Grafting Methionine on 1F1 Ab Increases the Broad-Activity on HA Structural-Conserved Residues of H1, H2, and H3 Influenza a Viruses

Evolutionary Bioinformatics
Volume 17: 1–13

DOI: 10.1177/11769343211003082

Hoa Thanh Le[1,2], Phuc-Chau Do[1,2] and Ly Le[1,2,3]

[1]School of Biotechnology, International University, Ho Chi Minh City, Vietnam. [2]Vietnam National University, Ho Chi Minh City, Vietnam. [3]Vingroup Big Data Institute, Hanoi, Vietnam.

ABSTRACT: A high level of mutation enables the influenza A virus to resist antibiotics previously effective against the influenza A virus. A portion of the structure of hemagglutinin HA is assumed to be well-conserved to maintain its role in cellular fusion, and the structure tends to be more conserved than sequence. We designed peptide inhibitors to target the conserved residues on the HA surface, which were identified based on structural alignment. Most of the conserved and strongly similar residues are located in the receptor-binding and esterase regions on the HA1 domain In a later step, fragments of anti-HA antibodies were gathered and screened for the binding ability to the found conserved residues. As a result, Methionine amino acid got the best docking score within the −2.8 Å radius of Van der Waals when it is interacting with Tyrosine, Arginine, and Glutamic acid. Then, the binding affinity and spectrum of the fragments were enhanced by grafting hotspot amino acid into the fragments to form peptide inhibitors. Our peptide inhibitor was able to form in silico contact with a structurally conserved region across H1, H2, and H3 HA, with the binding site at the boundary between HA1 and HA2 domains, spreading across different monomers, suggesting a new target for designing broad-spectrum antibody and vaccine. This research presents an affordable method to design broad-spectrum peptide inhibitors using fragments of an antibody as a scaffold.

KEYWORDS: Influenza A virus, hemagglutinin, drug design, broad-spectrum, peptide inhibitor, antibody, conserved residues, grafting

RECEIVED: November 5, 2020. **ACCEPTED:** February 24, 2021.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 108.06-2017.332. Hoa Le and Phuc-Chau Do are funded by Vingroup Joint Stock Company and supported by the Domestic Master/PhD Scholarship Programme of Vingroup Innovation Foundation (VINIF), Vingroup Big Data Institute (VINBIGDATA) under the codes VINIF.2020.ThS.88 and VINIF.2020.TS.121, respectively.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Ly Le, School of Biotechnology, International University, Ho Chi Minh City 700000, Vietnam. Email: ly.le@hcmiu.edu.vn

Introduction

Influenza A virus (IAV) was responsible for the pandemics in 1918, 1957, and 1968, killing over 50 million people worldwide.[1] The recent pandemic in 2009 claimed 18 631 lives,[2] but the total casualty was estimated to be around 10-fold higher.[3] The effort to treat IAV is complicated by a high rate of mutation caused by antigenic shift and antigenic drift[4] that enable the virus to evade the host immune system and lead to drug resistance against effective treatment, for example, the Oseltamivir.[5] Furthermore, it is difficult to predict which strain of IAV will cause the next epidemic, given that IAV subtypes are classified based on 18 hemagglutinin (HA) subtypes and 11 neuraminidase (NA). These problems lead to a persistent attempt to predict the next mutation's characteristics and the annual development of new antibiotics to replace antibiotics ineffective against new mutations.[6]

Currently, HA has become an alternative to NA as a popular target for drug discovery and design. By attaching to the viral capsid, the HA receptor recognizes and binds to sialic acid decorated receptors of host epithelial cells. After binding to sialic acid, HA is cleaved into HA1 and HA2 to facilitate viral particle-cell fusion and penetration of IAV into host cells.[1] Since HA is highly expressed on the viral surface, it is an excellent target for drug design. However, drug design targeting HA is not straightforward as different subtypes of HA have demonstrated a high level of sequence variability.[7] As HA plays a crucial role in the ability of IAV to infect cells, we hypothesized that HA has to keep a portion of its structure conserved to maintain its significant biological role unchanged through countless events of mutation. This hypothesis is supported by many experimental attempts that managed to identify broadly neutralizing antibodies (bnAb) capable of binding to conserved regions by X-ray crystallography. Conserved regions were found in the receptor binding site[8] of the head or region of the stem. Clade neutralization was generated by antibody germline genes, which give rise to HA stem-directed bnAbs, such as V_H1-69,[9-12] V_H1-18,[10,13-15] V_H3-30,[16-18] V_H6-1,[15,19,20] and D_H3-9.[21] Some bnAbs exert cross-clade neutralization,[11,20] and the ability to develop cross-clade neutralization from group 1 or group 2-specific germline precursors was observed in 2 clonotypes from VH6-1 of the same donor,[22] indicating the possibility of cross-clade neutralization through convergent evolution and bnAbs obtaining a far broader spectrum against influenza A virus. A novel class of antibody effective against a broad range of group 1 IAV, established from Turkish patients who recovered from H5N1 avian flu infection, was identified from phage-displayed combinatorial libraries.[8] A group of antibodies was found to neutralize multiple influenza subtypes (H1, H2, H5, H6, H8, and H9).[23] CR6261, the most potent of these antibodies, protected mice injected with lethal H5N1 or H1N1. A highly conserved helical region in the membrane-proximal stem of HA1 and HA2 was confirmed to be the site

of recognition by CR6261.[24] A variety of other antibodies cross-linking to multiple subtypes were also discovered.[13,16,25,26] Therefore, a hypothesis can lay the foundation for drug discovery to achieve broad-spectrum anti-HA drugs targeting conserved regions, hoping to bypass the high divergence of HA structure among various IAV strains and guarantee its usability as anti-HA drugs in many years and over many strains.

A practical and cheaper alternative to antibodies, peptide ligands, presents promising therapy by retaining high affinity and selectivity and overcoming multiple delivery challenges such as cell permeability, low bioavailability.[27] Peptide inhibitors (PI) could be synthesized automatically in a high-throughput manner, reducing the time and cost of production compared to full antibodies.[28] Generally, some PIs derived from protein segments have shown potential applications in medicine, gene expression, and sensor.[29-31] In previous studies, peptides derived from the amino acid sequences of antibody were proven to bind to their designated targets with moderate strength ($K_D = 1.3\,\mu M$)[31] or influenza HA ($K_D = 56.8\,\mu M$).[28] In the scope of influenza, linear peptide inhibitors[28,32] and circular peptide inhibitor[32,33] were designed and experimentally proven to have broad-spectrum inhibition to HA regions conserved across multiple subtypes. These studies open up a new way to tackle evolutionary divergence in influenza HA, consisting of screening, design, optimization, and characterization of the binding positions and binding modes of peptides derived from neutralizing antibodies.

The search for the next antibody or PI candidates binding to HA conserved region can be effectively aided by computational methods that are gaining broader interest in recent years thanks to the understanding of the structure of the antibody[34] and recent improvement in docking methods. For example, up to 2012, the best success rate in CAPRI was 46% on average for top ten performers (the percentage was based on the number of cases for which there is at least a correct prediction within the 10 models submitted from each participating group).[35] In antibody-antigen docking, a success rate (defined as having Cα RMSD within <10 Å) (RMSD: root mean square deviation) of up to 80% in the top 10 clusters of the lowest energy score of conformation could be achieved by incorporating potential asymmetric DARS (Decoys as the Reference State) and non-CDR fragment masking (CDR: complementarity determining region) into PIPER docking algorithm.[36] Modeling antibody structure has been extensively developed. For example, antibody modeling was drastically simplified by discovering that most of the CDR loops of antibody (all but H3) fold in a limited number of conformations called canonical structures.[37] Consequently, canonical structures can be predicted based on loop length and key residues within or outside the CDR regions. Notable applications of computational methods include the in silico design of proteins HB38, HB80 (HB: HA binder) capable of recognizing a hydrophobic patch on the stem region of HA.[38] These molecules showed moderate to low binding affinity in empirical tests. Computation methods also expanded to vaccine design against HA. Computationally Optimized Broadly Reactive Antigen (COBRA), an antigen design methodology establishing multiple rounds of consensus sequences, was utilized to develop to elicit an immune response against H5 and H1 subtypes.[39,40] Later years, a "re-epitoping" method had been introduced by Guy Nimrod and colleagues for designing a reactive antibody on a targeting antigen.[41] Or a different approach on targeting the membrane fusion region of the virus to the host proposed 2 pentapeptides, which are optimistic in action as inhibitors.[32] The most recent research by our group has an iBRAB protocol to design a broad-spectrum Ab/Fab based on structures of available Ab against HA.[42] These researches proved that computational methods could be an effective instrument in designing potential therapeutic agents. Among therapeutic agents, antibodies hold the advantage of specific binding and eliciting immune response. However, there has rarely been any attempt to utilize computational methods to design antibodies targeting conserved regions of HA molecules.

Our research aims to screen and modify the fragment of antibodies that can bind to the conserved region on HA to construct PIs using computational methods. Our method was derived from a procedure to design a hemagglutinin inhibitor that has been proven successful,[38] with a few simplifications to reduce intensive computation. Therefore, our method can be routinely applied with low to average computational facilities. Such fragments can be CDR, framework regions, or any other parts of the variable regions of the anti-HA antibody that are found to make contact with HA. The scope of the target was limited to H1, H2, and H3 HA subtypes because they all caused pandemics in the past and have high transmissibility and mortality among the human population. Although H1 and H2 share different clade from H3, some bnAbs were discovered to recognize these 3 subtypes.[8,20] These facts imply structural features conserved across these subtypes.

Result

Structural alignment

Alignment by Matt resulted in 190 core residues, core RMSD of Cα (RMSD for short) 1.220 Å, raw score 440.673 for HA1; and, 32 core residues, core RMSD 0.999 Å, raw score 30.415 for HA2. Conserved residues, strongly and weakly similar residues were classified and visualized on ClustalX2 (see Supplemental Figure S1). These residues were listed in Table 1. The conserved residues were numbered according to the PDB fasta file of 1RD8 (1918 H1N1). Only residues totally conserved in all structures were chosen to be docked against disembodied amino acids in the next step. The conserved residues were mapped onto the surface of the H1 protein (PDB ID 1RD8) and visualized by UCSF Chimera (Figure 1).

Table 1. Conserved residues, strongly and weakly similar residues structurally aligned by Matt and classified by ClustalX2. Residues were numbered as in the 1RD8 sequence to take account of deletion and insertion in the 1RD8 sequence. This numbering scheme also applies to H2 and H3 HA.

DEGREE OF CONSERVATION	AMINO ACID	
	HA1	HA2
Conserved	L59, C64, L71, G72, P74, C76, W84, E89, C97, Y98, P99, D104, Y105, L108, S114, E119, W127, G134, A138, C139, F147, W153, Y161, P162, N170, L177, W180, G181, H183, H184, P185, Q191, Y195, V204, P215, I217, R220, P221, W234, T235, D241, G249, N250, L251, P254, F258	D112, F138
Strongly similar	I/V/L66, L/I70, N/D73, D/E77, I/F87, V/I/M88, R/K109, L/V112, S/A113, T/S136, Y/F148, L/M151, L/I154, S/T167, E/Q/D175, L/I179, V/I182, D/E190, N/Q197, V/I202, S/T206, R/K/Q211, E/N216, V/I223, M/I230, L/I236, L/V237, I/L243, F/I245, E/N246, A/S247, I/V252, L/M/I260	N/E/K117, L/V126, K/E131, I/F/M133, F/M/I140
Weakly similar	Q/H/E/D60, K/D/N63, N/S/T65, G/D68, E/H/Q75, S/D85, F/V102, P/T/A128, S/G/D146, N/R150, K/E/N/H156, S/N/T/K160, S/N/K165, S/D/E/K172, D/N/G/S199, N/S/Q210, A/G218, R/N224, G/S228, N/E/S231, K/E/D238, T/N248, A/V253	K/E/R121, Q/S125, N/S129, C/A/137

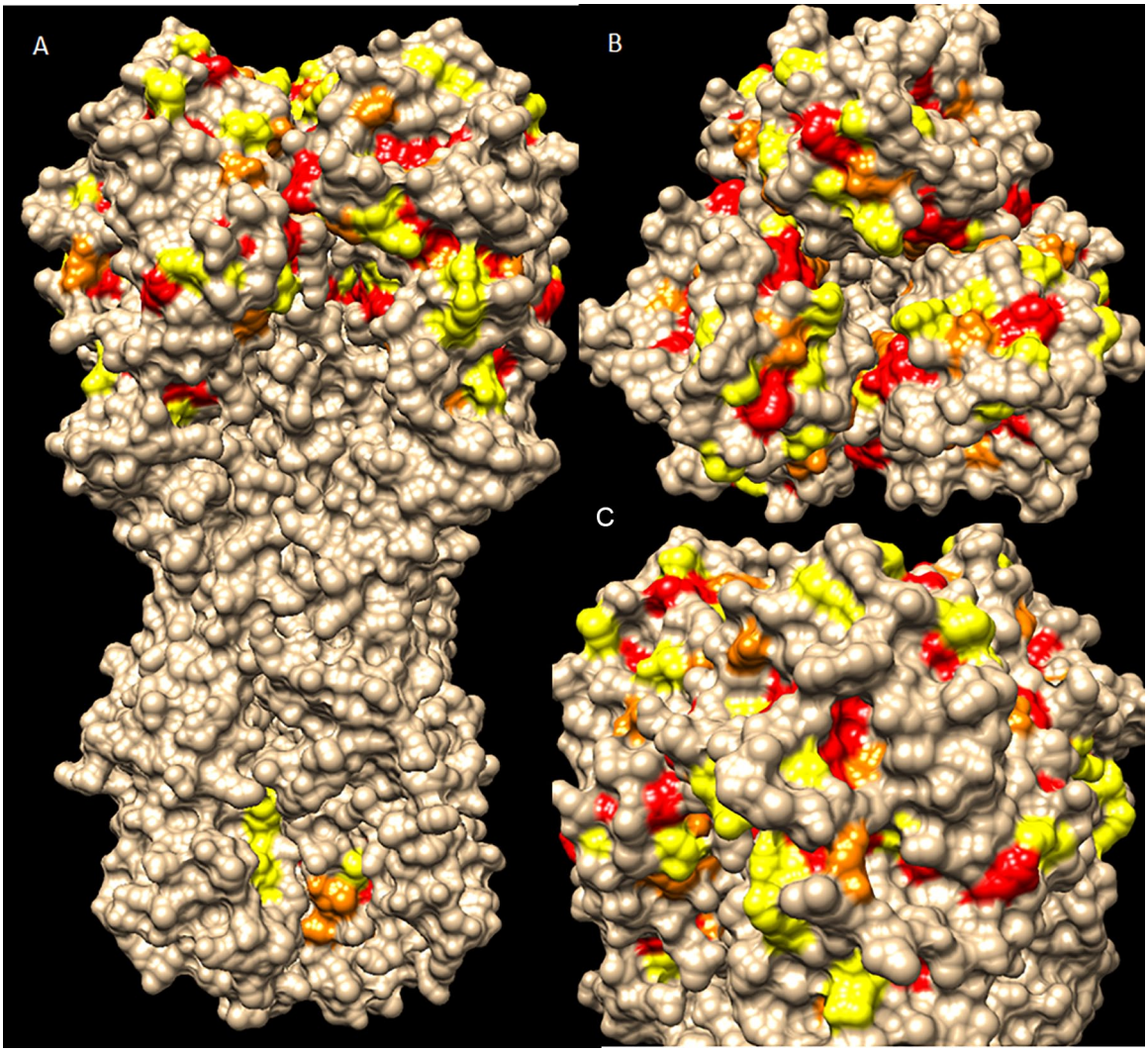

Figure 1. Conserved residues identified by structural alignment were mapped onto the surface of 1RD8 (H1) and visualized by UCSF Chimera, as seen from (A) the front, (B) the top, and (C) the rear. Conserved, strongly and weakly similar residues are labeled red, orange, and yellow, respectively. Most structurally conserved/similar residues lay on the HA1 domain.

Table 2. Hotspot residues were identified to form contact with conserved/similar residues on the HA structure. The hotspot residues were identified by docking disembodied amino acids to HA structure 1RD8. The corresponding conserved/similar residues in contact with hotspot residues and the docking score are listed in the second and third columns.

HOTSPOT	CONSERVED/SIMILAR RESIDUES	DOCKING SCORE
Methionine	Y105, R109	–218.3
Methionine	E89, R109	–213.1
Glutamic acid	N170, D241, K172	–175.2
Serine	C64, K63, Q60	–104.5

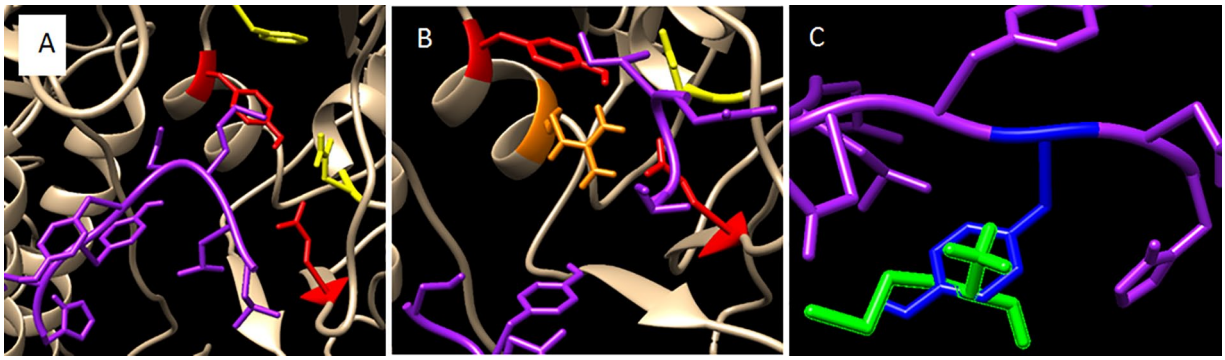

Figure 2. Antibody fragments of 4GXU (A) and 4M5Z (B) were found by docking to make contact with conserved/similar residues of HA. The antibody fragment is labeled purple, conserved, strongly, and weakly similar residues are labeled red, orange, and yellow, respectively. (C) Grafting of hotspot methionine (green) into antibody fragment on heavy chain M of 4GXU antibody. Tyrosine 100 A to be replaced by the hotspot methionine is marked blue.

Analysis of hotspot residues

Of twenty amino acids, 3 were found to gather in clusters and make weak or medium contact with at least 1 conserved residue and some physicochemically similar residues, identified by structural alignment, within Van der Waals radius of –2.8 Å (equivalent to the length of a row of 4 carbon atoms). The hotspot amino acids, nearby conserved or physiochemically similar residues and the corresponding docking score were listed in Table 2 (see Supplemental Figure S2 for the spatial position of the hotspot with respect to the conserved or similar residues).

Identifying antibody fragment scaffold and grafting hotspot residue into the scaffold

Two clusters of antibody fragments were found to bind close to conserved/similar residues on HA: 1 of 1F1 antibody (PDB: 4GXU) through contact with conserved amino acids Y105, F102, N65, and E89 and the other of 4M5Z through contact with conserved/similar amino acids Y105, N65, E89, and R109 (Figure 2A and B, Supplemental Table S2). A methionine hotspot fits the position of tyrosine numbered 100 A of the M chain of antibody heavy chain from 1F1 (Figure 2C). Tyrosine 100 A would be replaced by methionine hotspot as the hotspot would be grafted into the sequence of 1F1 antibody fragment.

Modeling antibody structures for grafted antibody fragment sequences

Modeled structures of 1F1 grafted antibodies achieved generally similar DOPE scores (Supplemental Table S3) with an average energy score of –21 815.2. The DOPE profiles followed a similar pattern and did not differ much from each other over the whole range of residues, except at the minima and maxima (Supplemental Figure S4). Ramachandran plots of all modeled antibody structures showed that most residues of each structure fell in the area of proper dihedral angles and low energy (Supplemental Figure S5).

Structural comparison of 5 grafted antibody models showed little deviation in structures and spatial orientation overall residues among modeled antibodies, modeled antibody fragments compared to the original antibody (Supplemental Figure S3), with 228 core residues, core RMSD 0.495 Å for whole antibody comparison; 8 core residues, core RMSD 0.149 Å for comparison among modeled antibody fragments.

Docking modeled antibody fragment structures to the HA

Within the top 20 clusters of lowest docking score, up to 5 clusters per PI were found binding to conserved residues of H1 HA close or directly to Y105. Every PI except PI.05 had at least 1 cluster bound directly to Y105. The clusters tend to

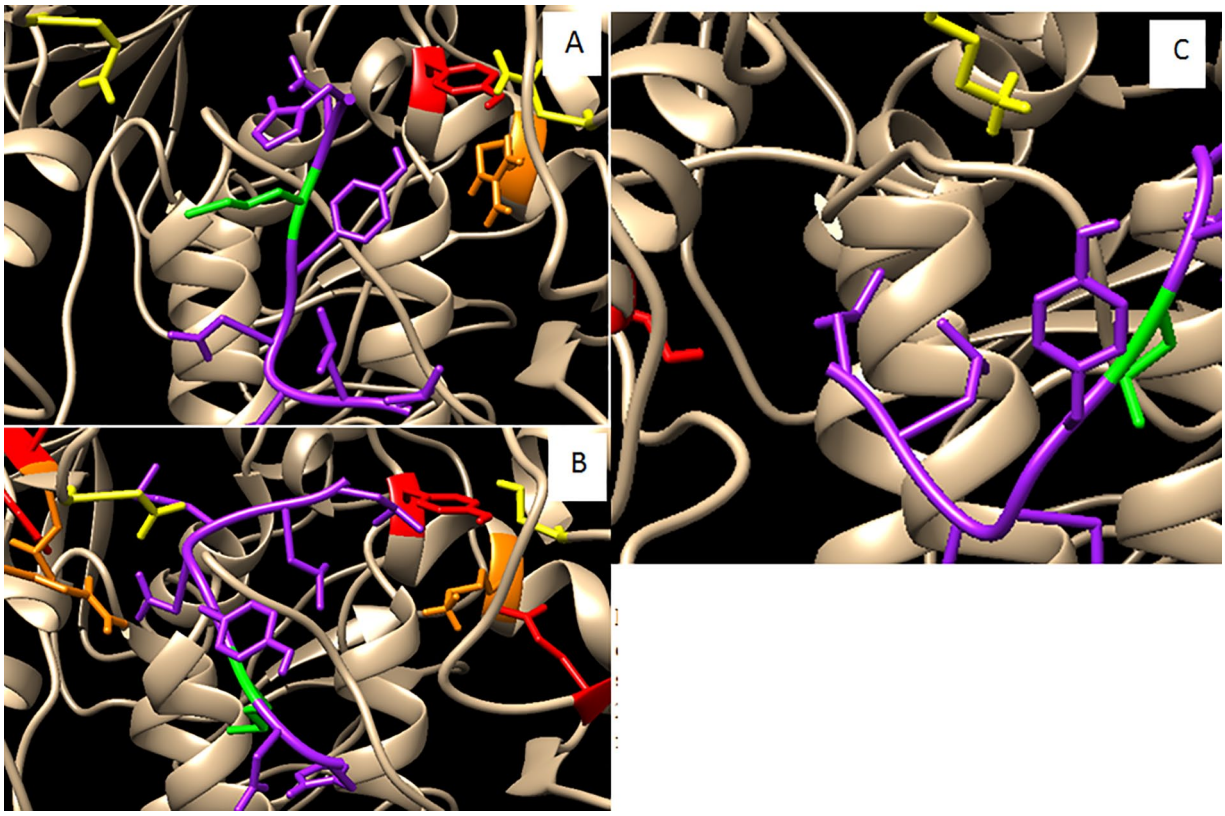

Figure 3. PIs of lowest docking score (purple) in contact with conserved (red), strongly (orange), and weakly (yellow) similar residues of H1 (A, 1RD8), H2 (B, 3QQO), H3 (C, 2HMG) HA. The hotspot methionine grafted into the antibody fragment scaffold is colored green.

Table 3. The docking result of clusters of PIs with the lowest docking score for each HA subtype. The PIs were found to form contact with conserved/similar residues on the structures of all 3 HA subtypes. The conserved/similar residues of HA structure in contact with PIs and the docking score are listed in the fourth and fifth columns.

HA SUBTYPE	PI	CLUSTER NUMBER	CONSERVED/SIMILAR RESIDUES	DOCKING SCORE
H1	PI.05	4	N65, Y105, R109, E238	–291.8
H2	PI.05	7	S65, E89, Y105, K109, Q175, W234, T235, L236, D238	–237.3
H3	PI.01	1	S114, K238	–282.8

gather around the region consisting of conserved residues N65, E89, F102, Y105, and R109. For H2 HA, up to 6 clusters per modeled antibody fragment bound either to conserved residues consisting of N65, E89, F102, Y105, R109, Q175, L236, and D238. PI.01 and PI.04 had 2 clusters that bound directly to Y105. In the H3 HA subtype, no contact between methionine and Y105 was found within the top 20 clusters of the lowest energy score. However, for each modeled antibody fragment, up to 12 clusters could be found binding to H3 HA at a position closed to conserved residues S114, D175, K238, and M260. The clusters of PIs with the lowest score of each subtype are shown in Figure 3 and Table 3. For detailed information on each PI, see Supplemental Tables S4 to S6.

Discussion

The rationale of our design scheme

Initially, our design scheme was intended to be different from other schemes so that our study may offer new data and insights. Several peptidic inhibitor design schemes have been carried out in recent years, such as the designed cyclic PI named P2 to P7 constructed by merging the HCDR3 loop of FI6v3 with the FR3 frame of CR9114 to target a highly conserved hydrophobic groove at the stem (HA2).[32,33] Both these schemes and ours utilized structural knowledge of pre-characterized antibodies and focused on targeting the regions on HA conserved across various subtypes. Nevertheless, our design scheme

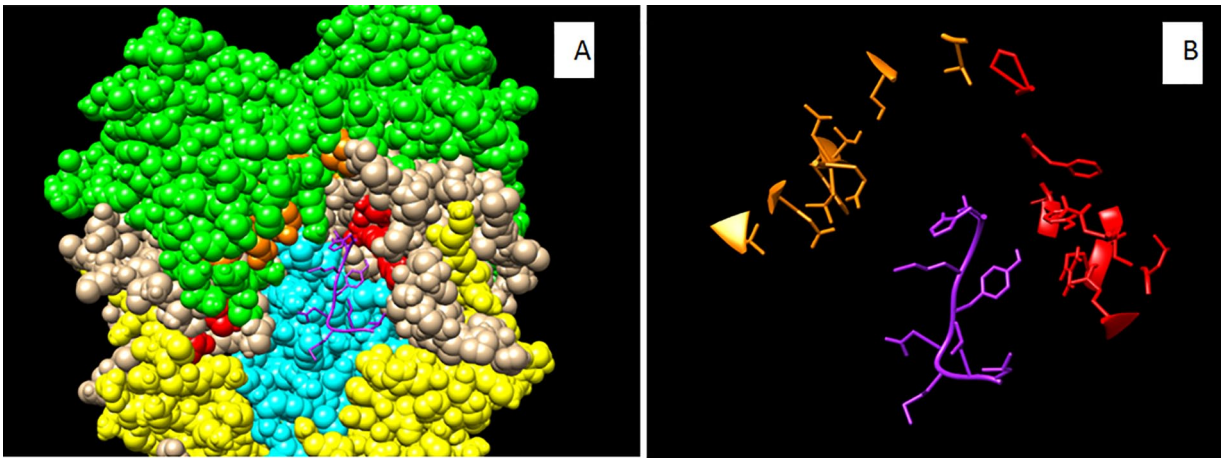

Figure 4. Binding position of all modeled PIs (here represented by PI.05 cluster 4) (purple) with respect to regions on HA structure (represented by 1RD8): receptor-binding site (green), HA1 fusion region (yellow), esterase (tan), and HA2 fusion region (cyan). The conserved/similar residues are visualized (A) with respect to HA regions and (B) separately and are marked red if they are located in the esterase region and orange in the receptor-binding site.

has a significant difference that may offer new data and new insights. While P2 to P7 were designed to bind to the conserved epitopes corresponding to their original bnAbs, we attempted to identify new potential conserved regions and screen for PIs from anti-hemagglutinin antibodies to other regions rather than our new conserved regions. We chose structure alignment on a set of HA structures rather than sequence alignment, which many previous studies did to identify new, structurally conserved regions. We aimed to discover new PI scaffolds by widening the sampling space of multiple anti-hemagglutinin antibodies' structures, which will include those epitopes and not overlap with any identified conserved regions. Then, the procedure was followed by hotspot grafting the epitope-binding sections of the antibody to increase the chance that the peptide inhibitor derived from that antibody can bind to novel conserved epitopes. Our method was not meant to be an alternative to the previous design schemes but complemented the original-bnAb-adhering schemes to attain the common goal: to achieve a universal influenza therapy based on HA conserved regions.

One advantage in the cyclic PI designation, as pointed by the authors, is the constraint of peptide backbone by cyclization, improving the peptidic ligands' performance against group 1 HA. More rigid cyclic peptide structures (made possible by incorporating nonproteinogenic amino acids) exhibited higher complex stability and enhanced binding and neutralization. We believe any form of structure stabilization (including cyclization) is desirable if the following 2s conditions are to be met: (1) the region to which the PIs were designed to bind has been identified to be conserved and a target for neutralizing antibody; (2) the designed PIs should bind only to this conserved region. Our structural alignment located new conserved regions that have not been observed to be the epitopes of any bnAbs, so there is no use in attempting to enhance the binding affinity of our PI to our newly found conserved regions through PI structure stabilization. The second condition was dropped from our consideration after the docking result showed that some conserved regions on the HA1 domain, which disembodied amino acids frequently clustered around, are close to each other, suggesting that our PI may also target conserved regions nearby provided the regions are also clustered with disembodied amino acids. This scenario is possible as our PI can bind to several structurally conserved regions located close to one another, depending on the subtypes of HA (Figure 4B). Because our conserved regions adopt different shapes, and the shape of a region may vary across different subtypes, structure stabilization might reduce the flexibility of PI structure, preventing PI from adapting to different shapes of the conserved regions nearby. This is shown to be true to our PI and its binding site on HA: the PI adapts different binding poses and orientations to accommodate the variation of the binding surface among different HA subtypes (Figure 5).

A twofold approach was used to make sure the PIs would bind to HA in the proximity of conserved residue: first, hotspot residues that were screened to determine the ones that could make contact with conserved residue were identified, and then antibody fragment scaffolds that could bind to the same conserved residue were identified. More clusters of PI within the top 20 clusters were found close to Y105 than the original antibody fragment scaffold (1F1) suggested that the PI had a higher chance to bind to Y105 than the original scaffold. Indeed, in terms of the number of clusters located close to Y105, the PIs outranked 1F1 by at most 6 to 1. The only difference between the PIs and 1F1 is the methionine hotspot, indicating that the hotspot played a role in bringing the PIs close to Y105. Therefore, the twofold approach was successful in the context of this research. Furthermore, this hotspot-grafting approach had been proven successful before.[38] It can be concluded that there is a certain degree of effectiveness in the twofold approach. The best docking scores of PI.05 to H1 and

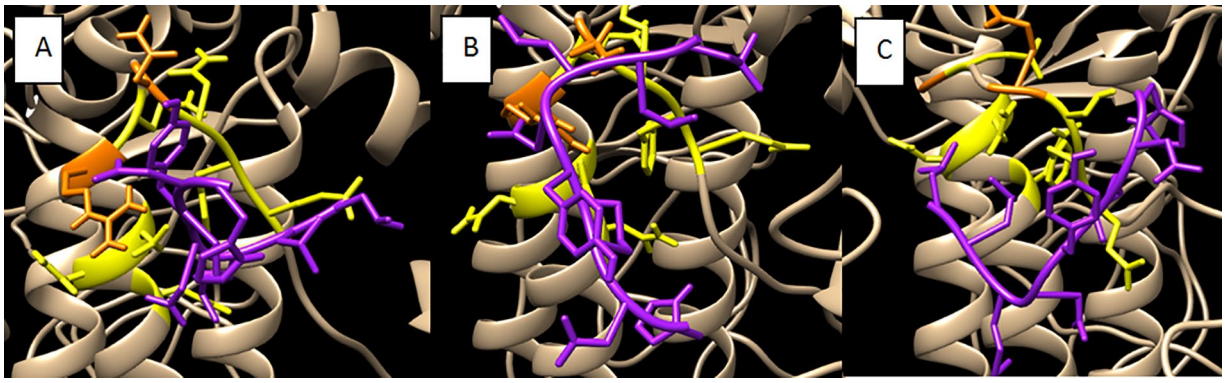

Figure 5. The binding site of PI.01 (purple) overlaps with the invariant residues (yellow) and synthetic lethal residues (orange) on the B loop that is proposed to drastically alter conformation during fusion. Notice the variation in the orientation of the side chains of invariant and synthetic lethal residues across (A) H1, (B) H2, and (C) H3 HA and the change in the orientation of PI.01 to adapt to different subtypes.

PI.01 to H3 were favorable as they were relatively close to or higher than the score of the original antibody fragment 1F1 (–287.7).

Moreover, the grafted scaffolds showed a significantly higher number of times (among top 20 clusters) of binding to HA to the site close to the arch-like cluster, and some grafted scaffolds showed a higher number of contact with conserved/similar residues than the original scaffold, regardless of H1, H2, or H3 subtypes (as compared between Supplemental Table S2 and S4-S6). For example, each grafted scaffold typically appeared 3 to 6 times among the top 20 clusters binding around the arch-like cluster of H2 (Supplemental Table S5) compared to 1 time of the original 1F1 antibody fragment (Supplemental Table S2), and each cluster of grafted scaffolds formed a contact with 4 to 12 conserved/similar residues (Supplemental Table S5) rather than 3 residues in the case of 1F1 antibody fragment (Supplemental Table S2, Supplemental Figure S9). These results suggest the hotspot residue could enhance the binding affinity to HA, the probability to appear among the top results and the breadth of recognition of HA of the original scaffold. By assessing docking results, 2 factors matter equally: docking score (representing a decrease in free energy) and the number of conserved/similar residues of HA in contact with PIs. The former was important as energy favorability is crucial to the formation and stability of the binding complex. The latter factor dictated the breadth of PIs to recognize conserved/similar residues across different subtypes. However, our docking result showed that these 2 factors did not necessarily go together (Supplemental Tables S4-S6). For example, in Supplemental Tables S4 and S5, most clusters binding to H1 HA generally had more favorable binding energy (reflected by the lower docking score) but fewer conserved residues in contact with them than the clusters binding to H2 HA. One possible explanation is that more conserved/similar residues recognized by the PIs cause the IPs to cover a wider area, and the energy landscape tends to vary across a wider area, resulting in less favorable binding energy.

In terms of modeling, the structures of our PIs generally adopted similar conformation to each other and the conformation of the original antibody fragment scaffold (1F1), as indicated by an RMSD of Cα 0.149 Å and lower than that of the whole modeled antibodies. The side chain of methionine hotspot of all PIs (PIs) showed different conformations but overall pointed outward from the modeled antibody's surface in 1 direction (Supplemental Figure S3C), allowing the hotspot to be accessible to the conserved residues on the HA molecule. Therefore, 5 modeled PIs can be viewed as 1 PI with flexible side chains. Ramachandran plots showed that most residues fell into the area of no steric clash. Otherwise, the grafted Y105M residue can be incorporated into the original 1F1 antibody as a way to improve the binding affinity and specificity of 1F1 to our new conserved residues. If the grafted 1F1 antibody is experimentally validated to be superior to the original antibody, we would gain a new neutralizing antibody for therapeutic use and immunological studies that identify new B-cell lines eliciting similar antibodies. Finally, the new conserved region to which improved 1F1 bind would be a potential candidate for broad-spectrum vaccine design. However, as a flexible macromolecule, the antibody would be better validated by molecular dynamics than rigid docking.

Identification of conserved regions using structural alignment

Different structural alignment algorithms adopt different mathematical approaches and emphases and produce different alignment results from the same batch of data. From this fact, it could be inferred that HA is more suitable for 1 algorithm than others. Conventionally, algorithms are tested by performing alignment with protein families derived from alignment databases. These databases contain groups of protein alignment that were manually curated and checked. Common databases for structural alignment include HOMSTRAD[43] containing the alignment of homologous families. SABmark,[44] despite its focus on sequence alignment, provides pairwise reference

structural alignment that can be used to benchmark structural alignment algorithms. However, no HA structures were found in HOMSTRAD, and a BLAST run on the whole SABmark database yielded only 3 HA structures, a sample size too small to afford good statistical power. To circumventing this problem, the algorithm was selected based on the performance test reported in past literature. Berbalk et al[45] tested 5 commonly used structural alignment algorithms: POSA,[46] MultiProt,[47] MASS,[48] MUSTANG,[49] and Matt[50]; using SISYPHUS,[51] a database whose protein members contain many features suitable for comprehensive test for the performance of structural alignment algorithms: circular permutations, segment-swapping, context-dependent folding or chameleon sequences that can adopt alternative secondary structures. Matt was ranked the best with the median values for the core alignment accuracy score of 75.83%.[45] Matt (short for "Multiple Alignment with Translations and Twists") utilizes the aligned fragment-pair chaining method and allows flexible protein backbone transformations, enabling the backbone to bend or rotate in order to produce optimal alignment.[50] Owing to its ability to make a structurally impossible twist in alignment, Matt is suitable for HA, which shows significant structural variation among IAV strains. Furthermore, Matt is a stable program; unlike other structural alignment algorithms, Matt was reported to cause no technical failures during runtime.[45]

As the sequences of HA are very diverse, whereas structures tend to be more conserved than sequences,[52] we believe that sequence alignment may not achieve a result that matches the result of structural analysis of HA in terms of secondary and tertiary structural levels. To combine structural alignment with scoring matrix from sequence alignment (Gonnet Pam250 matrix) of sequence alignment, our group performed structural alignment using 3D structural information, bring peptide segments of HA with similar secondary or tertiary structural characteristics close together to identify structurally conserved regions. Then, we analyzed the aligned amino acid sequences of the structurally conserved regions with the aid of ClustalX2 (using its scoring matrix and visualization) to locate residue positions that held physicochemically similar or conserved amino acids. Note that ClustalX2 was used not to perform sequence alignment but to visualize the degree of conservation (analyzed by the scoring matrix of ClustalX2) of each amino acid position.

The result of structural alignment showed that many conserved residues are buried deep in the core of the HA structure, particularly in the case of HA2, while some conserved residues of HA1 were located close to the outer environment but partially shielded by neighboring residues. Some residues belong to or are located close to regions of special interest, such as the receptor binding sites: A138, G134, T/S136 (130 loop); D/E190, Q191, Y195, N/Q197 (190 helix); P221, V/I223, R/N224, G/S228 (220 loop).[53] Such crucial positions showed that the conserved residues employed a clever tactic: by situating as deep into the core as possible, these residues could provide structural support to maintain the general structure of HA from inside while being shielded from the immune system and therapeutic agents.

Most conserved/similar residues were located in the receptor-binding region of H1 HA1 (Supplemental Table S7). Esterase ranked second in the number of conserved/similar residues. Coming next was the HA2 fusion and, finally, the HA1 fusion region. The location of some residues in correlation with special regions of HA implies that some conserved residues may play an active part in maintaining the function of HA. Conserved/similar residues that are also receptor-binding sites, as previously identified,[53] included G134, A138, W153, H183, P185, T/S136, and G/S228. D/E77 was also previously identified as a highly conserved residue that may affect receptor binding specificity.[53] This implication is also in agreement with the hypothesis that HA must keep a portion of its structure conserved to maintain its biological role. Some antigenic sites hold conserved or physicochemically similar residues: E19, W127, P162, S/T167, N/Q197, K/E/N/H156, S/D/E/K172, R/N224.[53] These sites hold promise as future therapeutic targets such as a universal vaccine.

A large portion of conserved residues was exposed on the surface of HA (Figure 1). These conserved residues tend to cluster together, creating small patches of conserved residues scattering throughout the head region, whereas a few patches are situated on the stem region. These patches could be a good target for the design of new anti-influenza antibodies. On the other hand, our conserved/similar residues tend to cluster in the head region rather than the stem, which implied a high level of structural conservation in the head domain and a low degree of structural conservation in the stem domain (Table 1, Figure 1 and Supplemental Figure S1). Nevertheless, the head domain usually undergoes a higher level of mutations to enable immune evasion of IVA while the stem is conserved across several pathogenic influenza A subtypes.[13] This discrepancy stems from the fact that past research focused on sequence alignment more than structural alignment, whereas structural alignment was emphasized in this research. Consequentially, the result of our structural alignment does not necessarily contradict with the result of sequence alignment of past literature; instead, our finding added a new light to the degree of conservation of HA2: HA2 is sequentially conserved (as stated in past literature) but structurally divergent (as found in our research). It means that HA2 of different subtypes may adopt different conformations. However, it should be noted that the number of HA structures used in this research was low. New structures need to be identified in the future to fully grasp the extent of structural conservation in the head and the stem region.

A new target for therapeutic design

The result of the final docking step between PI and HA structures showed that the grafted antibody fragment structures were able to bind to conserved residues of HA of 3 subtypes. Notably, many clusters were found binding to a recurrent set of

conserved/similar residues, such as N65, E89, F102, Y105, R109 of the H1 subtype; S114, D175, K238, and M260 of the H3 subtype and a combination of conserved regions in the H2 HA subtype. Such recurrent binding regions imply that the PIs have a high chance of being highly specific to these regions. The binding part on the H2 subtype is a combination of both H1 and H3 regions, suggesting that the binding region of H2 shares similar characteristics. Altogether, these binding regions comprise an arch-like cluster made of residues structurally conserved/similar among H1, H2, and H3. It should be noted that H1 and H2 HA belong to a different clade from H3 HA. Also, most known broad-spectrum antibodies were effective against only 1 clade, which meant they could only target either both H1 and H2, or H3 HA. Therefore, this arch-like cluster may be a potential target for new antibodies that can cross-react 2 clades.

One notable characteristic of the arch-like cluster is that it lies at the intersection of various regions of hemagglutinin: receptor-binding site (red), HA1 fusion region (yellow), esterase (orange), and HA2 fusion region (cyan) (Figure 4A) (the regions were identified based on H1N1)[53]; stretching from the right side of HA monomer (as viewed from the rear side) to the left side of the adjacent monomer (Figure 4B). This intersection is known to undergo a considerable conformational change upon virion-endosome fusion, for example, the shedding of HA1 and the reorganization of a buried loop into an exposed extended coiled-coil (also known as the spring-loaded mechanism).[54-56] Due to being composed of conserved residues and situated at a strategic area of fusogenic activity, the arch-like cluster may hold a significant role during fusogenic conformational change. This region can be a novel promising target for therapeutic agents, as no known broad-spectrum antibodies or inhibitors have bound to this area.

Interestingly, some PI orientations were found to simultaneously make contact with the arch-like cluster and a region consisting of the B loop connecting 2 helices (Figure 5A) which is proposed by the spring-loaded mechanism to undergo drastic conformational change during fusion.[55] A recent study employing over 10 000 HA sequences found multiple invariant residues and 2 lethal synthetic residues (represent mutations that are lethal to the virus when present together) in this spring-loaded region, forming a pocket resistant to mutation and an attractive target for small-molecule inhibitors.[57] Therefore, our PI may simultaneously bind to 2 conserved regions, one of the arch-like clusters and the other invariant residues. This region was a target for designing pentapeptide inhibitors. In silico validation method using molecular dynamics and quantum chemistry revealed that the pentapeptide blocked the spring-loaded mechanism by binding to the region around the B loop and altering the protonation states of key residues in the loop.[32] Taking together the arch-like cluster, the spring-loaded region, and encouraging results from the pentapeptide and our studies, we hypothesize the newly found arch-like cluster acts as a seal-able position. Any antibodies or PIs that could bind strongly to this conjoined area may act as a sealing tape that ties HA1 of different monomers, effectively blocking the shedding of HA1 and the exposition of the fusogenic HA2 domain.

A simple, light, and fast method for designing therapeutic peptides

The scaffold design method implemented by our group relies mostly on rigid protein docking. Although the downsides of rigid docking, compared to molecular dynamics, is that it discards the structural flexibility of proteins, it is less computationally intensive and can be performed all day, in short to medium amount of time and on average computers that can be found in the market at an affordable price. The encouraging results of our study prove that our design method using rigid docking can be used to conduct extensive screening or designing of small molecular candidates (such as therapeutic peptides of around 10 to 30 residues). Alternatively, rigid docking can be performed for a fast, rough global sampling of all possible docking poses, which will then be shortlisted for validation by molecular dynamics. Such an approach combines the advantages and mitigates the shortcomings of both methods.

Methods

The procedure of this article was derived from a procedure previously described.[38] In brief, HA structures were gathered for structural alignment to identify conserved residues of HA. Twenty types of amino acids (in disembodied form, disconnected from any protein structure) were docked against HA to determine the hotspot residues. Sample antibody fragments in complex with HA were gathered and docked against HA to determine the antibody fragment scaffolds that can bind to conserved residues. The hotspot residue would be grafted into that antibody fragment sequence. Finally, the grafted antibody fragment scaffolds were modeled to form PIs, which would then be docked against HA to check whether the hotspot residue made contact with conserved residues. The procedure was summarized in Figure 6.

Preparing HA structures and antibodies

HA structures were searched from Research Collaboratory for Structural Bioinformatics Protein Database (RCSB-PDB) with the following criteria: sequence length ranged between 100 and 450, title contained "hemagglutinin," the title did not have "complex," "antibody," "Fab," "bound," "epitope," and "receptor" to collect HA native structures which are not in complex with any other molecules. The PDB files were randomly collected and sorted into H1, H2, H3 groups. Any PDB files that merged chain A and chain B of HA into 1 single chain were left out due to the algorithm's limitation for structural alignment software (Matt) to distinguish protein chains, which will be explained in the next step. Antibody scaffolds were selected from antibody structures in complex with HA of

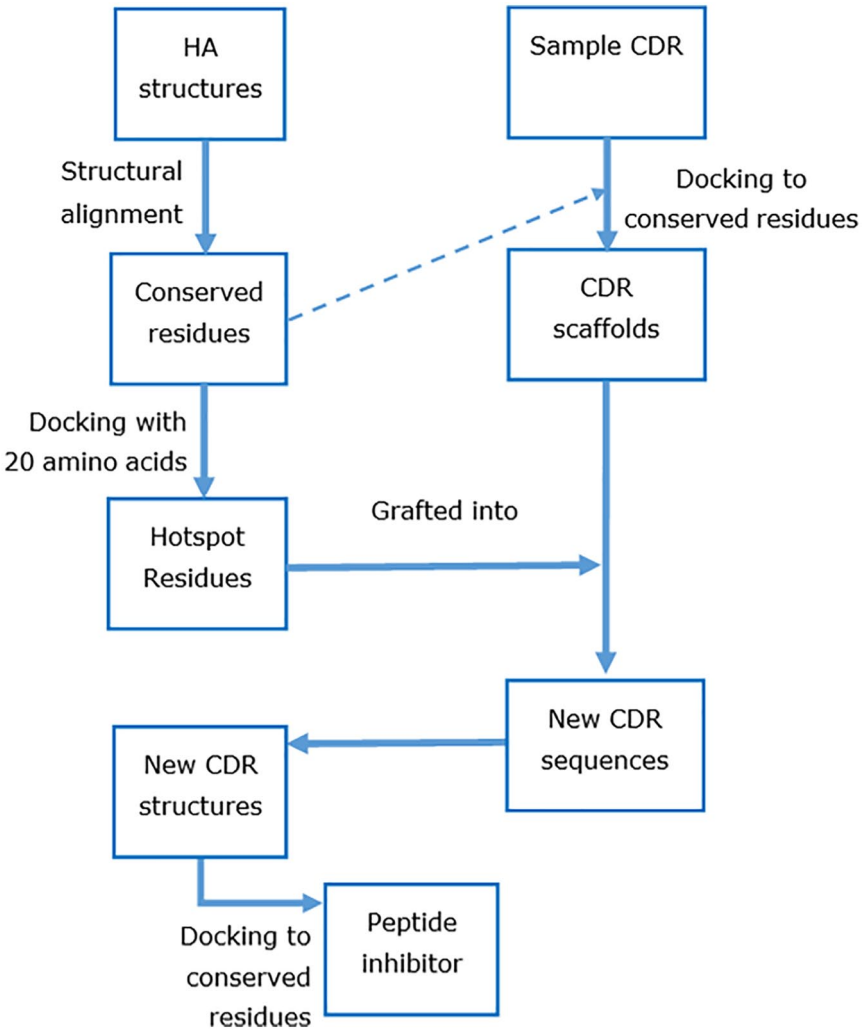

Figure 6. Overview of the design of PI to be employed in this paper.

H1, H2, and H3 subtypes from PDB. PDB was searched using the structure title "hemagglutinin" and "antibody." Supplemental Table S1 listed all PDB IDs of HA structures and anti-influenza antibodies used in this article.

Structural alignment to identify conserved residues

The structural alignment was preferred to sequence alignment because the former can exploit structural information, which tends to be better conserved than sequence.[52] This fact made structural alignment a good choice for proteins with high sequence variation among various subtypes like HA.

Before alignment, it had been decided to perform separate alignment on each HA chain, HA1 and HA2, because Matt rejected multiple-chain input; that is, Matt treated each protein chain as 1 different protein. As such, HA structures would not be aligned against each other as a whole, but rather, HA1 chains would be aligned against HA2 chains. An advantage of aligning separate domains is that domain-domain alignment might achieve a higher level of structural conservation, but not necessarily sequential conservation.

In the form of fasta, the alignment result was visualized by simply being loaded on ClustalX2[58] (no further analysis was needed in ClustalX2). ClustalX2 helped classify and visualize all residue positions into 3 degrees of physiochemical similarity: conserved residue (marked by asterisk "*"), strongly similar residue (":") and weakly similar residue ("."). Positions with conserved residue contained only 1 type of amino acid, while more than 1 type of amino acid was associated with strongly and weakly conserved residues. Strongly and weakly similar residues are groups with a positive matrix score based on the Gonnet Pam250 matrix,[59] which results in the residues in each group sharing similar physiochemical properties. In ClustalX2, the strongly and weakly similar groups were defined as having substitution scores higher than 0.5 and less or equal to 0.5, respectively. One implication from the Gonnet Pam250 matrix score is the degree of conservation of the physiochemical property, and thus the degree of conservation of each residue decreases from conserved to strongly and, finally, weakly similar residue. Only conserved residues were collected to identify hotspot residues to increase their reliability. When hotspot residue forming contact with a conserved

residue of HA structure was found, the hotspot residue would be further analyzed for contact with strongly and weakly similar residues nearby.

Identifying hotspot residue

Hotspot residues contribute to binding more than other residues in the same binding interface.[60] Each disembodied amino acid was docked against the HA structure using Hex 8.0.[61] Hex docked protein based on rigid-body, shape complementarity, and refines docked complexes by electrostatic potential. Docking parameters of Hex were set as default, with the exceptions listed as follows: correlation type "Shape+Electro" (corresponding to hydrophobic energy and electrostatic energy, respectively), Post Processing "OPLS Minimization" (Newton-like energy minimization and "soft" Lennard-Jones and hydrogen bond potentials adapted from the OPLS force-field parameters and explicit charge-charge electrostatic contribution); solutions 50 000; clustering RMSD was set 1.5 Å for docking small amino acids, and 3.0 Å for antibody scaffold, because docking of small molecules is more effective at smaller RMSD of Cα, as noted by the authors of Hex (for more information, see Hex user manual at http://hex.loria.fr/manual800/hex_manual.pdf). Crystal structure 1RD8 (1918 human H1 HA) was chosen as a receptor from this step and other docking steps. Docking results were narrowed down to the top 200 clusters, because the energy score of the 200th ranked cluster is usually 20% to 33% greater than the 1st ranked cluster, which has the lowest docking score. This shortlisting would increase the chance to collect clusters with a high probability of forming contact with HA structure in terms of free energy. Any disembodied amino acids that were either not located close to the conserved residues or not among the top 200 clusters were eliminated. The contact (polar, hydrophobic, or clash) between hotspots and conserved residues was analyzed by UCSF Chimera software,[62] with van der Waals radii set at −2.8 Å (equivalent to the length of 4 carbon atoms in a row).

Identifying antibody fragment Scaffolds

Since Hex cannot distinguish antibody fragment regions from other parts of the antibody structure, the antibodies would be docked against HA arbitrarily, irrespective of antibody fragment regions. To alleviating this problem, atomic coordinates of antibody fragment regions were extracted from the contact interface of HA-antibody complexes within a van der Waals radius of −2.8 Å with the help of UCSF Chimera. The extraction was done by copy-pasting the atomic coordinates of the antibody fragment (thus preserving the arrangement of atoms in the antibody fragment structure) from the PDB file of HA-antibody complexes into a separate, newly formed PDB file. Instead of docking the whole antibody, antibody fragment regions were docked against HA, as if non-antibody fragment regions of the antibody had been masked. Docking results were narrowed down to the cluster with the energy score that was up to 20% greater than that of the first ranked cluster.

Grafting hotspot residue into antibody fragment scaffold

When Hex found any antibody fragment region located close to any conserved residues, the atomic coordinates of antibodies, hotspot residues, and sample antibodies were merged into 1 single PDB file. The file was then inspected by the RasMol tool to see if any hotspot residues overlapped with residues of antibody fragment regions. If overlapping happened, new antibody fragment sequences would be formed by replacing those overlapped residues of antibody fragment regions with the hotspot residues.

Predicting the structure of the PIs

PIs were modeled from the grafted antibody fragment sequence based on homology modeling by Modeller.[63] The grafted hotspot residue was incorporated into the sequence of initial antibody structures from which the grafted antibody fragment sequence originated, then that antibody would be used as the template for modeling grafted antibody. The quality of the models was checked by comparing the DOPE score (Discrete optimized protein energy) by GNUPlot,[64] Ramachandran plot[65] of the Visual Molecular Dynamics (VMD) tool,[66] and structurally comparison by Matt. Finally, atomic coordinates of the structures of PIs were extracted from the antibody fragment of the grafted antibody structure to perform docking in the final step.

Docking PIs to HA

PIs were docked against the representative HA structure of H1, H2, and H3 (1RD8, 3QQO, and 2HMG, respectively). For each PI structure, twenty clusters with the lowest docking score were checked for interaction with HA structures by UCSF Chimera.

Conclusion

Our research managed to identify the hotspot residues and model grafted antibody fragment structures into peptide inhibitors with a high probability of contacting conserved or physiochemically similar residues on the HA structures of subtypes H1, H2, and H3. We drew attention to an arch-like cluster on HA composed of conserved/similar residues at the intersection of various regions of HA and stretching through the heads of adjacent HA monomers. This article consolidates the potential and affordable application of hotspot residues and antibody fragment scaffolds in designing PIs to treat other diseases.

Author Contributions

HL and LL conceived and designed the experiments. HL and LL performed the experiments and analyzed the data. HL, P-CD, and LL wrote the paper. All authors have read and agreed to the published version of the manuscript.

ORCID iD

Ly Le https://orcid.org/0000-0002-3182-0007

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Bouvier NM, Palese P. The biology of influenza viruses. *Vaccine*. 2008;26:D49-D53.
2. World Health Organization. Pandemic (H1N1) 2009 - update 112; 2010. Accessed April 6, 2020. https://www.who.int/csr/don/2010_08_06/en/
3. Simonsen L, Spreeuwenberg P, Lustig R, et al. Global mortality estimates for the 2009 Influenza Pandemic from the GLaMOR project: a modeling study. *PLoS Med*. 2013;10:e1001558.
4. Carrat F, Flahault A. Influenza vaccine: the challenge of antigenic drift. *Vaccine*. 2007;25:6852-6862.
5. de Jong MD, Thanh TT, Khanh TH, et al. Oseltamivir resistance during treatment of influenza A (H5N1) infection. *N Engl J Med*. 2005;353:2667-2672.
6. Smith DJ. Predictability and preparedness in influenza control. *Science*. 2006;312:392-394.
7. Mazzocco G, Łazniewski M, Migdał P, Szczepińska T, Radomski JP, Plewczynski D. 3DFlu: database of sequence and structural variability of the influenza hemagglutinin at population scale. *Database (Oxford)*. 2016;2016:baw130.
8. Ekiert DC, Kashyap AK, Steel J, et al. Cross-neutralization of influenza A viruses mediated by a single antibody loop. *Nature*. 2012;489:526-532.
9. Wrammert J, Koutsonanos D, Li GM, et al. Broadly cross-reactive antibodies dominate the human B cell response against 2009 pandemic H1N1 influenza virus infection. *J Exp Med*. 2011;208:181-193.
10. Nakamura G, Chai N, Park S, et al. An in vivo human-plasmablast enrichment technique allows rapid identification of therapeutic influenza A antibodies. *Cell Host Microbe*. 2013;14:93-103.
11. Lang S, Xie J, Zhu X, Wu NC, Lerner RA, Wilson IA. Antibody 27F3 broadly targets influenza A group 1 and 2 hemagglutinins through a further variation in V_H1-69 antibody orientation on the HA stem. *Cell Rep*. 2017;20:2935-2943.
12. Yamayoshi S, Yasuhara A, Ito M, Uraki R, Kawaoka Y. Differences in the ease with which mutant viruses escape from human monoclonal antibodies against the HA stem of influenza A virus. *J Clin Virol*. 2018;108:105-111.
13. Ekiert DC, Friesen RHEE, Bhabha G, et al. A highly conserved neutralizing epitope on group 2 influenza A viruses. *Science*. 2011;333:843-850.
14. Wu Y, Cho MS, Shore D, et al. A potent broad-spectrum protective human monoclonal antibody crosslinking two haemagglutinin monomers of influenza A virus. *Nat Commun*. 2015;6:7708.
15. Andrews SF, Gordon Joyce M, Chambers MJ, et al. Preferential induction of cross-group influenza A hemagglutinin stem–specific memory B cells after H7N9 immunization in humans. *Sci Immunol*. 2017;2:eaan2676.
16. Corti D, Voss J, Gamblin SJ, et al. A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. *Science*. 2011;333:850-856.
17. Fu Y, Zhang Z, Sheehan J, et al. A broadly neutralizing anti-influenza antibody reveals ongoing capacity of haemagglutinin-specific memory B cells to evolve. *Nat Commun*. 2016;7:12780.
18. Wyrzucki A, Dreyfus C, Kohler I, Steck M, Wilson IA, Hangartner L. Alternative recognition of the conserved stem epitope in influenza A virus hemagglutinin by a VH3-30-encoded heterosubtypic antibody. *J Virol*. 2014;88:7083-7092.
19. Joyce MG, Wheatley AK, Thomas PV, et al. Vaccine-induced antibodies that neutralize group 1 and group 2 influenza A viruses. *Cell*. 2016;166:609-623.
20. Kallewaard NL, Corti D, Collins PJ, et al. Structure and function analysis of an antibody recognizing all influenza A subtypes. *Cell*. 2016;166:596-608.
21. Wu NC, Yamayoshi S, Ito M, Uraki R, Kawaoka Y, Wilson IA. Recurring and adaptable binding motifs in broadly neutralizing antibodies to influenza virus are encoded on the D3-9 segment of the Ig gene. *Cell Host Microbe*. 2018;24:569-578.e4.
22. Wu NC, Andrews SF, Raab JE, et al. Convergent evolution in breadth of two V_H6-1-encoded influenza antibody clonotypes from a single donor. *Cell Host Microbe*. 2020;28:434-444.e4.
23. Throsby M, van den Brink E, Jongeneelen M, et al. Heterosubtypic neutralizing monoclonal antibodies cross-protective against H5N1 and H1N1 recovered from human IgM+ memory B cells. *PLoS One*. 2008;3:e3942.
24. Ekiert DC, Bhabha G, Elsliger MA, et al. Antibody recognition of a highly conserved influenza virus epitope. *Science*. 2009;324:246-251.
25. Okuno Y, Isegawa Y, Sasao F, Ueda S. A common neutralizing epitope conserved between the hemagglutinins of influenza A virus H1 and H2 strains. *J Virol*. 1993;67:2552-2558.
26. Sui J, Hwang WC, Perez S, et al. Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nat Struct Mol Biol*. 2009;16:265-273.
27. Craik DJ, Fairlie DP, Liras S, Price D. The future of peptide-based drugs. *Chem Biol Drug Des*. 2013;81:136-147.
28. Memczak H, Lauster D, Kar P, et al. Anti-hemagglutinin antibody derived lead peptides for inhibitors of influenza virus binding. *PLoS One*. 2016;11:e0159074.
29. Kim DW, Hwang HS, Kim DS, et al. Effect of silk fibroin peptide derived from silkworm Bombyx mori on the anti-inflammatory effect of Tat-SOD in a mice edema model. *BMB Rep*. 2011;44:787-792.
30. Kim JH, Lee SR, Li LH, et al. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One*. 2011;6:e18556.
31. Okochi M, Muto M, Yanai K, et al. Array-based rational design of short peptide probe-derived from an anti-TNT monoclonal antibody. *ACS Comb Sci*. 2017;19:625-632.
32. Perrier A, Eluard M, Petitjean M, Vanet A. In silico design of new inhibitors against hemagglutinin of influenza. *J Phys Chem B*. 2019;123:582-592.
33. Kadam RU, Juraszek J, Brandenburg B, et al. Potent peptidic fusion inhibitors of influenza virus. *Science*. 2017;358:496-502.
34. North B, Lehmann A, Dunbrack RL. A new clustering of antibody CDR loop conformations. *J Mol Biol*. 2011;406:228-256.
35. Grosdidier S, Fernández-Recio J. Protein-protein docking and hot-spot prediction for drug discovery. *Curr Pharm Des*. 2012;18:4607-4618.
36. Brenke R, Hall DR, Chuang G-Y, et al. Application of asymmetric statistical potentials to antibody–protein docking. *Bioinformatics*. 2012;28:2608-2614.
37. Chothia C, Lesk AM. Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol*. 1987;196:901-917.
38. Fleishman SJ, Whitehead TA, Ekiert DC, et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science*. 2011;332:816-821.
39. Giles BM, Ross TM. A computationally optimized broadly reactive antigen (COBRA) based H5N1 VLP vaccine elicits broadly reactive antibodies in mice and ferrets. *Vaccine*. 2011;29:3043-3054.
40. Carter DM, Darby CA, Lefoley BC, et al. Design and characterization of a computationally optimized broadly reactive hemagglutinin vaccine for H1N1 influenza viruses. *J Virol*. 2016;90:4720-4734.
41. Nimrod G, Fischman S, Austin M, et al. Computational design of epitope-specific functional antibodies. *Cell Rep*. 2018;25:2121-2131.e5.
42. Do PC, Nguyen TH, Vo UTM, Le L. iBRAB: in silico based-designed broad-spectrum fab against H1N1 influenza A virus. Preprint. Posted online September 1, 2020. BioRxiv. doi:10.1101/2020.09.01.277335
43. Mizuguchi K, Deane CM, Blundell TL, Overington JP. HOMSTRAD: a database of protein structure alignments for homologous families. *Protein Sci*. 1998;7:2469-2471.
44. Van Walle I, Lasters I, Wyns L. SABmark—a benchmark for sequence alignment that covers the entire known fold space. *Bioinformatics*. 2004;21:1267-1268.
45. Berbalk C, Schwaiger CS, Lackner P. Accuracy analysis of multiple structure alignments. *Protein Sci*. 2009;18:2027-2035.
46. Ye Y, Godzik A. Multiple flexible structure alignment using partial order graphs. *Bioinformatics*. 2005;21:2362-2369.
47. Shatsky M, Nussinov R, Wolfson HJ. A method for simultaneous alignment of multiple protein structures. *Proteins*. 2004;56:143-156.
48. Dror O, Benyamini H, Nussinov R, Wolfson H. MASS: multiple structural alignment by secondary structures. *Bioinformatics*. 2003;19(suppl 1):i95-i104.
49. Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM. MUSTANG: a multiple structural alignment algorithm. *Proteins*. 2006;64:559-574.
50. Menke M, Berger B, Cowen L. Matt: local flexibility aids protein multiple structure alignment. *PLoS Comput Biol*. 2008;4:e10.
51. Andreeva A, Prlić A, Hubbard TJPP, Murzin AG. SISYPHUS—structural alignments for proteins with non-trivial relationships. *Nucleic Acids Res*. 2007;35(suppl 1):D253-D259.
52. Illergård K, Ardell DH, Elofsson A. Structure is three to ten times more conserved than sequence - a study of structural response in protein cores. *Proteins*. 2009;77:499-508.
53. Sriwilaijaroen N, Suzuki Y. Molecular basis of the structure and function of H1 hemagglutinin of influenza virus. *Proc Jpn Acad Ser B Phys Biol Sci*. 2012;88:226-249.
54. Hamilton BS, Whittaker GR, Daniel S. Influenza virus-mediated membrane fusion: determinants of hemagglutinin fusogenic activity and experimental approaches for assessing virus fusion. *Viruses*. 2012;4:1144-1168.

55. Carr CM, Kim PS. A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell*. 1993;73:823-832.
56. Gruenke JA, Armstrong RT, Newcomb WW, Brown JC, White JM. New insights into the spring-loaded conformational change of influenza virus hemagglutinin. *J Virol*. 2002;76:4456-4466.
57. Lao J, Vanet A. A new strategy to reduce influenza escape: detecting therapeutic targets constituted of invariance groups. *Viruses*. 2017;9:38.
58. Larkin MA, Blackshields G, Brown NP, et al. Clustal W and clustal X version 2.0. *Bioinformatics*. 2007;23:2947-2948.
59. Gonnet GH, Cohen MA, Benner SA. Exhaustive matching of the entire protein sequence database. *Science*. 1992;256:1443-1445.
60. Morrow JK, Zhang S. Computational prediction of protein hot spot residues. *Curr Pharm Des*. 2012;18:1255-1265.
61. Ritchie DW, Venkatraman V. Ultra-fast FFT protein docking on graphics processors. *Bioinformatics*. 2010;26:2398-2405.
62. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25:1605-1612.
63. Fiser A, Šali A. Modeller: generation and refinement of homology-based protein structure models. In: Carter CW, Sweet RM, eds. *Macromolecular Crystallography, Part D*. Vol. 374. Academic Press; 2003:461-491.
64. Racine J. gnuplot 4.0: a portable interactive plotting utility. *J Appl Econom*. 2006;21:133-141.
65. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. *J Mol Biol*. 1963;7:95-99.
66. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996;14:33-38.