/2005(H1N1)	–	40
	3HTP_AB	A/Wdk/Jx/12416/2005(H1N1)	LSTa	40
	3HTQ_AB	A/Wdk/Jx/12416/2005(H1N1)	LSTc	40
	3GBN_AB	A/SouthCarolina/1/1918(H1N1)	FabCR6261	45
	1RVZ_AB	A/Puertorico/8/1934(H1N1)	LSTc	46
	1RV0_LM	A/Swine/Iowa/1930	DANA	46
H2	3KU3_AB	A/Japan/305/1957(Q226,G228)	–	47
H3	1EO8_AB	A/Aichi/1968(H3N2)	FabBH151	48
	2VIU_AB	A/X-31(H3N2)	–	49
H5	1JSM_AB	A/Duck/Singapore/3/1997	–	50
	2FK0_AB	A/Vietnam/1203/2004(H5N1)	–	51
	3FKU_AB	A/Vietnam/1203/2004(H5N1)	F10	33
	3GBM_AB	A/Vietnam/1203/2004(H5N1)	FabCR6261	45
	4FQI_AB	A/Vietnam/1203/2004(H5N1)	FabCR9114	52
H7	1TI8_AB	A/Turkey-Italy/2002	–	53
	3M5H_AB	A/Environment/NewYork/30732−1/2005(H7N2)	LSTa	54
	4DJ6_AB	A/Netherlands/219/2003(H7N7)		55
	4FQV_AB	A/Netherlands/219/2003(H7N7)	FabCR9114	52
H9	1JSD_AB	A/Swine/HongKong/9/1998	–	50
H14	3EYK_AB	A/Mallard/Astrakhan/263/1982(H14N5)		56

Helix A and Helix C at the stem domain (see the scheme of HA in Figure 1). The region around Site 2 (Helix C) is a known target for membrane fusion inhibitor, the currently only effective inhibitor against Group 2 of HA. However, ProBiS suggested that the conserved Site 2, which could become a new target for membrane fusion inhibitor across HA subtypes, is located slightly above the already known TBHQ binding site. Lastly, Site 3 (Helix A) has recently been targeted by a broad active antibody. Following this, we extended the prediction into the analysis of conserved antibody binding sites (epitopes) around Site 3 on HA across different strains and subtypes. The conservation of these amino acid residues are represented by the average of their structural conservation degrees in all of input strains and subtypes. Our results showed that the conserved antibody binding site around Site 3 is composed of S/A33, M/L317, T/V319, G320, W521, T541, Q542, I/F/V545, D/N546, I/T/V548, T/N549, V/L552, N553, and I/V556 residues that are located on chains A and B of HA's stem domain.

The three identified conserved binding sites revisited by ProBiS are expected to provide valuable insights for the experimentalists developing better drugs and vaccines against HA. Structurally conserved residues identified in Site 1 can be used for optimization of multivalent ligands or antibodies that could exploit these conserved residues for improved binding. Site 2 is proposed as a target for new membrane fusion inhibitors, especially to overcome HA's specific group problem. Finally, the results of our conservation analysis of Site 3 are of value for the development of antibodies with broad activity against various strains and subtypes.

METHODS

Four different binding site prediction tools collectively with ProBiS-2012 (http://probis.cmm.ki.si/)[2] were selected for the binding site prediction comparison analysis. These were Q-Site Finder (http://www.modelling.leeds.ac.uk/qsitefinder/),[35] Pocket Finder (http://www.modelling.leeds.ac.uk/pocketfinder/),[35,36] Binding Site tools of Accelrys Discovery Studio 2.5.5,[37] and ConSurf (http://consurf.tau.ac.il/).[38] These tools were selected because they employ different concepts to binding site prediction, such as explicit energy-based, geometric-based, and sequence conservation. Hemagglutinin (PDB ID 3HTP)[40] as the query protein was used as the test subject and was submitted into each predictor. Chain A and B of the protein structure were used as separated queries in ProBiS and ConSurf. Default prediction parameters in Q-Site Finder, Pocket Finder, and Binding Site tools of Accelrys Discovery Studio were used, whereby additional definition of heteroatoms as ligands was needed for both Q-Site Finder and Pocket Finder. As for ConSurf, prediction parameters were set according to the latest recommendation from its manual, e.g., UNIREF90 database with a number of putative homologous sequences of 150 and a CSI-BLAST E-value cutoff of 0.0001(http://consurf.tau.ac.il/faq.php). This was to ensure that the comparison was done without any specific preferences to avoid any bias.

In addition, ProBiS was used to further explore its potential capability in conserved epitope prediction (conserved antibody binding site). The collection of HA crystal structures from the PDB with different subtypes and strains were used. The representatives of these subtypes and strains are listed in Table 1. Twenty-four different HAs were submitted to the ProBiS-2012 prediction server (http://probis.cmm.ki.si/). Highly conserved residues as well as structural conservation degrees were determined using local structural alignment integrated within ProBiS algorithm.[3] Consensus structural conservation degrees were calculated by averaging the structural conservation degrees across 24 available HA structures. Finally, Connolly molecular surfaces in Accelrys Discovery Studio 2.5.5 were used to filter out those conserved residues that cannot be accessed

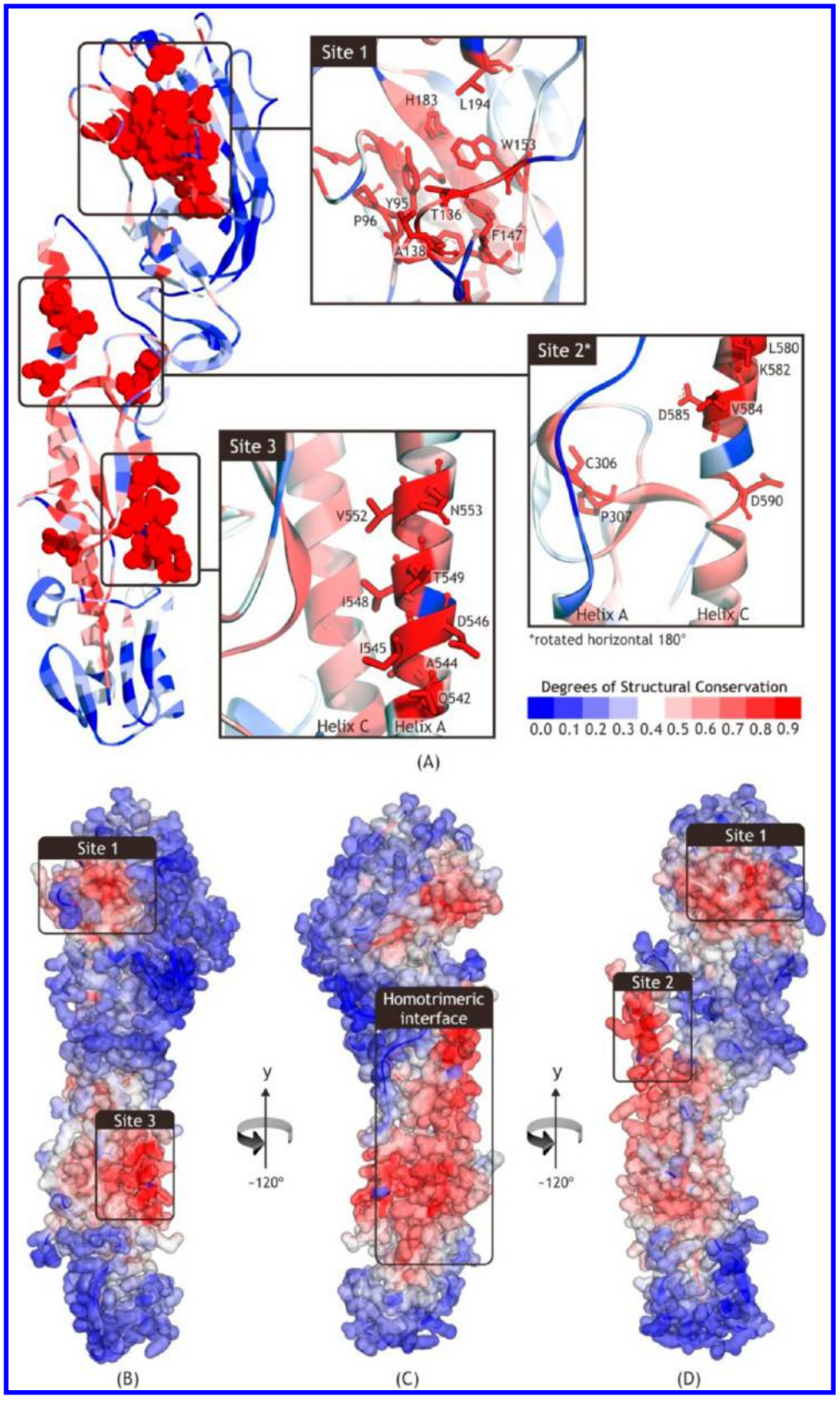


Figure 2. Structural conservation analysis of HA by ProBiS. (A) Three conserved sites located by ProBiS having the highest conservation degrees (score of 0.9) are represented as CPK models (left) and stick models (right). (B) Site 1 and Site 3. (C) Homotrimeric interface. (D) Site 1 and Site 2 are clearly distinct from other nonconserved regions when projected to the van der Waals surface of HA.

by the antibody molecule. All results were visualized and analyzed with the aid of Accelrys Discovery Studio Visualizer 3.5.[41]

RESULTS AND DISCUSSION

The crystal structure of HA (PDB ID 3HTP)[40] was selected as the input to the comparison study. This structure represents

dx.doi.org/10.1021/ci400421e | J. Chem. Inf. Model. 2013, 53, 2423−2436

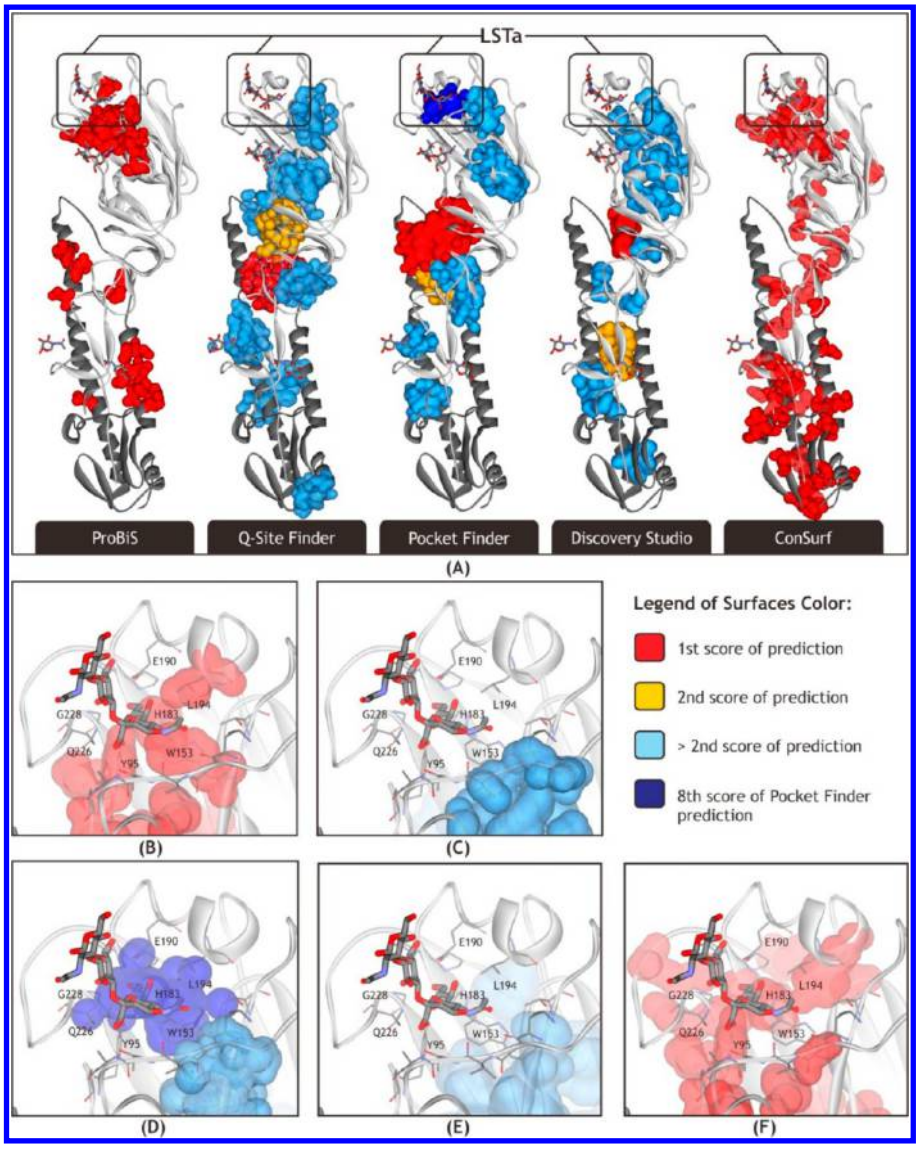


Figure 3. Prediction of receptor binding site of HA (3HTP). (A) Comparison of different binding site predictors: (B) ProBiS, (C) Q-SiteFinder, (D) Pocket-Finder, (E) Accelrys Discovery Studio, and (F) ConSurf. Red colored surfaces indicate the highest ranked prediction from each method followed by orange surfaces and cyan surfaces. Blue colored surfaces on (D) shows that Pocket-Finder finds the binding site only at rank 8, where rank 10 is the lowest.

the H1 subtype (from H1N1) complexed with its natural host receptor analog, Sialyl-Lactose-Tetraose a (LSTa). In ProBiS, the structural conservation degrees range from 0 to 0.9. Figure 2 shows that at the highest conservation degree, ProBiS identified three sites: Site 1 at the surrounding of the co-crystallized LSTa at the globular head domain, Site 2 at the region between Helix C of HA2 and HA1 at the stem domain, and Site 3 at the Helix A of HA2 at the stem domain. In addition, the homotrimeric interface at stem domain is also assigned relatively high conservation degrees.

Site 1: Receptor Binding Site. Figure 3 shows the results of the comparison among the selected binding site predictors. In general, the highest score is considered as binding site prediction (highest score represents the best confidence level). The predicted binding site observed in ProBiS is located at Y95 (Y98 in H3 numbering), W153, H183, E190, and L194 with the scoring ranging from 0.8 to 0.9 (Figures 3B and 4B). This is in agreement with Wiley and Skehel that stated such residues are largely conserved in many strains of influenza virus.[15] In addition, there are several residues detected at the high conservation score by ProBiS, i.e., C94 (C97 in H3 numbering), P96 (P99 in H3 numbering), F147 (P147 in H3 numbering), Y195, and R229 with the score ranging from 0.6 to 0.9. These conserved residues are located behind the receptor-binding pocket, and their role is to stabilize the binding pocket without contributing any interaction with the receptor.[15] ProBiS also identified the residues Q542, A544, I545, D546, I548, T549, V552, and N553 that are part of Helix A, L580, K582, V584, D585, D590, and D609 that are part of Helix C, and C306−P307 that are part of HA1 at the highest 0.9 conservation level. These residues are located at the stem domain.

As for Q-Site Finder, the highest-scoring predicted binding site is in between Helix C of the HA2 and HA1 domains, which consists of H299, P300, T302, I303, G304, E305, C306, P307, K308, Y309, T561, Q562, T564, V566, V584, D585, F588, and L589 (Figure 3A and C; Figure S1, Supporting Information). Q-Site Finder therefore failed to identify the actual receptor-

dx.doi.org/10.1021/ci400421e | J. Chem. Inf. Model. 2013, 53, 2423−2436

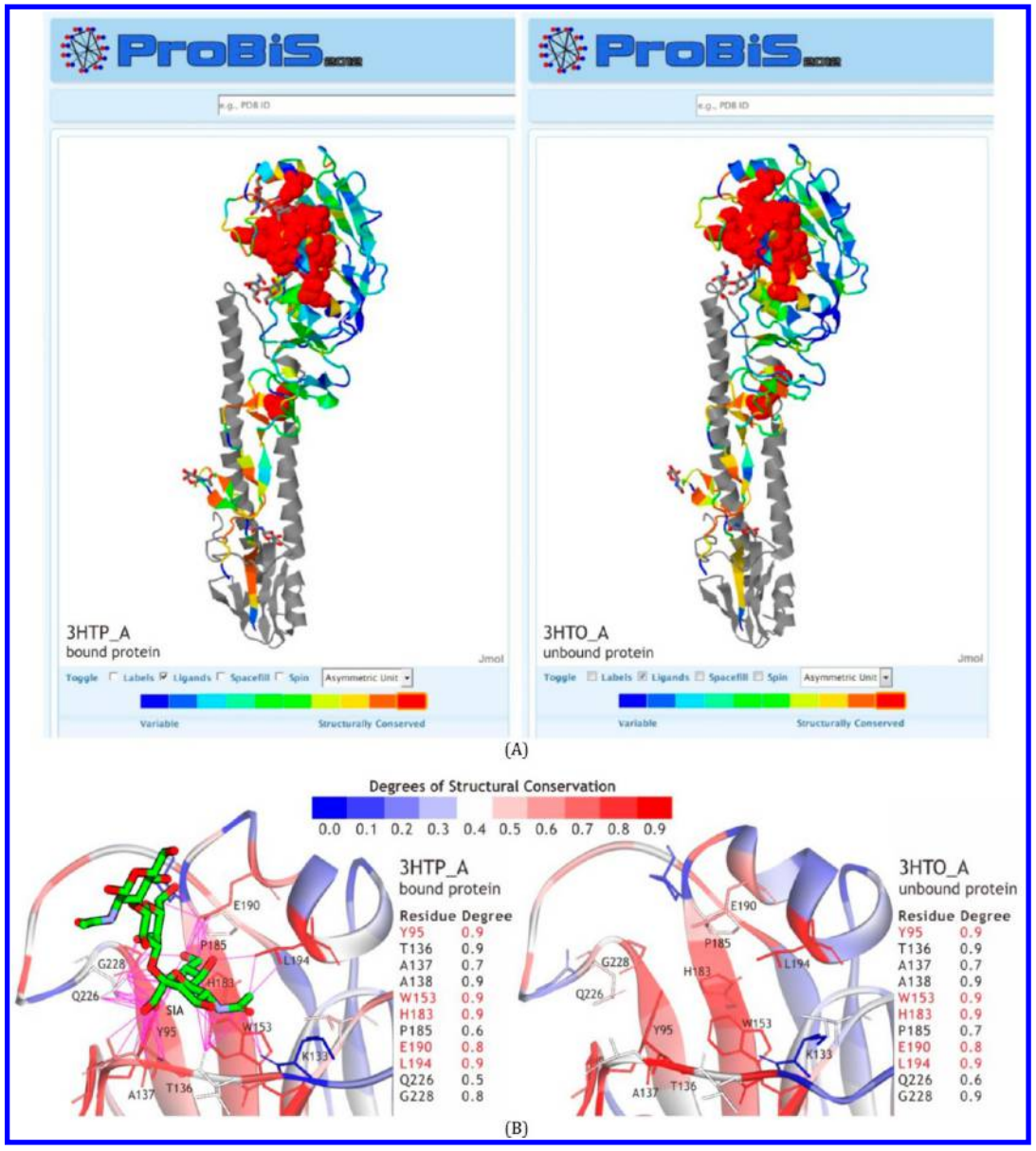


Figure 4. Receptor-binding site detection on bound (PDB ID 3HTP_A) and unbound (3HTO_A) HA with ProBiS. (A) HA residues are colored by the structural conservation degrees on the scale 0 (blue) to 0.9 (red). The sialic acid binding site, which is in the head globular region (top), is highly structurally conserved, and its residues are the red colored CPK model. (B) Conserved residues in a radius of 5 Å around the LSTa (green sticks model) are shown as sticks and are colored red. Purple thin lines indicate interactions between LSTa and the receptor.

binding site of HA. Q-Site Finder uses the concept of van der Waals surface interaction to probe the protein's interaction surface to find the energetically favorable interaction sites,[35] and the fact HA's receptor-binding site has a moderate activity (less energetic)[15] might explain why it failed to recognize this receptor-binding site.

Both Pocket Finder and Binding Site tools in Discovery Studio use a similar concept for binding site prediction by identifying the cavities (geometric-based) on the surface of protein. Pocket-Finder recognized the largest cavity with a volume of 496 Å^3 in between HA1 and Loop B−Helix C of the HA2 region, which consists of E103, E104, R106, E107, S110, D265, S266, G267, I268, I269, P300, I301, T302, I303, A565, V566, G567, K568, E569, F570, E578, N581, K582, D585, and D586. Interestingly, Pocket-Finder found a part of the actual binding site, which is made up of Y95 (Y98 in H3 numbering), W153, T155, H183, E190, L194, and Q226, but ranked it in eighth place with precision values of 75.4 with a cavity volume of 56 Å^3 (Figure 3D; Figure S2, Supporting Information). On the other hand, Binding Site tools in Discovery Studio failed to predict the receptor-binding site completely but recognized at the first rank a site in between loop B of the HA2 and HA1 domain, which consists of E103, R106, E107, S110, S266, G267, I268, I269, Q283, I301, T302, I303, A565, V566, G567, K568, and E569 (Figure 3E; Figure S3, Supporting Information). The receptor-binding site of HA is a shallow depression.[15,28] Pocket-Finder detects clefts by clustering probes around a protein's surface and correlates the probe cluster sizes with cleft sizes.[35] Shallow clefts may not accommodate large probe clusters, and thus, the detection of shallow binding sites is hindered.

Lastly, with ConSurf, the actual binding site was detected at the highest conservation score due to their high evolutionary conservation between the sequence homologues. However,

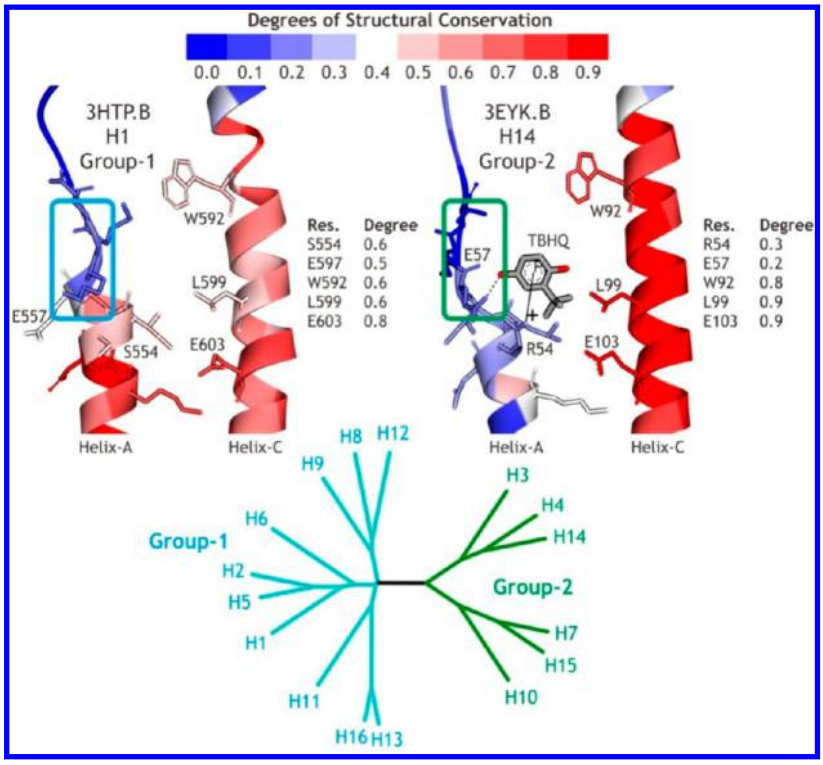


Figure 5. Conserved sites at positions 596−602 (H1 numbering) as predicted by ProBiS depicted as red colored ribbon and sticks. In Group 2 of hemagglutinin, this site is targeted by TBHQ, an inhibitor of membrane fusion, whereas in Group 1 HA, the additional turn of helix blocks TBHQ binding (compare the blue box with the green box).

ConSurf also detected many more conserved residues in the hemagglutinin sequence, e.g., D5, I7, C8, H12, N14, S16, V20, T22, V30, T31, C46, G67, P69, S81, E85, C94, Y95, P96, L105, G134, C139, F147, N170, W180, H183, H184, Y195, P215, R220, R229, C637, H642, and T656 (Figure 3F; Figure S4 and Table S1, Supporting Information). This made ConSurf less specific in defining the actual binding site compared to ProBiS at the highest score of prediction. All the numbering used here is based on H1 in PDB ID 3HTP. The conversion into the H3 numbering scheme is provided in Figure S5 in the Supporting Information. ConSurf, which uses sequence conservation, found the receptor-binding site at the highest conservation score, but it also predicted many other sites at this conservation score, thus making its prediction quite unspecific and unrealistic.

The output from 3HTP with 3HTO (LSTa bound and unbound HA, respectively) were compared in order to evaluate the independency of ProBiS prediction toward bound structure. Figure 4 shows how ProBiS defines the structurally conserved receptor-binding site residues in 3HTP and 3HTO (Figure S6, Supporting Information. The degrees of structural conservation for the receptor-binding site in each crystal structure are shown in Figure 4B. It is evident from the results that ProBiS has a high ability to define the receptor-binding site of HA as compared to the other methods used in the study. It is also demonstrated that regardless whether the structure is a complex holo structure (with the ligand) or apo structure (without ligand), ProBiS still locates the binding site distinctively.

Site 2: New Potential Target for Membrane Fusion Inhibitor. Upon examination of structural conservation analysis of HA by ProBiS (Figure 2A), we found two additional conserved sites at the HA stem region (Site 2 and Site 3). Site 2 is located around Helix C at an interface between HA1 and HA2. At the highest conservation degrees, this site is composed of residues C306 and C307 (from HA1 subunit) and L580, K582, V584, and D585 (from HA2 subunit).

The region of Site 2 is a target of tert-butylhydroquinone (TBHQ), an inhibitor of membrane fusion, as seen in the crystal structure 3EYK (3HTP is an H1, whereas 3EYK is an H14 subtype; these structures represent Group 1 and Group 2 HA, respectively).[56] Hemagglutinins in Group 1 have an additional turn of helix in Helix A (Figure 5), which renders the hydrophobic pocket in Helix C inaccessible to TBHQ and prevents its binding.[56] The group-specific character of this site is also indicated by the low structural conservation degrees of the upper part of Helix A, as shown in Figure 5 (blue colored ribbon). Vanderlinden et al.[57] proposed a novel class of membrane fusion inhibitors, and their best compounds have EC50 values of 3−23 μM. Unfortunately, these compounds only show activity against H3 and are inactive toward H1, H5, and H7. In addition, substitutions appear to make HA resistant toward these type of compounds (reviewed in ref 58).

On the other hand, Site 2 identified by ProBiS is located slightly above the known TBHQ binding site. This conserved site is composed of residues from both subunits (HA1 and HA2); therefore, ligand binding to this site could stabilize the HA structure and prevent dissociation during the membrane fusion process at the attachment step of the influenza virus onto the host receptor. For this reason, Site 2 may be suitable as a target for membrane fusion inhibitors. Interestingly, Q-Site Finder also detected this site at the first rank (shown in Figure 2A), which indicates that Site 2 may form energetically favorable interactions with ligands. Amino acid residues classified as the first rank in Q-Site Finder prediction are summarized in Table S1 of the Supporting Information. Therefore, we suggest Site 2 as a new potential target for the

dx.doi.org/10.1021/ci400421e | J. Chem. Inf. Model. 2013, 53, 2423−2436

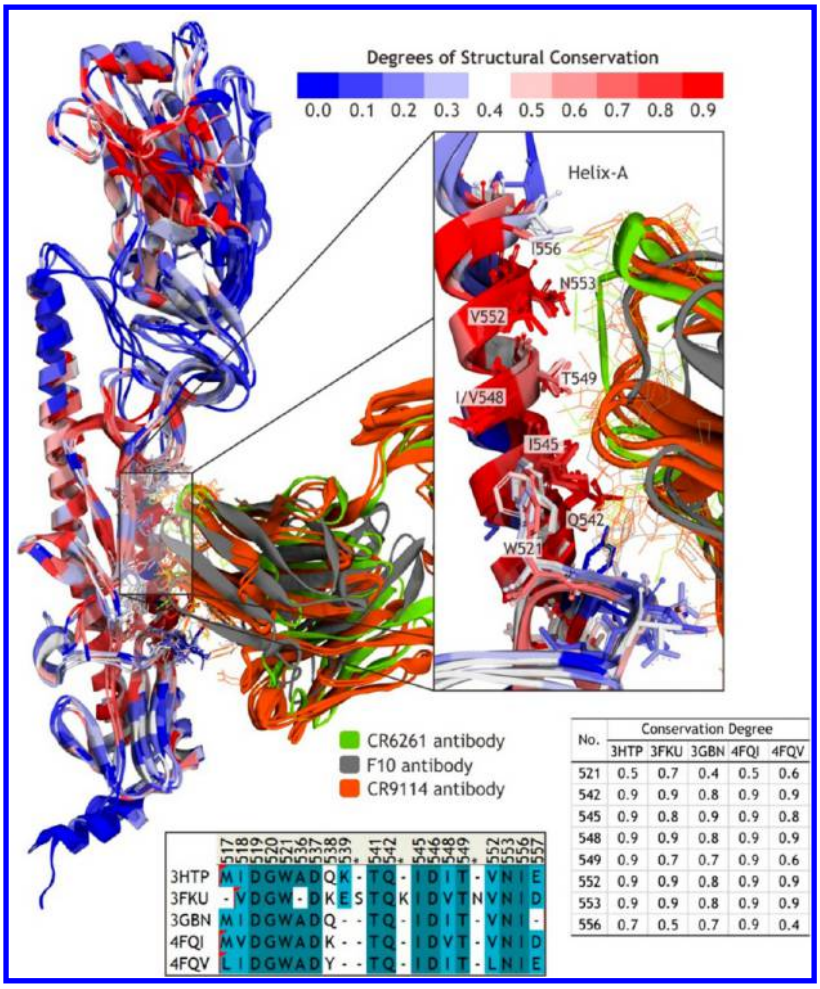
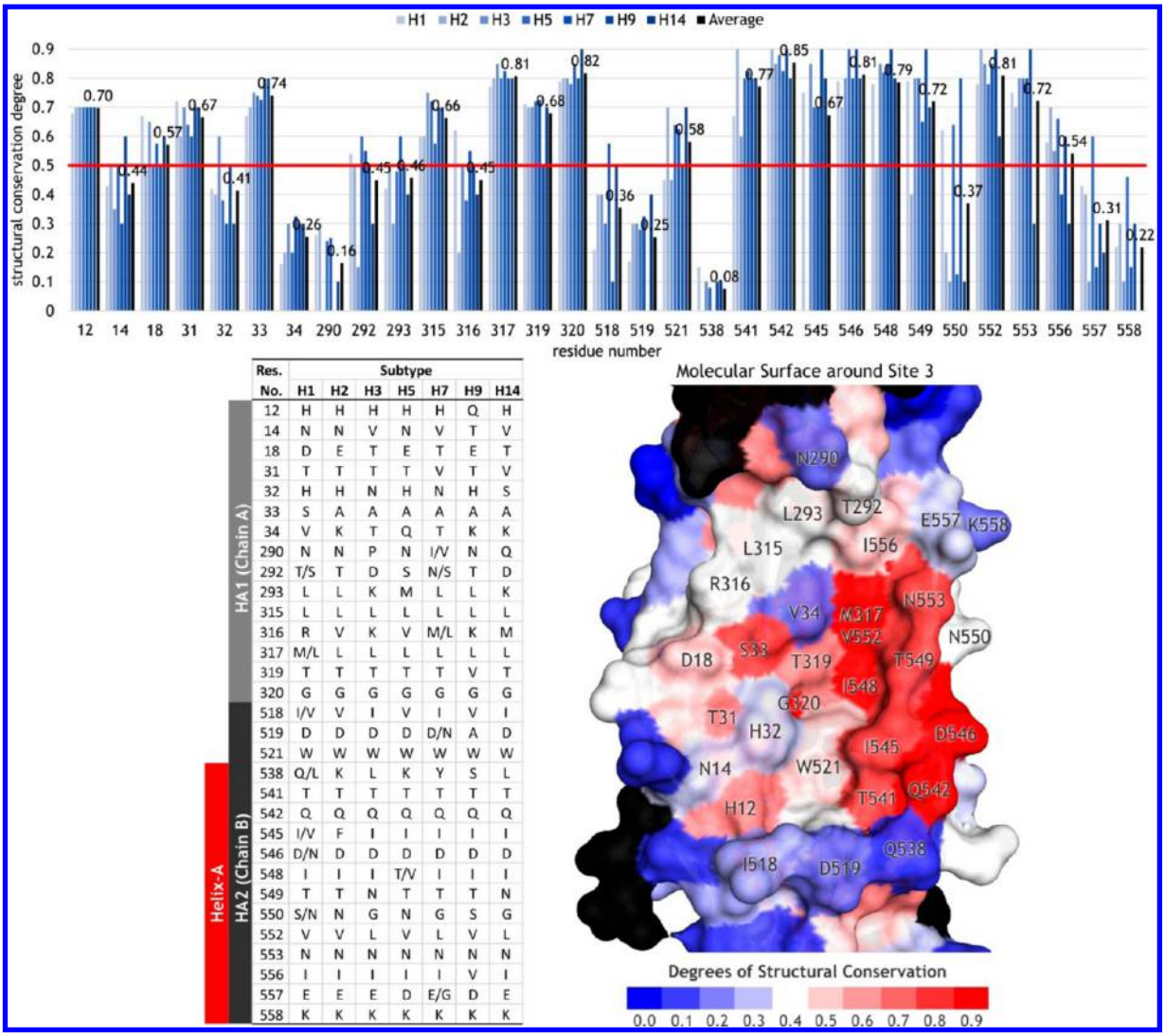


Figure 6. Superimposition of antibody-binding site located by ProBiS in the Helix A of stem region of HA (3HTP) with 3GBN (H1−CR6261 complex), 3FKU (H5−F10 complex), 4FQI (H5−CR9114 complex), and 4FQV (H7−CR9114 complex) shows that this site is conserved in H1, H5, and H7 subtypes. The degrees of structural conservation by ProBiS are shown in a blue−white−red gradient, with the red color indicating the highest degree of structural conservation.

development of new membrane fusion inhibitors, which may overcome the group-specific problem of current membrane fusion inhibitors.

Site 3: Target for Broad Effective Vaccine Development. The predicted Site 3 is composed of highly structurally conserved residues (score of 0.9), i.e., Q542, A544, I545, D546, I548, T549, V552, and N553, and is located at Helix A, consisting of residues 538−558 (Figures 1 and 4). Interestingly, this region is the target of recent broad effective antibodies such as CR6261, F10, and CR9114[33,45,52] as observed in the available crystal structures of the HA-antibody complexes, i.e., H1−CR6261 (3GBN), H5−F10 (3FKU), H5−CR9114 (4FQI), and H7−CR9114 (4FQV) complexes. The antibody binding site is located at the junction between Helix A of HA2 and the stem region of HA1. Superimposition of the calculated conserved residues of 3HTP with those of 3GBN, 3FKU, 4FQI, and 4FQV using ProBiS showed that residues in Helix A interacting with F10, CR6261, and CR9114 in 3HTP are highly conserved, i.e., Q542, I545, I548, T549, V552, N553, and I556 have degrees of structural conservation of 0.9, 0.9, 0.9, 0.9, 0.9, and 0.7, respectively (Figure 6). This result supports the fact that CR6261 and F10 antibodies are active against H1 and H5.[59] Besides H1 and H5, CR6261 is also predicted to neutralize H2, H4, H6, H8, H9, H11−H14, and H16.[45] In addition, CR6261 and F10 are more active against Group 1 HA than Group 2.[45,59] The latter CR9114 protects mice against a lethal challenge with influenza A and B virus.[52] However, although isoleucine at position 556 is perfectly conserved in multiple sequence alignment (Figure 6), the structural conservation degree by ProBiS in Group 2 member (H7 in 4FQV) is rather low (0.4). This indicates that the position of isoleucine in three dimensions is not conserved, most likely due to the loss of one helix turn in Helix A of Group 2 of HA. This explains the inefficacy of these antibodies against Group 2 HA.

Prediction of Conserved Antibody Binding Site. To attain a more comprehensive conserved antibody binding site around Site 3, the structural conservation degrees of all HA subtypes as listed in Table 1 were also calculated. Group 1 HA is represented by H1, H2, H5, and H9 and Group 2 by H3, H7, and H14. In general, similar to 3HTP, all HA subtypes demonstrated the presence of three conserved sites: receptor-binding sites at globular head (Site 1), proposed membrane fusion inhibitor binding site around Helix C (Site 2), and conserved antibody binding site around Helix A (Site 3). However, Helix A itself is not sufficient to form a proper binding surface for antibody molecule, as also suggested by the crystal structures of HA−antibody complexes, i.e., PDB ID 3GBN, 3FKU, 4FQI, and 4FQV (Figure 6). Hence, we



Res. No.	Subtype H1	H2	H3	H5	H7	H9	H14
12	H	H	H	H	H	Q	H
14	N	N	V	N	V	T	V
18	D	E	T	E	T	E	T
31	T	T	T	T	V	T	V
32	H	H	N	H	N	H	S
33	S	A	A	A	A	A	A
34	V	K	T	Q	T	K	K
290	N	N	P	N	I/V	N	Q
292	T/S	T	D	S	N/S	T	D
293	L	L	K	M	L	L	K
315	L	L	L	L	L	L	L
316	R	V	K	V	M/L	K	M
317	M/L	L	L	L	L	L	L
319	T	T	T	T	T	V	T
320	G	G	G	G	G	G	G
518	I/V	V	I	V	I	V	I
519	D	D	D	D	D/N	A	D
521	W	W	W	W	W	W	W
538	Q/L	K	L	K	Y	S	L
541	T	T	T	T	T	T	T
542	Q	Q	Q	Q	Q	Q	Q
545	I/V	F	I	I	I	I	I
546	D/N	D	D	D	D	D	D
548	I	I	I	T/V	I	I	I
549	T	T	N	T	T	T	N
550	S/N	N	G	N	G	S	G
552	V	V	L	V	L	V	L
553	N	N	N	N	N	N	N
556	I	I	I	I	I	V	I
557	E	E	E	D	E/G	D	E
558	K	K	K	K	K	K	K

Figure 7. Mapping of the average structural conservation degrees of all HA subtypes and strains listed in Table 1. The Connolly molecular surface of the region around Site 3 is shown with conserved solvent accessible residues indicated by red color.

extended analysis of the structural conservation degree to cover the residue radius of 10 Å around Helix A as well, excluding the homotrimeric interface, e.g., Helix C and Helix D (residues 592−623). The homotrimeric interface is assumed to be inaccessible by the antibody molecule due to its position at the interior of tertiary structure of homotrimeric HA.

The calculated average degrees of structural conservation from all HA subtypes mapped to the Connolly molecular surface of the HA are shown in Figure 7. The multiple sequence alignment and inset tables in Figures 6 and 7 show that the Helix A (538−558) residues share striking similarities. It is evident that amino acids in this region are highly conserved among the various subtypes and strains of influenza virus. As potential epitope for designed antibodies, we considered only amino acids that are conserved and solvent exposed (thus accessible to the antibody).[60,61] For this reason, we used the structural conservation degrees from ProBiS and solvent accessibility defined by the Connolly molecular surfaces as a filter. The qualitative threshold was set to >60% of the structural conservation degrees, i.e., score of >0.5. Residues with average structural conservation degrees of >0.5 that were below the Connolly molecular surface were considered to be inaccessible and were thus discarded. The bar chart in Figure 7 demonstrates that the amino acid residues in positions (3HTP numbering) 12, 18, 31, 33, 315, 317, 319, 320, 521, 541, 542, 545, 546, 548, 549, 552, 553, and 556 are qualitatively conserved. Therefore, we would like to propose these conserved binding site residues as potential candidates for conserved target in broad effective antibody design.

Analysis of the crystal structures 3GBN and 3FKU showed that the HA1 amino acid residues at positions 12, 14, 15, 31, 32, 34, 35, 36, 292, 293, and 319 participate in the interactions with both CR6261 and F10 (these residues are located in a radius of 5 Å from the antibodies). Among these, only the amino acids at positions 12, 31, and 319 are structurally conserved in all subtypes listed in Table 1 (Figure 7). Histidine at position 12 is conserved in most subtypes except in H9, where it is replaced by glutamine. Threonine is conserved at position 31, except in H7 and H14 subtypes, in which it is substituted by valine. Threonine at position 319 is conserved except in H9 subtype, where it is substituted by valine. At position 32, histidine and asparagine are for Group 1 and Group 2 HA, respectively (except serine in H14 subtype). Asparagine at position 32 (38 in H3 numbering) is subjected to N-linked glycosylation.[49,62] Interestingly, for Group 2 HA (H3 and H7, except H14), threonine occupies position 34 (note that unlike Group 1 HAs,

dx.doi.org/10.1021/ci400421e | J. Chem. Inf. Model. 2013, 53, 2423−2436

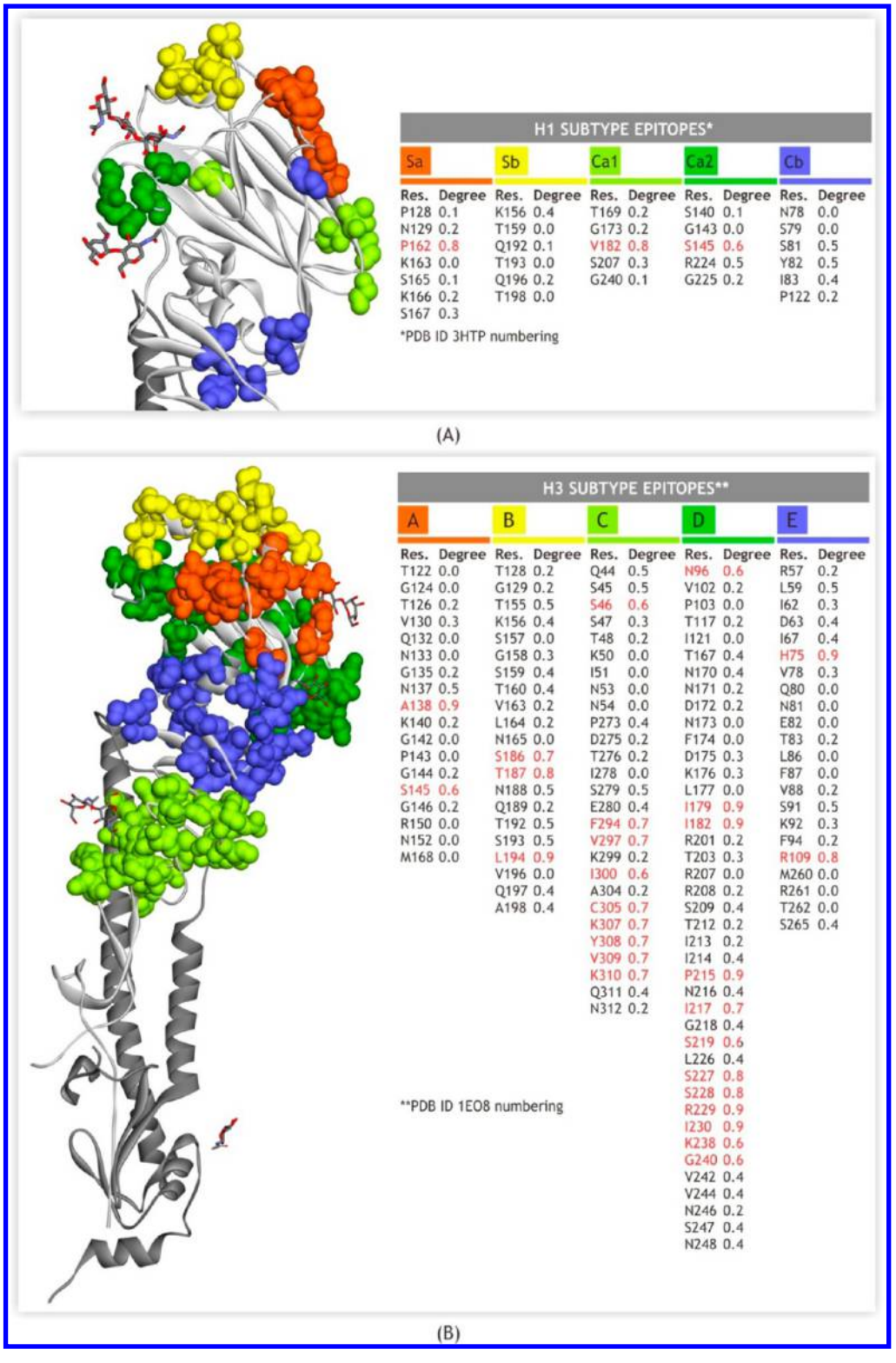


Figure 8. Mapping of structural conservation to epitopes of HA structures. Structurally conserved amino acid residues are colored red, and the structural conservation degree assigned by ProBiS is provided for each residue. (A) H1 subtype epitopes (Sa, Sb, Ca1, Ca2, Cb) mapped to the crystal structure PDB ID 3HTP. (B) H3 subtype epitopes (A–E) mapped to the crystal structure PDB ID 1EO8.

where position 32 is occupied by histidine, for H3 and H7, this position is occupied by asparagine), thus satisfying the N-linked glycosylation pattern of NxT (the pattern is NxS/T where x can be any residue except proline[63]). Hence, for Group 2 (except H14), this pattern most likely encourages N-linked glycosylation, thus making the region inaccessible to antibody binding.[45] It is therefore not surprising that CR6261 and F10 antibodies do not work as effectively toward the H3 subtype of HA.[33,45] In crystal structures 3GBN and 3FKU, residues 292 and 293 interact with phenylalanine and valine of CR6261 and F10 antibodies, respectively. Interestingly, ProBiS calculation showed that at these positions H3 and H14 have relatively

dx.doi.org/10.1021/ci400421e | J. Chem. Inf. Model. 2013, 53, 2423−2436

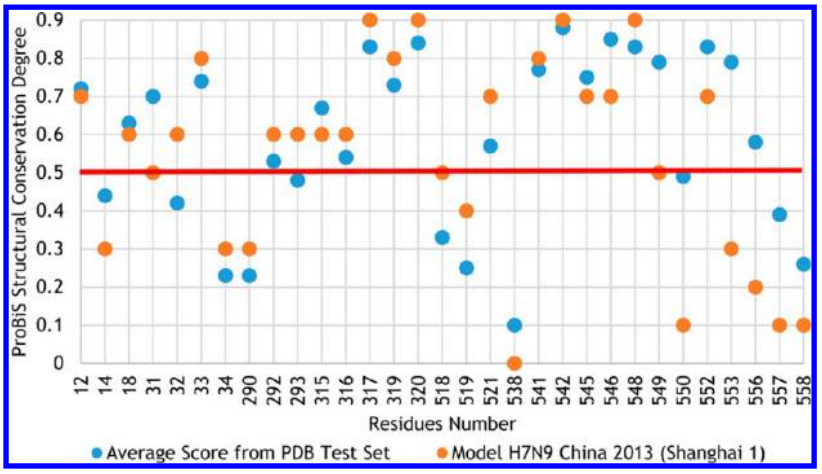


Figure 9. Structural conservation scores of the H7N9 model compared to the average scores from the PDB test set (Table 1). Threshold (0.5), above which we considered residues to be structurally conserved, is depicted by the red continuous line.

lower structural conservation degrees compared to others (Figure 7). At these positions, H3 and H14 have aspartic acid and lysine, whereas other HAs have uncharged residues, i.e., threonine, serine, asparagine, methionine, and leucine. Negatively charged aspartic acid cannot form favorable interactions with the π-electron-rich phenylalanine's ring side chain from the antibody molecule. Moreover, these charged residues (aspartic acid and lysine) may also form unfavorable interactions with the hydrophobic side chain of phenylalanine and valine of the antibody. This could provide extra explanation for the low affinity of CR6261 and F10 toward H3 subtype of HA.[33,45]

We propose that this reasoning is also applicable to the H14 subtype. Previously, Ekiert et al.[45] assumed that CR6261 likely neutralizes the H14 subtype due to the nonpresence of a glycosylation site on it. However, Corti et al.[64] found that H4, which is in close similarity with H14, was negatively neutralized by CR6261. This indicates that glycosylation at the position of 32 might not be the only factor affecting the negative affinity toward CR6261. Sequence alignment between H4 strain (A/duck/CZ/1956(H4N6)) and H14 in PDB 3EYK shows that in position of 293 (H1 3HTP numbering), H4 has lysine, identical to H14 (and H3). Hence, we propose that not only glycosylation but also the presence of charged residues at positions 292 and 293 cause low affinity of CR6261 (and F10) toward H3 and H4. Accordingly, we predict that the H14 subtype will likely also not be inhibited by antibodies for this reason.

Our data suggests that the structurally conserved amino acids that make up the epitope are S/A33, M/L317, T/V319, G320, W521, T541, Q542, I/F/V545, D/N546, I/T/V548, T/N549, V/L552, N553, and I/V556 (red surface in Figure 7). These results are in agreement with Ekiert et al.,[45] who analyzed 5261 sequences from H1 to H16 subtypes. That study showed that Helix A, which is part of the CR6261 antibody interacting surface (epitope), is highly conserved, consisting of T541, Q542, I545, D546, I548, T549, V552, N553, V555, I556, and E557 (3HTP numbering). We find that for designing a cross-reactive antibody, extra concern should be given to position 545, which according to our results, is not structurally conserved in H2 (bar chart in Figure 7).

In H2, a bulkier phenylalanine is at position 545 compared to isoleucine or valine in other subtypes. Although phenylalanine, valine, and isoleucine are all hydrophobic residues, the aromatic ring of phenylalanine is much bulkier than the aliphatic chain of valine or isoleucine. Moreover, the phenyl ring is electron-rich, which is unfavorable for interactions with negatively charged groups on the antibody molecule.

Epitope Mapping of Hemagglutinin. A great challenge in influenza treatment is to develop broad effective vaccines that are able to recognize a conserved epitope on hemagglutinin across the different virus subtypes. Thus, ProBiS was used in our attempt to detect such conserved potential antibody binding sites (epitopes) in a series of HA crystal structures. Out of three conserved regions in the HA structure identified by ProBiS (Figure 2A), we proposed as a potential target for universal vaccine development only Site 3. The antibody targeting this identified region should have a broader protection compared to those targeting other epitopes.

Hemagglutinin is the major surface antigen of the influenza virus. Antigenic sites in HA of the H1, H2, and H3 subtypes are characterized and are mainly located in the globular head of the HA1 subunit of the HA molecule.[28,65] Hemagglutinin has five antigenic sites in the H3 subtype, named A–E, whereas in the H1 subtype there are Sa, Sb, Ca_1, Ca_2, and Cb antigenic sites. Therefore, we collected the information regarding these five epitopes from Ndifon et al.[31] and Shen et al.[66] and mapped them to the PDB ID 1EO8 and 3HTP structures, respectively, as shown in Figure 8 (alignment of H3 to H1 numbering schema is shown in Figure S5, Supporting Information). Most amino acids in each epitope have low structural conservation degrees. Using the threshold of 60% of scoring (>0.5), only 3 out of 29 residues in H1 epitopes were found as relatively conserved. In H1 epitopes Sa, Ca1, and Ca2, only P162, V182, and S145 are structurally conserved, respectively. In Ca2 and Cb, only R224, S81, and Y82 have conservation scores of 0.5, whereas other residues are not conserved (Figure 8A). In H3 epitopes, more residues have high conservation degrees. Residues A138 (epitope A), L194 (epitope B), S228, R229, I230 (epitope D), and H75 (epitope E) have the highest structural conservation degrees (0.9) (Figure 8B). A138 and L194 are part of the conserved residues in the receptor binding site, while S228, R229, and I230 are located behind it, and their role is to stabilize the binding pocket.[27] Lastly, H75 interacts with the BH151 antibody in the 1EO8 structure. However, Figure 8A in general shows that the scattered structurally conserved residues in epitopes A–E do not make it into a specific conserved region, not like as seen in the receptor

dx.doi.org/10.1021/ci400421e | J. Chem. Inf. Model. 2013, 53, 2423–2436

binding site that consists of many conserved residues with high degrees of structural conservation ranging from 0.8 to 0.9.

These antigenic sites are recognized by virus-neutralizing antibodies preventing attachment of the virus to the receptor of the host cell.[15] Escape variants with mutation in the antigenic site easily avoid neutralization by existing host antibodies, leading to seasonal influenza outbreaks.[59] ProBiS, however, was unable to locate many conserved residues in these antigenic sites (or epitopes) at the HA1 globular head domains (Figure 8). This is unsurprising, as HA, particularly in the HA1 segment, is subjected to continual antigenic changes, and the degree of sequence diversity between the HA subtypes is substantial. In contrast, ProBiS successfully predicted the conserved binding site at Site 3, in the region of Helix A and the stem domain composed of HA1 and HA2. Unlike the antigenic sites in the head domain, antibodies targeting this site neutralize virus infectivity through inhibition of the fusion process.[67] High structural conservation of the HA2 domain makes it an attractive target for vaccine design.[25]

Submitting H7N9 Protein Model into ProBiS. To verify if the conserved residues at Site 3 are also present in HA of other influenza strains, we obtained the most recent novel H7N9 strain (A/Shanghai/1/2013(H7N9)) from the GISAID database (http://platform.gisaid.org/). Protein models for this strain were built using MODELLER.[68] Each Chain A (HA1) and Chain B (HA2) of the resulting model were submitted to the ProBiS-2012 server separately.

Resulting structural conservation degrees assigned by ProBiS were plotted in comparison with the average values from the entire PDB test set (Table 1), as shown in Figure 9. Most conservation degrees follow a similar trend, except for residues at positions 553 and 556. At position 553, all subtypes in our PDB test set have asparagine (Figure 7). At position 556, all except H9 have isoleucine. The difference in scores at position 553 supports the idea that in the structural conservation assignment, the different orientations/positions of the same residues result in low structural conservation degrees. As for position 556, the low conservation score in the H7N9 model resulted from the loss of one turn helix in Group 2 of HA. This is supported by the bar chart in Figure 7 showing that the structural conservation degrees of position 556 in Group 2 (H3, H7, H14) are relatively lower than those in Group 1.

The conserved amino acids in Helix A defined by ProBiS should be of value in designing new antibodies with broader affinity against HA subtypes. Recently, Fleishman et al. designed a protein that targets the conserved stem domain in HA based on the identified CR6261 epitope. The designed protein, HB80.3, was able to bind and inhibit multiple HA subtypes, similar to CR6261.[69] However, HB80.3, CR6261, and F10 were not able to inhibit the second group (Group 2) of hemagglutinin, e.g., H3 and H7.[45,69] Because of the conservation of the stem region, Steel et al.[67] also targeted this conserved stem for the design of a vaccine. They developed a "headless HA" that protects against lethal viral challenge. Furthermore, they suggest that the immune response induced by this headless HA is promising for further development toward a universal influenza vaccine.

Sequence analysis and structural bioinformatics have been used extensively in identifying potential residues as epitope candidates. Precise determination of conformational epitopes of neutralizing antibodies represents a key step in the rational design of novel vaccines. Knowledge of the epitope 3D structure can offer important insights into understanding virus neutralization, predicting epitope conservancy across different strains, and rationally designing new vaccine candidates. However, only 22 epitope−receptor complexes, which represent an average of 4% of all reported epitopes, have been determined up to 2010.[70] As more and more structures being deposited in the Protein Data Bank or other public domain databases, there will be opportunity for ProBiS to be applicable in therapeutic, diagnostic, and vaccine research. This work will thus serve as an example of how ProBiS can be used in a more comprehensive investigations of the antibody binding sites across various strains and subtypes of antigens.

CONCLUSION

As far as we are aware, this is the first time HA of various subtypes are subjected to local structural alignment against ∼30,000 3D structures in a nonredundant PDB database. The results show that ProBiS locates three conserved sites on HA's structure: Site 1, Site 2, and Site 3. Site 1 is identified later as the receptor binding site. Compared to other binding site predictors, ProBiS is the only one that accurately located the receptor binding site of HA. The geometrically shallow and energetically low affinity HA receptor binding site proves to be a challenge for other geometry- and energy-based predictors to detect this binding site at the highest prediction score. In locating Site 1, ProBiS is also more accurate than ConSurf, which is the tool that maps the evolutionary sequence conservation to the protein structure. Site 2 is located close above the known membrane fusion inhibitor (TBHQ) binding site. Therefore, due to its conservation, we propose this site as a novel conserved target for inhibiting the membrane fusion process, which might overcome the specificity problem of current inhibitor (TBHQ) that is ineffective against Group 1 of HA. Moreover, Site 2 may form favorable interactions with designed ligands, as identified by Q-Site Finder. Lastly, Site 3 is recently known as a target for broad effective antibodies, e.g., CR6261, F10, and CR9114. On the basis of the average of structural conservation degrees among various strains and subtypes of HA mapped to the molecular surface of HA, we propose residues that are structurally conserved and accessible to a potential antibody, i.e., S/A33, M/L317, T/V319, G320, W521, T541, Q542, I/F/V545, D/N546, I/T/V548, T/N549, V/L552, N553, and I/V556. These residues represent a conserved epitope across H1, H2, H5, H9 (Group 1) and H3, H7, H14 (Group 2) subtypes of HA and should be considered as targets for future antibody design. Further, by mapping the structural conservation degrees of known H1 and H3 epitopes, we suggest that these epitopes are not conserved among various strains/subtypes; therefore, designing the influenza vaccine against these epitopes could lead to a specificity problem. Finally, the test on the recent H7N9 China strain revealed that Site 3 is structurally conserved in this strain as well. We propose that the conserved region around Site 3 is the only potential target on HA for universal influenza vaccine development.

ASSOCIATED CONTENT

Supporting Information

Figure S1−S4: further details on binding sites predictions. Figure S5: conversion between H1 and H3 numbering. Figure S6: alignment of HAs by ProBiS. Table S1: predicted residues for each predictor. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: dusa@cmm.ki.si.

Author Contributions

#These two authors equally contributed to the paper.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

D.J. and J.K. acknowledge financial support through Grant P1-0002 of the Slovenian Research Agency. H.A.W. gratefully acknowledges the Malaysian Ministry of Science, Technology and Innovation for this work funded by Grant No. 09-05-lfn-meb-004. We also acknowledge the Laboratory of WHO Chinese National Influenza Center for sharing and submitting the sequence of A/Shanghai/1/2013 (H7N9) into the GISAID database.

REFERENCES

(1) Pérot, S.; Sperandio, O.; Miteva, M. A.; Camproux, A.-C.; Villoutreix, B. O. Druggable pockets and binding site centric chemical space: A paradigm shift in drug discovery. *Drug Discovery Today* **2010**, *15*, 656−667.

(2) Konc, J.; Janežič, D. ProBiS-2012: Web server and Web services for detection of structurally similar binding sites in proteins. *Nucleic Acids Res.* **2012**, *40*, W214−W221.

(3) Konc, J.; Janežič, D. ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment. *Bioinformatics* **2010**, *26*, 1160−1168.

(4) Konc, J.; Janežič, D. ProBiS: A web server for detection of structurally similar protein binding sites. *Nucleic Acids Res.* **2010**, *38*, W436−W440.

(5) Wong, M. T.; Choi, S. B.; Kuan, C. S.; Chua, S. L.; Chang, C. H.; Normi, Y. M.; Too, W. C.; Wahab, H. A.; Few, L. L. Structural modeling and biochemical characterization of recombinant KPN_02809, a zinc-dependent metalloprotease from *Klebsiella pneumoniae* MGH 78578. *Int. J. Mol. Sci.* **2012**, *13*, 901−917.

(6) Helal, M. A.; Chittiboyina, A. G.; Avery, M. A. New insights into the binding mode of melanin concentrating hormone receptor-1 antagonists: Homology modeling and explicit membrane molecular dynamics simulation study. *J. Chem. Inf. Model.* **2011**, *51*, 635−646.

(7) Jaroszewski, L.; Li, Z.; Krishna, S. S.; Bakolitsa, C.; Wooley, J.; Deacon, A. M.; Wilson, I. A.; Godzik, A. Exploration of uncharted regions of the protein universe. *PLoS Biol.* **2009**, *7*, e1000205.

(8) Musiani, F.; Bellucci, M.; Ciurli, S. Model structures of *Helicobacter pylori* UreD(H) domains: A putative molecular recognition platform. *J. Chem. Inf. Model.* **2011**, *51*, 1513−1520.

(9) Kar, G.; Keskin, O.; Nussinov, R.; Gursoy, A. Human proteome-scale structural modeling of E2−E3 interactions exploiting interface motifs. *J. Proteome Res.* **2012**, *11*, 1196−1207.

(10) Ashburn, T. T.; Thor, K. B. Drug repositioning: Identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discovery* **2004**, *3*, 673−683.

(11) Defranchi, E.; Schalon, C.; Messa, M.; Onofri, F.; Benfenati, F.; Rognan, D. Binding of protein kinase inhibitors to synapsin I inferred from pair-wise binding site similarity measurements. *PLoS One* **2010**, *5*, e12214.

(12) Vankayala, S. L.; Hargis, J. C.; Woodcock, H. L. Unlocking the binding and reaction mechanism of hydroxyurea substrates as biological nitric oxide donors. *J. Chem. Inf. Model.* **2012**, *52*, 1288−1297.

(13) Zhang, T.; Liu, L. A.; Lewis, D. F.; Wei, D. Q. Long-range effects of a peripheral mutation on the enzymatic activity of cytochrome P450 1A2. *J. Chem. Inf. Model.* **2011**, *51*, 1336−1346.

(14) Tomašić, T.; Kovač, A.; Klebe, G.; Blanot, D.; Gobec, S.; Kikelj, D.; Mašič, L. Virtual screening for potential inhibitors of bacterial MurC and MurD ligases. *J. Mol. Model.* **2012**, *18*, 1063−1072.

(15) Skehel, J. J.; Wiley, D. C. Receptor binding and membrane fusion in virus entry: The influenza hemagglutinin. *Annu. Rev. Biochem.* **2000**, *69*, 531−569.

(16) Johnson, N. P.; Mueller, J. Updating the accounts:Global mortality of the 1918−1920 "Spanish" influenza pandemic. *Bull. Hist. Med.* **2002**, *76*, 105−115.

(17) Potter, C. W. A history of influenza. *J. Appl. Microbiol.* **2001**, *91*, 572−579.

(18) Ellebedy, A. H.; Webby, R. J. Influenza vaccines. *Vaccine* **2009**, *27* (Supplement 4), D65−D68.

(19) WHO. Number of Confirmed Human Cases of Avian Influenza A(H7N9) Reported to WHO. http://www.who.int/influenza/human_animal_interface/influenza_h7n9/Data_Reports/en/index.html (accessed June 6, 2013).

(20) CDC. Avian Influenza A (H7N9) Virus. http://www.cdc.gov/flu/avianflu/h7n9-virus.htm (accessed September 6, 2013).

(21) WHO. Background and summary of human infection with influenza A(H7N9) virus− as of 5 April 2013. http://www.who.int/influenza/human_animal_interface/latest_update_h7n9/en/index.html (accessed September 6, 2013).

(22) Hien, T. T.; de Jong, M.; Farrar, J. Avian influenza–A challenge to global health care structures. *N. Engl. J. Med.* **2004**, *351*, 2363−2365.

(23) Grebe, K. M.; Yewdell, J. W.; Bennink, J. R. Heterosubtypic immunity to influenza A virus: Where do we stand? *Microbes Infect.* **2008**, *10*, 1024−1029.

(24) Palese, P.; Garcia-Sastre, A. Influenza vaccines: Present and future. *J. Clin. Invest.* **2002**, *110*, 9−13.

(25) Gerhard, W.; Mozdzanowska, K.; Zharikova, D. Prospects for universal influenza virus vaccine. *Emerging Infect. Dis.* **2006**, *12*, 569−574.

(26) Stanekova, Z.; Vareckova, E. Conserved epitopes of influenza A virus inducing protective immunity and their prospects for universal vaccine development. *Virol. J.* **2010**, *7*, 351.

(27) Wiley, D. C.; Skehel, J. J. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu. Rev. Biochem.* **1987**, *56*, 365−394.

(28) Wilson, I. A.; Skehel, J. J.; Wiley, D. C. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 [angst] resolution. *Nature* **1981**, *289*, 366−373.

(29) Matsubara, T.; Onishi, A.; Saito, T.; Shimada, A.; Inoue, H.; Taki, T.; Nagata, K.; Okahata, Y.; Sato, T. Sialic acid-mimic peptides as hemagglutinin inhibitors for anti-influenza therapy. *J. Med. Chem.* **2010**, *53*, 4441−4449.

(30) Gambaryan, A.; Tuzikov, A.; Pazynina, G.; Desheva, J.; Bovin, N.; Matrosovich, M.; Klimov, A. 6-Sulfo sialyl Lewis X is the common receptor determinant recognized by H5, H6, H7 and H9 influenza viruses of terrestrial poultry. *Virol. J.* **2008**, *5*, 85.

(31) Ndifon, W.; Wingreen, N. S.; Levin, S. A. Differential neutralization efficiency of hemagglutinin epitopes, antibody interference, and the design of influenza vaccines. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 8701−8706.

(32) Wilson, I. A.; Cox, N. J. Structural basis of immune recognition of influenza virus hemagglutinin. *Annu. Rev. Immunol.* **1990**, *8*, 737−787.

(33) Sui, J.; Hwang, W. C.; Perez, S.; Wei, G.; Aird, D.; Chen, L.-m.; Santelli, E.; Stec, B.; Cadwell, G.; Ali, M.; Wan, H.; Murakami, A.; Yammanuru, A.; Han, T.; Cox, N. J.; Bankston, L. A.; Donis, R. O.; Liddington, R. C.; Marasco, W. A. Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nat. Struct. Mol. Biol.* **2009**, *16*, 265−273.

(34) Tan, K. P.; Varadarajan, R.; Madhusudhan, M. S. DEPTH: A web server to compute depth and predict small-molecule binding cavities in proteins. *Nucleic Acids Res.* **2011**, *39*, W242−W248.

(35) Laurie, A. T. R.; Jackson, R. M. Q-SiteFinder: An energy-based method for the prediction of protein−ligand binding sites. *Bioinformatics* **2005**, *21*, 1908−1916.

(36) Hendlich, M.; Rippmann, F.; Barnickel, G. LIGSITE: Automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graphics Modell.* **1997**, *15*, 359−363.
(37) Accelrys Discovery Studio Modeling Environment, Release 2.5.5. Accelrys Software, Inc: San Diego, 2007.
(38) Ashkenazy, H.; Erez, E.; Martz, E.; Pupko, T.; Ben-Tal, N. ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **2010**, *38*, W529−W533.
(39) Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. LigandFit: A novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graphics Modell.* **2003**, *21*, 289−307.
(40) Lin, T.; Wang, G.; Li, A.; Zhang, Q.; Wu, C.; Zhang, R.; Cai, Q.; Song, W.; Yuen, K.-Y. The hemagglutinin structure of an avian H1N1 influenza A virus. *Virology* **2009**, *392*, 73−81.
(41) Accelrys Discovery Studio Modeling Environment, Release 3.5. Accelrys Software, Inc: San Diego, 2007.
(42) Zhang, W.; Qi, J.; Shi, Y.; Li, Q.; Gao, F.; Sun, Y.; Lu, X.; Lu, Q.; Vavricka, C.; Liu, D.; Yan, J.; Gao, G. Crystal structure of the swine-origin A (H1N1)-2009 influenza A virus hemagglutinin (HA) reveals similar antigenicity to that of the 1918 pandemic virus. *Protein Cell* **2010**, *1*, 459−467.
(43) Yang, H.; Carney, P.; Stevens, J. Structure and receptor binding properties of a pandemic H1N1 virus hemagglutinin. *PLoS Curr.* **2010**, RRN1152.
(44) Liu, J.; Stevens, D. J.; Haire, L. F.; Walker, P. A.; Coombs, P. J.; Russell, R. J.; Gamblin, S. J.; Skehel, J. J. Structures of receptor complexes formed by hemagglutinins from the Asian influenza pandemic of 1957. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 17175−17180.
(45) Ekiert, D. C.; Bhabha, G.; Elsliger, M.-A.; Friesen, R. H. E.; Jongeneelen, M.; Throsby, M.; Goudsmit, J.; Wilson, I. A. Antibody recognition of a highly conserved influenza virus epitope. *Science* **2009**, *324*, 246−251.
(46) Gamblin, S. J.; Haire, L. F.; Russell, R. J.; Stevens, D. J.; Xiao, B.; Ha, Y.; Vasisht, N.; Steinhauer, D. A.; Daniels, R. S.; Elliot, A.; Wiley, D. C.; Skehel, J. J. The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science* **2004**, *303*, 1838−1842.
(47) Xu, R.; McBride, R.; Paulson, J. C.; Basler, C. F.; Wilson, I. A. Structure, receptor binding, and antigenicity of influenza virus hemagglutinins from the 1957 H2N2 pandemic. *J. Virol.* **2010**, *84*, 1715−1721.
(48) Fleury, D.; Daniels, R. S.; Skehel, J. J.; Knossow, M.; Bizebard, T. Structural evidence for recognition of a single epitope by two distinct antibodies. *Proteins: Struct., Funct., Bioinf.* **2000**, *40*, 572−578.
(49) Fleury, D.; Wharton, S. A.; Skehel, J. J.; Knossow, M.; Bizebard, T. Antigen distortion allows influenza virus to escape neutralization. *Nat. Struct. Mol. Biol.* **1998**, *5*, 119−123.
(50) Ha, Y.; Stevens, D. J.; Skehel, J. J.; Wiley, D. C. H5 avian and H9 swine influenza virus haemagglutinin structures: possible origin of influenza subtypes. *EMBO J.* **2002**, *21*, 865−875.
(51) Stevens, J.; Blixt, O.; Tumpey, T. M.; Taubenberger, J. K.; Paulson, J. C.; Wilson, I. A. Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. *Science* **2006**, *312*, 404−410.
(52) Dreyfus, C.; Laursen, N. S.; Kwaks, T.; Zuijdgeest, D.; Khayat, R.; Ekiert, D. C.; Lee, J. H.; Metlagel, Z.; Bujny, M. V.; Jongeneelen, M.; van der Vlugt, R.; Lamrani, M.; Korse, H. J. W. M.; Geelen, E.; Sahin, Ö.; Sieuwerts, M.; Brakenhoff, J. P. J.; Vogels, R.; Li, O. T. W.; Poon, L. L. M.; Peiris, M.; Koudstaal, W.; Ward, A. B.; Wilson, I. A.; Goudsmit, J.; Friesen, R. H. E. Highly conserved protective epitopes on influenza B viruses. *Science* **2012**, *337*, 1343−1348.
(53) Russell, R. J.; Gamblin, S. J.; Haire, L. F.; Stevens, D. J.; Xiao, B.; Ha, Y.; Skehel, J. J. H1 and H7 influenza haemagglutinin structures extend a structural classification of haemagglutinin subtypes. *Virology* **2004**, *325*, 287−296.
(54) Yang, H.; Chen, L.-M.; Carney, P. J.; Donis, R. O.; Stevens, J. Structures of receptor complexes of a North American H7N2 influenza hemagglutinin with a loop deletion in the receptor binding site. *PLoS Pathog.* **2010**, *6*, e1001081.
(55) Yang, H.; Carney, P. J.; Donis, R. O.; Stevens, J. Structure and receptor complexes of the hemagglutinin from a highly pathogenic H7N7 influenza virus. *J. Virol.* **2012**, *86*, 8645−8652.
(56) Russell, R. J.; Kerry, P. S.; Stevens, D. J.; Steinhauer, D. A.; Martin, S. R.; Gamblin, S. J.; Skehel, J. J. Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17736−17741.
(57) Vanderlinden, E.; Göktaş, F.; Cesur, Z.; Froeyen, M.; Reed, M. L.; Russell, C. J.; Cesur, N.; Naesens, L. Novel inhibitors of influenza virus fusion: Structure−activity relationship and interaction with the viral hemagglutinin. *J. Virol.* **2010**, *84*, 4277−4288.
(58) Rungrotmongkol, T.; Yotmanee, P.; Nunthaboot, N.; Hannongbua, S. Computational studies of influenza A virus at three important targets: Hemagglutinin, neuraminidase and M2 protein. *Curr. Pharm. Des.* **2011**, *17*, 1720−1739.
(59) Wang, T. T.; Palese, P. Universal epitopes of influenza virus hemagglutinins? *Nat. Struct. Mol. Biol.* **2009**, *16*, 233−234.
(60) Nyambi, P. N.; Mbah, H. A.; Burda, S.; Williams, C.; Gorny, M. K.; Nádas, A.; Zolla-Pazner, S. Conserved and exposed epitopes on intact, native, primary human immunodeficiency virus type 1 virions of group M. *J. Virol.* **2000**, *74*, 7096−7107.
(61) Novotny, J.; Haber, E. Static accessibility model of protein antigenicity: The case of scorpion neurotoxin. *Biochemistry* **1986**, *25*, 6748−6754.
(62) Weis, W. I.; Brünger, A. T.; Skehel, J. J.; Wiley, D. C. Refinement of the influenza virus hemagglutinin by simulated annealing. *J. Mol. Biol.* **1990**, *212*, 737−761.
(63) Mellquist, J. L.; Kasturi, L.; Spitalnik, S. L.; Shakin-Eshleman, S. H. The amino acid following an Asn-X-Ser/Thr sequon is an important determinant of N-linked core glycosylation efficiency. *Biochemistry* **1998**, *37*, 6833−6837.
(64) Corti, D.; Voss, J.; Gamblin, S. J.; Codoni, G.; Macagno, A.; Jarrossay, D.; Vachieri, S. G.; Pinna, D.; Minola, A.; Vanzetta, F.; Silacci, C.; Fernandez-Rodriguez, B. M.; Agatic, G.; Bianchi, S.; Giacchetto-Sasselli, I.; Calder, L.; Sallusto, F.; Collins, P.; Haire, L. F.; Temperton, N.; Langedijk, J. P. M.; Skehel, J. J.; Lanzavecchia, A. A neutralizing antibody selected from plasma cells that binds to Group 1 and Group 2 influenza A hemagglutinins. *Science* **2011**, *333*, 850−856.
(65) Wiley, D. C.; Wilson, I. A.; Skehel, J. J. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. *Nature* **1981**, *289*, 373−378.
(66) Shen, J.; Ma, J.; Wang, Q. Evolutionary trends of A(H1N1) influenza virus hemagglutinin since 1918. *PLoS ONE* **2009**, *4*, e7789.
(67) Steel, J.; Lowen, A. C.; Wang, T. T.; Yondola, M.; Gao, Q.; Haye, K.; García-Sastre, A.; Palese, P. Influenza virus vaccine based on the conserved hemagglutinin stalk domain. *mBio* **2010**, *1*.
(68) Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779−815.
(69) Fleishman, S. J.; Whitehead, T. A.; Ekiert, D. C.; Dreyfus, C.; Corn, J. E.; Strauch, E.-M.; Wilson, I. A.; Baker, D. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **2011**, *332*, 816−821.
(70) Bui, H. H.; Peters, B.; Assarsson, E.; Mbawuike, I.; Sette, A. Ab and T cell epitopes of influenza A virus, knowledge and opportunities. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 246−251.