Deciphering the rule of antigen-antibody amino acid interaction

Min Jiang[†], Changyin Fang[†], Yongping Ma[*]

Short title: **The rule of Ag-Ab interaction**

Keywords: amino acid interaction; ELISA; antigenic drift; immune escape; IEMM; reverse antibody technique;

*__Correspondence__

Yongping MA

Department of Biochemistry and Molecular Biology, Molecular Medicine & Cancer Research Center, Basical Medical Collage, Chongqing Medical University Yuzhong District, Yi XueYuan Road, No 1, Chongqing, 400016, China

[†] **Co-first author**

Abstract

Antigenic drift is the biggest challenge for mutagenic RNA virus vaccine development. The only way to overcome antigenic drift is to update vaccine simultaneously with virus mutation. Consequently, the most fundamental but neglected thing is to determine the immune escape mutation map (IEMM) of 20 amino acids to reveal the rule of the viral immune escape. Thus, we used universal protein-tags as linear epitope model to determine the relationship of the epitope mutation and immune escape. In order to describe and draw amino acid interaction maps, mutations of protein-tags (linear epitopes) were classified into four types: IEM (immune escape mutation), ADEM (antibody-dependent enhancement mutation), EQM (equivalent mutation), and IVM (invalid mutation)) according to antibody binding affinity by ELISA. The amino acid interaction pairs determined by 3D structure through literature search were used to form a more systematic and universal antigen-antibody (Ag-Ab) interaction maps so as to make up for the ELISA data limitation. The results were: (i) One residue interacted with multiple amino acids in Ag-Ab interaction (one-to-multiple rule). (ii) Most of the amino acid replacement were IVM and EQM (harmless mutation predomination rule). (iii) Once non-aromatic amino acids were replaced by aromatic amino acids, the mutation was often inactive (aryl inactive rule). (iv) The substitution of residues with the same physical and chemical properties

easily led to IEM (homogeneous substitution inactive rule). (v) The IEMM of 20 amino acids was established. Therefore, this study had important theoretical significance for future research on antigenic drift and vaccine renewal design.

Introduction

Because RNA polymerase lacks the error-correcting mechanism of 5'-3' exonuclease, the high-speed replication of RNA virus leads to low accuracy of genetic stability, causing the genetic variation of the virus, which is defined as genetic drift(*1*). When this mutation produces amino acid substitution in the neutralizing antigen (Ag), which will lead to typical antigen drift and immune escape during the virus circulating in the population(*2*). Once the immunodominant epitope of the virus surface protein is mutated to form a new subtype, the existing neutralizing antibody (Ab) is no longer effectively recognize the mutated epitope, and the re-exposed host becomes re-infected(*3*).This phenomenon occurs repeatedly in seasonal influenza virus and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic(*4-6*). SARS-CoV-2 is the pathogen of novel coronavirus pneumonia named COVID-19, which has resulted in a significant amount of morbidity and

mortality (*7*, *8*). The pandemic has seriously affected the quality of life, health and even the global economic growth (*9*).

Like most mutagenic RNA viruses, SARS-CoV-2 has high genetic variability and rapid evolution(*10*, *11*). Because the SARS-CoV-2 mutants in the current epidemic are resistant to neutralizing antibodies, including clinical therapeutic antibodies, how to solve antigenic drift is a huge theoretical and practical problem (4, 5). In the past three years, strains B.1.617.2 (Delta) and B.1.1.529 (Omicron) have swept the world (*12*). The mutation leads virus to evade the immune response induced by the host, vaccine or monoclonal antibody (mAb) (*13*-*16*). This has forced the redesign and production of new vaccines to cope with the constant emergence of new variants (*17*). However, the dilemma is that the vaccine development cannot keep pace with viral mutations. Therefore, the vaccine strains must be updated and re-administered without delay. Normally, it takes about 6 months to manufacture matched vaccines for the circulating viruses (*18*).

Antigenic drift is the typical process of RNA virus. According to antigenic drift successfully model, new strains may be generated constantly through mutation, but most of these cannot expand in the host population due to pre-existing immune responses against epitopes of limited variability (*19*). Notably this natural selection is biased, *e.g.*, the

amino acid at position E484 of SARS-CoV-2 is the highest frequency of mutation site and tends to mutate from E (Glu, original strain) to K (Lys, β, γ, and μ variants) to escape from immune, not random occurrence of the other amino acids(*20*). The mutation frequency of E484K is 5.5-fold that of E484Q(*20*).

Immune recognition occurs in the process of matching and anastomosis between specific positions and specific fragments of antibody and antigen molecules. This specific interaction is completely based on the tertiary structure of the folded first-order sequence, and the corresponding first-order sequence changes are the basis to reveal the molecular recognition rule. Nevertheless, due to the complex spatial structure of proteins and the diversity of organisms, it is very difficult to predict exactly how the antigen-determined amino acid will mutate. Anyhow, it is of great significance to explore the rule of amino acid interaction between antigens and antibodies, and then to summarize the interaction (recognition and binding) rule of amino acids for current virus immunity and vaccine preparation. The best way is to detect changes in the ability of the antigen to bind to the mAb by mutating the key amino acid on the epitope, so as to summarize the regular amino acid interaction spectrum.

In order to describe the relationship between linear epitope mutation and immune escape, we cautiously assumed four concepts: (i) Immune

escape mutation (IEM) meant that the residue substitution caused the antigen to lose its affinity (recognition) to pre-existing antibody, or to remain less than the assumed 30% affinity without neutralization. (ii) Antibody-dependent enhancement risk mutation (ADERM) referred to that the residue substitution caused the antigen to remain pre-existing antibody affinity more than the assumed 30% but less than 50%. However, pre-existing antibody could not neutralize the mutated antigen (pathogen). Rather, the virus-Ab complex with low affinity enhanced virus uptake resulting from the attachment of immune complexes to Fcγ receptor and enhanced the infection. (iii) Equivalent mutation (EQM) meant that the residue substitution led to the antigen to remain pre-existing antibody affinity beyond the assumed 50% but less than 80%. Fortunately, the pre-existing antibody still completely neutralizes the antigen. (iv) Invalid mutation (IVM) indicated that the residue substitution did not affect the pre-existing antibody affinity and the antibody completely neutralized the antigen.

Given the complexity of antigen spatial epitopes, we only focused on the relationship between linear epitope mutation and immune escape. The enzyme linked immunosorbent assay (ELISA) was employed to reveal the precise mechanism of the immune escape by measuring the degree of binding affinity between anti-protein-tags mAbs and protein-tag mutants. Specifically, the amino acid sites of universal protein-tags were

mutated one by one and the variation of their OD_{450} values were determined to deplore the rules of antigen-mAb interaction at amino acid level.

In order to make up for the data limitation caused by the detection of universal protein-tags mutants, we also summarized the amino acid interaction pairs determined by 3D structure through literature search to draw a more systematic and universal antigen-antibody (Ag-Ab) interaction map.

Results

Results of protein-tags ELISA

We screened the key residues in HA-tag to mAb TA180128. At first, HA-tag (Y1P2Y3D4V5P6D7Y8A9) was substituted one-by-one from Y1 to A9 with G, E, or H, respectively. Except A9 residue, mAb TA180128 reacted to the other residues from Y1 to Y8 in HA-tag. Then the residues were substituted with other 19 amino acids from Y1 to Y8 and the affinity to mAb TA180128 was measured by ELISA. The results were that Y1H, Y1E, Y3H, V5A, V5I, V5K, V5L, V5M, V5S, V5T, D7G, D7H, Y8E, Y8G, and Y8H mutants were IVM for maintaining more than 80% affinity to mAb TA180128. Instead, P2E, P2H, D4E, D4H, V5F, and P6G mutants were kept less than 8% affinity, and P2G, V5D, V5E, V5W, V5Y, P6A, P6C, P6D, P6E, P6F, P6H, P6I, P6K, P6L, P6M, P6N, P6Q, P6R,

P6S, P6T, P6V, P6W, and P6Y mutants were lost the affinity to mAb TA180128 completely as IEM (Figure 1). Meanwhile, Y1G, Y3E, Y3G, and V5G maintained about 16-18% affinity as IEM, accordingly, these 33 mutants were IEM to mAb TA180128. However, V5C, V5N, and V5R maintained about 35%, 41%, and 48% affinity, respectively, and were ADERM. Anyway, D4G, V5H, V5P, V5P, V5Q, and D7E were EQM for maintaining about 66%-79% affinity to mAb TA180128 (Figure 1).

Since IVM and EQM did not affect the protective effect of current vaccine, the aim was to reveal the IEM of V5 and P6 substitutions replaced by the other 19 amino acids, respectively. As a summary, V5 was substituted by F and G maintained less than 16% affinity and by A, E, D, W, and Y were lost affinity as IEM accounted for 31.6% of the 19 amino acids. Otherwise, V5 was substituted by H, P, and Q were EQM accounted for 15.8% and maintained about 66%-79% affinity to mAb TA180128 (Figure 1). That was, more than 47% V5 substitutions were completely IVM and EQM. However, V5 was substituted by I, K, L, M, S, and T led to IVM, which remained more than 80% affinity to mAb TA180128 accounted for 36.8%. Meanwhile, V5 was substituted by C, N, and R maintained about 35%, 41%, and 47.8% affinity and were ADERM, accounted for 15%.

Similarly, P6 was substituted by I maintained 42% affinity as an ADERM. P6 was substituted by A, C, D, F, H, K, N, Q, R, W, and Y,

were lost the affinity completely as an IEM accounted for 57.8%. Nevertheless, P6 was substituted by E and M were EQM for maintained more than 71% affinity accounted for 10.5%. P6 was substituted by T, S, V, L, and G maintained more than 80% affinity accounted for 26.3% and were IVM to mAb TA180128 (Figure 1). That was, more than 36.8% P6 substitutions were completely IVM and EQM. Therefore, about a half of antigenic drift did not affect the immunity of mAb TA180128.

In order to obtain substitution data covering 20 amino acids, we added the MYC-tag, Flag-tag, and VSV-tag to the experimental catalogue. Specifically, we replaced the Q2, S6, and E8 residues of MYC-tag with G, and replaced its L4 and I5 with other 19 amino acids respectively. While the K3 of Flag-tag was replaced by other 19 amino acids. The N7, R8, and I4 in VSV-tag were replaced by G and the T2 and M6 was mutated to the other 19 amino acids respectively.

For c-MYC-tag mutants, except for the affinity of Q2G, I5S, and E8G > 80%, which belonged to IVM, I5L, SG6, and L4C belonged to EQM. The other mutants had low affinity or could not be combined, and defined as IEM (Figure 2).

Among the mutants of Flag-tag, only K8G shared similar mAb binding affinity with Flag-tag and defined as EQM. The affinity of K3D was 43% and belonged to ADERM. Anyway, the other seventeen K3 mutants were IEM (Figure 3).

Similarly, most of the VSV-tag mutants had low affinity and belonged to IEM. Unexpectedly, the level of affinity of the four mutants replaced by M6C, M6H, M6K, and M6R was higher than that of VSV-tag, and the affinity of N7G was more than 80%, while the affinity of T2R and M6Q was slightly higher, which belonged to EQM. However, M6N and R8G were ADERM (Figure 4).

In order to display the mutant interaction visually, we used Microsoft Excel to draw the heatmap of amino acid residue substitution. From the selected and substituted nine amino acids, I, K, L, and T were tend to lose their affinity with mAb after substitution by other 19 amino acids (Figure 5). Except that the binding affinity substituted by C, H, K, and R could be maintained, the substitution of M by other amino acids led to IEM. Similarly, when D was replaced by E, H, and G, it could still recognize and bind to mAb, the other substitutions of D led to IEM. Conversely, about 84% substitutions of V and Y showed IVM, that was, most substitutions of V and Y retained their original immune affinity. Only V/F, V/G, V/W, V/Y, Y/V, Y/D, Y/I, Y/P, and Y/K replacement models produced IEM (Figure 5). Nevertheless, only 42% substitutions of P showed IVM, and the other 11 substitutions were IEM (Figure 5).

Notably, all amino acids replaced by A, D, F, P, W, and Y led to inactive, and 98% amino acids replaced by V and I might be inactive (Figure 5). However, 66% amino acids replaced by G might be inactive

(Figure 6).

Obviously, the position of the original amino acid also had a great influence on the outcome of the substitution of another identical amino acid. For instance, the substitution of MYC-E8G resulted in EQM, and the substitution of MYC-E7G led to IEM (Figure 6). Of course, the result indicated that MYC-E7 was the key residue to MYC epitope.

Literature data analysis of antigen-antibody interactions

Seventy articles related to Ag-Ab interaction were screened through the NCBI PubMed database and Web of Science. The interactions of amino acid residues at the Ag-Ab binding interface were extracted from the 3D structures reported in the literature, and the amino acid binding matrix was drawn and colored according to the number of relevant literatures (Figure 7A). The X-axis is the amino acids of antibody and the Y-axis is antigenic residues. According to the matrix, the binding rules of amino acids in antigen and antibody tended to be consistent. Among the 20 types of amino acids, Y, S, D, and R had affinity with other amino acids and could recognize almost all amino acids in the literature. Instead, C and M were almost silent and only interact with relatively few amino acids, which coincided with some of the findings in our ELISA experiments. Explicitly, residue M did not find in antibody interface and only 4 in antigen interface (Figures 7B and C). Residue C only found one in antibody interface and only 3 in Ag interface (Figures 7B and C).

Otherwise, P was rarely found in Ab interfaces and V rarely found in Ag interfaces (Figures 7B and C).

Next, the histogram of the interaction capacity of the 20 amino acids was plotted according to the literatures. The top10 amino acids with the highest detection rate in antibodies were Y, R, D, N, S, K,T, F, L, and E, respectively (Figure 7B). Nevertheless, the top10 detection amino acids in antigens were R, Q, D, N, E, Y, L, A, H, K, S, P, S, and I (Figure 7C). It must explain that the H, K, S, P, S, and I shared the 10th rank for their same detection rate (Figure 7C).

Amino acid interaction network analysis

The data were imported into Cytoscape to draw interaction network. According to the physical and chemical properties, 20 amino acids were divided into 8 categories, which were aliphatic amino acids, aromatic amino acids, acidic amino acids, basic amino acids, amide amino acids, sulfur-containing amino acids, hydroxyl amino acids, and sub-amino acids. Then, the interaction diagram was constructed (Figure 8A). Except for a few aliphatic amino acids and all sulfur-containing amino acids, the side chains of hydroxy-amino acids S and T could interact with 14 amino acids in the literatures (Figure 8B). The aliphatic amino acids included A,V, I, and L, and only I could interact with C (Figure 8C). The alkaline amino acids K, R, and H could interact with 17 kinds of amino acids in antibodies (Figure 9A). Notably, antigenic P interacted with 11 amino

acids in antibodies, and P in Abs only interacted with antigenic N, Q, and R (Figures 8A and C; Figure 9C). The antigenic C (interacted with D, T, and V) and M (interacted with L,R, W, and Y) contained sulphur in their side chains, causing them not to cross-link (Figure 9D). Otherwise, the residue of antibodies C only recognized antigenic I and N (Figures 8A and C; Figure 9C). Acidic amino acids could interact with 16 amino acids (Figure 9E). Similarly, antigenic aromatic amino acids seemed to be interacted with more than 10 amino acids (Figure 9F).

In order to quickly view the amino acid interaction between Ag-Ab, Figure 8A was plotted individually in alphabetical order of amino acid abbreviations. Amino acids without interaction were listed in boxes below the main figure (Figures S1A-Y).

Verification of experimental data and literature data

In order to reveal the effect of mutation on the peptide structure and verify the experimental results, we used the computer ab initio folding algorithm to predict the 3D structure of peptide tags and their mutants, and combined with the interaction rule to fit the structure of some mutants. We predicted the 3D structure of HA-tag, MYC-tag and Flag-tag and their mutants HA-Y1G, HA-Y3G, HA-Y8G, MYC-E8G, MYC-E7G, Flag-K8G, and Flag-K3G respectively by online tool PEP-FOLD3. After obtaining the PDB file, we compared the mutant with the original structure. The fit degree of HA-Y8G, and MYC-E7G to the original tag

structure was lower than that of other mutants (Figure 10). As proof, HA-tag was at an angle of 10°, however, HA-Y8G at 120°. MYC-tag at -24° and MYC-E8G at 11°, and MYC-E7G at 30°. Surprisingly, Flag-K3G mutation disrupted the original α-helix and curled into a semi-o-ring in the opposite direction of the original Flag-K3 peptide, which affected the antibody binding capacity (Figures 3 and 10). Consequently, the low fit mutation resulted in IEM due to loss of interaction. The same IEM mutations were happened in HA-D4E, HA-P6F, HA-P6W, HA-P6Y, VSV-M6F, VSV-M6Y, and VSV-M6W (Figures 11 and 12).

Protein-tags mutation experiment found that the substitution of amino acids with similar physical and chemical properties residue tended to IVM, *e.g.*, aliphatic amino acids V, I, and L substituted each other (HA-V5I, HA-V5L; MYC-I5L) (Figures 1, 2 and 4). Similarly, aromatic amino acids Y, W, and F mutual substitution were also IVM, *e.g.*, HA-Y3F and HA-Y3W (Figures 1B). The results indicated that the amino acids with the same physical and chemical properties of the side chain group shared high similarity in 3D structures (Figure 11). In addition, The non-aromatic amino acids were replaced by aromatic amino acids (F, Y, and W) often lost their affinity to antibodies (Figures 1-4). It might be related to the benzene ring on the side chain of aromatic amino acids. We consequently predicted the structure of some non-aromatic amino acid mutants replaced by aromatic amino acids and compared their structural

changes (Figure 12). To visually describe the structural changes of the mutated tags, the RMSD values were obtained in UCSF ChimeraX for the fitted structures of the above mentioned mutants and the original protein-tags were summarized (Table S5). RMSD was a measure of the coordinate deviation of a particular part of the atom with respect to a reference structure. If the RMSD of two structures was 0 Å, then they were considered to be of the same structure and can fully coincide. When RMSD was less than 3 Å, the two structures were considered similar. Conversely, if the RMSD was greater than 3 Å, indicating that the structure of the mutant was inconsistent with that of the original type. Thus, our test suggested that the protein-tag mutants with RMSD > 3.0 Å were trended to IEM (Table S5).

Conclusion

This study had important theoretical significance for the future research of antigenic drift and vaccine renewal design. And the following conclusions were summarized:

(I) Ag-Ab interactions were classified into four types: IEM, ADEM, EQM, and IVM according to antibody binding affinity by ELISA.

(II) For most antigenic amino acids, the pattern of recognition and binding was one-to-many. The amino acids on antibodies were also in one-to-many mode (one-to-many recognition rule). The conclusion obtained from the experiment was supported and verified by the literature

data.

(III) For specific antigenic amino acid, most of the replacement were IVM and EQM (harmless mutation predomination rule).

(IV) Once a antigenic non-aromatic amino acid was replaced by aromatic amino acids, the mutation was often inactive and resulted in IEM (aryl mutation harmful rule).

(V) Antigenic amino acids were replaced by other amino acids with similar physical and chemical properties. Most of the replacement mutations were easily led to IEM (homogeneous mutation harmful rule).

Discussion

The goal of this study is to establish the rule of the amino acid interaction during Ag-Ab recognition and use this rule to guide scientists to rescue mAb with good efficacy and safety in clinical application, saving time and money for secondary research and development in the future. Antibody rescue means when a key antigenic amino acid is mutated, leading to the ineffectiveness of mAb in clinical application, the original amino acid on the mAb is replaced with a new amino acid according to the Ag-Ab recognition rule, so as to recognize and neutralize the mutated antigen again. We called it reverse antibody technique.

For instance, the E484K mutation of SARS-CoV-2 RBD disabled mAb P17 neutralization (*21*, *22*). 3D data demonstrated that the

negatively charged E484 interacted with positively charged H35, H99, and R96 residues, which formed a strong electrostatic interaction (*22*). Naturally, E484K mutation was repulsive to the positively charged binding site of the mAb P17. Similarly, E484A mutation was reduced polar interactions with mAb 2-15, which failed to interacted with $Y33_{HC}$ and $N52_{HC}$ (*23*, *24*). Guided by our reverse antibody theory, mAb P17 might be theoretically rescued by mutating part or all of the H35, H99, and R96 into negatively charged residues such as E and/or D. With current technology, single-base directed mutation of an existing antibody is much easier and has a high probability of success than a new antibody screening for a variant antigen.

In RNA viruses, the extremely high mutation rate is advantageous because they may allow for increased adaptability (*25*). Research suggests that replication speed is more important than accuracy in RNA viruses. Because the spread of the virus within the host is related to the rate of replication of the virus, not to genetic diversity (*26*). The trade-off between speed and fidelity determines the mutation rate and virulence of RNA viruses. Low fidelity leads to antigen drift and immune escape while the random mutation occurs in pathogen epitopes (*27*).

In fact, viral immune escape caused by antigenic drift has always been a great challenge for vaccination and prevention (*28*). It is the best way to fully reveal the molecular mechanism of antigen drift and immune

escape by studying the universal rule of amino acid interaction of Ag-Ab. In this study we summarized 4 rules to illustrate the antigenic amino acid mutation and the rules of Ag-Ab recognition, which helped us to design more efficient vaccine or to develop therapeutic mAb (*29*, *30*).

The IEM rule obtained from protein-tags was confirmed in the real-world data. For example, we found that HA-tag P6L mutation resulted in IEM (Figure 1). In SARS-CoV, P462L mutation was led to escape from mAb CR3014 (*31*). Similarly, antigenic amino acids replaced with P was IEM in protein-tags mutation (Figures 1-4), the P384 mutation on SARS-CoV-2 also led to IEM for SARS-CoV (A384) mAb CR3022 (*32*). We found that antigens were often not recognized by antibody when residues K, L, or T were replaced by other residues. It was confirmed that the spike L452R mutation conferred SARS-CoV-2 escape from immune (*33*). In this study, the mutations of V to A were led to IEM (Figure 1). Many studies have indicated that the A180V mutation in the influenza virus promotes the virus to escape from antibody-based immune in vivo and in vitro (*34-36*). Studies have confirmed that A475V and V483A mutations in SARS-CoV-2 variants are resistant to some neutralizing antibodies (*37*). Hence, the mutual mutations of antigenic A and V are inactive, leading to IEM. The amino acid mutation type of antigen and the corresponding IEM pattern are termed IEMM (Figure 14).

In theory, antigenic residue determines antibody structure. So far, it

cannot be precisely predicted which kinds of amino acid substitution lead to immune escape to a given amino acid in a linear epitope. Only IEMM can answer the question of which kinds of antigenic substitution at critical binding sites will significantly disrupt the Ag-Ab binding affinity and subsequently cause immune escape. The purpose of IEMM establishment is to deal with antigenic drift and immune escape of human viruses (*30*). Consequently, it was a significant work to determine IEMM of 20 amino acids to a given residue.

Theoretically, the substitution of epitope key residues seems to be a random event, but it has its own secrets. A great deal of mutations does not exist in nature for immune pressure (*38*). That is, all strains found to have mutations in neutralizing antigens are more or less immune escape strains. In particular, mutations at spike protein L452 are recurrently observed SARS-CoV-2 (*3, 39, 40*). Tan et al characterized all 19 possible mutations at L452 and revealed that five mutants expressing the amino acids Q, K, H, M, and R gained greater infectivity and immune escape, whereas other mutants failed to maintain steady-state expression levels and/or pseudovirus infectivity (*39*). Nevertheless, the a single nucleotide mutations of L452 are limited to Q, M, and R in nature (*3, 33, 40, 41*).

As mentioned above, the amino acid interaction between Ag-Ab was not a one-to-one cross-linking of individual amino acids. We thought it was the mutual attraction and matching of the overall net charge value

between the interaction interface centered around the key amino acid on the epitope of antigen and the antibody subepitope, so as to achieve the equilibrium state of the interaction interface microenvironment with the lowest energy and the most stable structure. The 3D structure of Ab showed that the V_H and V_L regions each contain three hypervariable or complementarity-determining regions (CDRs). Six CDRs composed of a subepitope (*42*, *43*).

Protein-protein interaction (PPI) interface can be divided into planar and non-planar interfaces, and the heterocomplexes have interfaces that are more planar than the homodimers (*44*). However, only a small subset of all interface residues is actually essential for recognition or binding, which often refer to as "hotspots" (*45*). The protein interactions will be blocked if these hotspots (residues) were mutated. The four binding forces (electrostatic forces, Van der Waals forces, hydrogen bonds, and hydrophobic interaction forces) provide stability to antigens and antibodies (*46*).

Our research suggested that a half of amino acid mutations were IVM and EQM in a specific site. Meanwhile, the specific amino acid caused by a single nucleotide mutation are limited in specific types in nature (*39*). Therefore, following the IEMM, vaccinologists just focused on the IEM hotspots to design different epitope mutants as epitope library and to determine their antigenicity and safety. In the event of IEM strain

pandemic, the matching epitope library sequences will be immediately constructed into pre-existing vector to substitute of the original (parental) sequence for purified subunit vaccine and/or recombinant live vaccine development (Figure 13). Manufacturers could develop reverse mAb according to IEMM to improve the current mAb. In this way, the strategy for mutagenic RNA viruses prevention will be changed from “guessing and chasing” to “predicting and preparing in advance” in the future.

Although supported by experimental and literature data, our study still has some shortcomings. First, the Ag-Ab amino acids interaction pattern was not covered all 20 kinds amino acids. Our experimental data came from random mutations of linear epitopes, and some mutations were still not involved yet, which need to be further explored. Even so, the protein-tags and their mutants were detected by ELISA, a golden standard method, and consequently the credibility and reliability of the results were reliable. Furthermore, the joint 3D database analysis made up for this deficiency to some extent. Second, the protein-tags linear epitopes lack the complexity of spatial epitopes and the Ag-Ab interaction data derived from them may not fully reflect the true spatial conformation of complete antigen molecules. Therefore, the reverse antibody test was the main task to verify this theory in the future.

Acknowledgements

No

Contributions

Y.M. conceived of the paper and written the manuscript, proposed the four concepts: IEM, IVM, EQM and ADERM, and proposed the hypothesis of reverse monoclonal antibody technology. M.J. and C.F. contributed, as co-first author, to perform the experiments, and analyze data. M.J. prepared all of the figures and written the draft.

Competing interests

No competing interests.

Supplementary materials

Supplementary materials and methods

Supplementary table S1. Sequences of HA-tag and its mutants.

Supplementary table S2. Sequences of MYC-tag and its mutants.

Supplementary table S3. Sequences of Flag-tag and its mutants.

Supplementary table S4. Sequences of VSV-tag and its mutants.

Supplementary table S5. RMSD of the fitted structures of mutant tags and original protein-tags

References

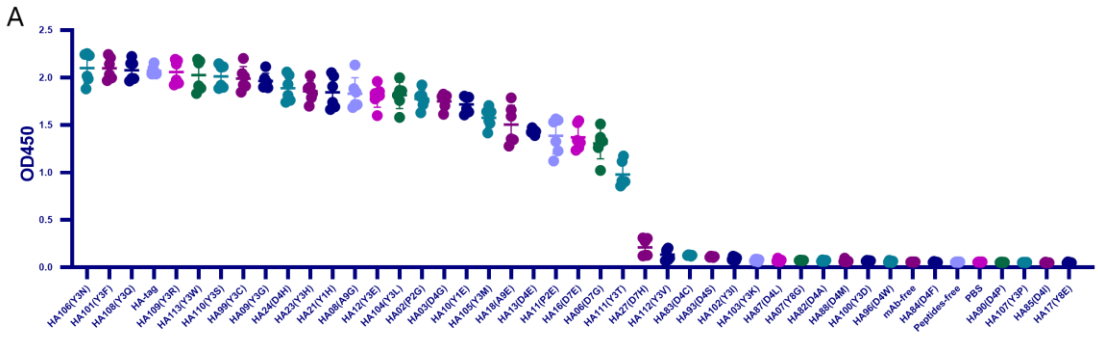
1. J. W. Yewdell, Antigenic drift: Understanding COVID-19. *Immunity* **54**, 2681-2687 (2021).
2. M. Auladell *et al.*, Recalling the Future: Immunological Memory Toward Unpredictable Influenza Viruses. *Frontiers in immunology* **10**, 1400 (2019).
3. Y. Cao *et al.*, Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature* **602**, 657-663 (2022).
4. L. Deng *et al.*, Risk of SARS-CoV-2 reinfection: a systematic review and meta-analysis. *Scientific reports* **12**, 20763 (2022).
5. L. R. Goes *et al.*, New infections by SARS-CoV-2 variants of concern after natural infections and post-vaccination in Rio de Janeiro, Brazil. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **94**, 104998 (2021).
6. V. N. Petrova, C. A. Russell, The evolution of seasonal influenza viruses. *Nature reviews. Microbiology* **16**, 47-60 (2018).
7. D. Wang *et al.*, Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *Jama* **323**, 1061-1069 (2020).
8. A. W. K. Yeung *et al.*, Official Websites Providing Information on COVID-19 Vaccination: Readability and Content Analysis. *JMIR public health and surveillance* **8**, e34003 (2022).
9. J. J. Malin, I. Suárez, V. Priesner, G. Fätkenheuer, J. Rybniker, Remdesivir against COVID-19 and Other Viral Diseases. *Clinical microbiology reviews* **34**, (2020).
10. B. Hu, H. Guo, P. Zhou, Z. L. Shi, Characteristics of SARS-CoV-2 and COVID-19. *Nature reviews. Microbiology* **19**, 141-154 (2021).
11. D. Singh, S. V. Yi, On the origin and evolution of SARS-CoV-2. **53**, 537-547 (2021).
12. R. T. Eguia *et al.*, A human coronavirus evolves antigenically to escape antibody immunity. *PLoS pathogens* **17**, e1009453 (2021).
13. M. M. DeGrace *et al.*, Defining the risk of SARS-CoV-2 variants on immune protection. *Nature* **605**, 640-652 (2022).
14. L. Swadling *et al.*, Pre-existing polymerase-specific T cells expand in abortive seronegative SARS-CoV-2. *Nature* **601**, 110-117 (2022).
15. E. Callaway, H. Ledford, How bad is Omicron? What scientists know so far. *Nature* **600**, 197-199 (2021).
16. D. Focosi, F. Maggi, Neutralising antibody escape of SARS-CoV-2 spike protein: Risk assessment for antibody-based Covid-19 therapeutics and vaccines. *Reviews in medical virology* **31**, e2231 (2021).
17. E. Buntinx *et al.*, Immunogenicity of an adjuvanted SARS-CoV-2 trimeric S-protein subunit vaccine (SCB-2019) in SARS-CoV-2-naïve and exposed individuals in a phase 2/3, double-blind, randomized study. *Vaccine* **41**, 1875-1884 (2023).
18. F. Krammer, A. García-Sastre, P. Palese, Is It Possible to Develop a "Universal" Influenza Virus Vaccine? Potential Target Antigens and Critical Aspects for a Universal Influenza Vaccine. *Cold Spring Harbor perspectives in biology* **10**, (2018).
19. M. Recker, O. G. Pybus, S. Nee, S. Gupta, The generation of influenza outbreaks by a network of host immune responses against a limited set of antigenic types. *Proceedings*

of the National Academy of Sciences of the United States of America **104**, 7711-7716 (2007).

20. A. J. Greaney *et al.*, Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell host & microbe* **29**, 463-476.e466 (2021).
21. P. A. G. Ferrareze *et al.*, E484K as an innovative phylogenetic event for viral evolution: Genomic analysis of the E484K spike mutation in SARS-CoV-2 lineages from Brazil. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **93**, 104941 (2021).
22. L. Wu *et al.*, Exploring the immune evasion of SARS-CoV-2 variant harboring E484K by molecular dynamics simulations. **23**, (2022).
23. G. Cerutti *et al.*, Cryo-EM structure of the SARS-CoV-2 Omicron spike. *Cell reports* **38**, 110428 (2022).
24. M. Rapp *et al.*, Modular basis for potent SARS-CoV-2 neutralization by a prevalent VH1-2-derived antibody class. *Cell reports* **35**, 108950 (2021).
25. E. C. Smith, N. R. Sexton, M. R. Denison, Thinking Outside the Triangle: Replication Fidelity of the Largest RNA Viruses. *Annual review of virology* **1**, 111-132 (2014).
26. W. J. Fitzsimmons *et al.*, A speed-fidelity trade-off determines the mutation rate and virulence of an RNA virus. **16**, e2006459 (2018).
27. D. Forni *et al.*, Antigenic variation of SARS-CoV-2 in response to immune pressure. *Molecular ecology* **30**, 3548-3559 (2020).
28. C. Nypaver, C. Dehlinger, C. Carter, Influenza and Influenza Vaccine: A Review. *Journal of midwifery & women's health* **66**, 45-53 (2021).
29. D. R. Burton, L. M. Walker, Rational Vaccine Design in the Time of COVID-19. *Cell host & microbe* **27**, 695-698 (2020).
30. N. L. Miller, R. Raman, T. Clark, R. Sasisekharan, Complexity of Viral Epitope Surfaces as Evasive Targets for Vaccines and Therapeutic Antibodies. *Frontiers in immunology* **13**, 904609 (2022).
31. J. ter Meulen *et al.*, Human monoclonal antibody combination against SARS coronavirus: synergy and coverage of escape mutants. *PLoS medicine* **3**, e237 (2006).
32. N. C. Wu *et al.*, A natural mutation between SARS-CoV-2 and SARS-CoV determines neutralization by a cross-reactive antibody. *PLoS pathogens* **16**, e1009089 (2020).
33. C. Motozono *et al.*, SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. *Cell host & microbe* **29**, 1124-1136.e1111 (2021).
34. S. E. Hensley *et al.*, Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. *Science (New York, N.Y.)* **326**, 734-736 (2009).
35. J. E. Sealy *et al.*, Association of Increased Receptor-Binding Avidity of Influenza A(H9N2) Viruses with Escape from Antibody-Based Immunity and Enhanced Zoonotic Potential. *Emerging infectious diseases* **25**, 63-72 (2018).
36. R. Zhu *et al.*, HA gene amino acid mutations contribute to antigenic variation and immune escape of H9N2 influenza virus. *Veterinary research* **53**, 43 (2022).
37. Q. Li *et al.*, The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell* **182**, 1284-1294.e1289 (2020).
38. M. O. Altman, D. Angeletti, J. W. Yewdell, Antibody Immunodominance: The Key to

Understanding Influenza Virus Antigenic Drift. *Viral immunology* **31**, 142-149 (2018).
39. T. S. Tan *et al.*, Dissecting Naturally Arising Amino Acid Substitutions at Position L452 of SARS-CoV-2 Spike. *J Virol* **96**, e0116222 (2022).
40. V. Tchesnokova *et al.*, Acquisition of the L452R Mutation in the ACE2-Binding Interface of Spike Protein Triggers Recent Massive Expansion of SARS-CoV-2 Variants. *J Clin Microbiol* **59**, e0092121 (2021).
41. S. Lusvarghi *et al.*, Key Substitutions in the Spike Protein of SARS-CoV-2 Variants Can Predict Resistance to Monoclonal Antibodies, but Other Substitutions Can Modify the Effects. *J Virol* **96**, e0111021 (2022).
42. S. Sheriff *et al.*, Three-dimensional structure of an antibody-antigen complex. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 8075-8079 (1987).
43. A. G. Amit, R. A. Mariuzza, S. E. Phillips, R. J. Poljak, Three-dimensional structure of an antigen-antibody complex at 2.8 A resolution. *Science (New York, N.Y.)* **233**, 747-753 (1986).
44. S. Jones, J. M. Thornton, Principles of protein-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 13-20 (1996).
45. Y. Ofran, B. Rost, Protein-protein interaction hotspots carved into sequences. *PLoS computational biology* **3**, e119 (2007).
46. B. C. Braden, R. J. Poljak, Structural features of the reactions between antibodies and protein antigens. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **9**, 9-16 (1995).

Figure legends



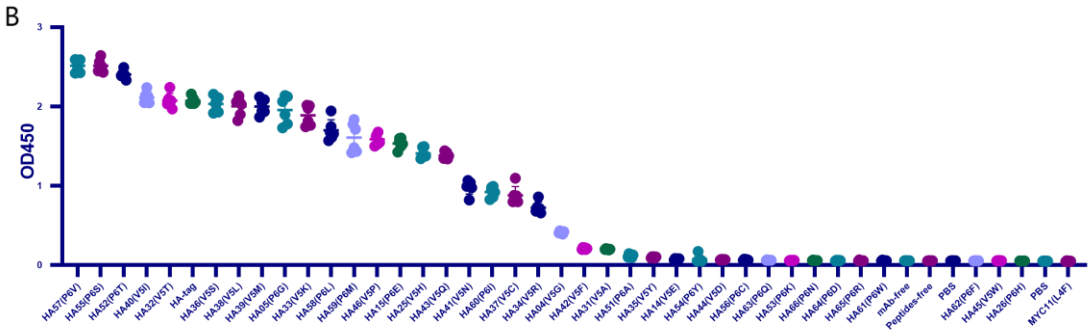


Figure 1. Scanning the binding affinity of HA-tag mutants to mAb TA180128.

(A.B.) HA-tag mutants P2X, Y3X, D4X, V5X, P6X, D7X, Y8X, A9X.

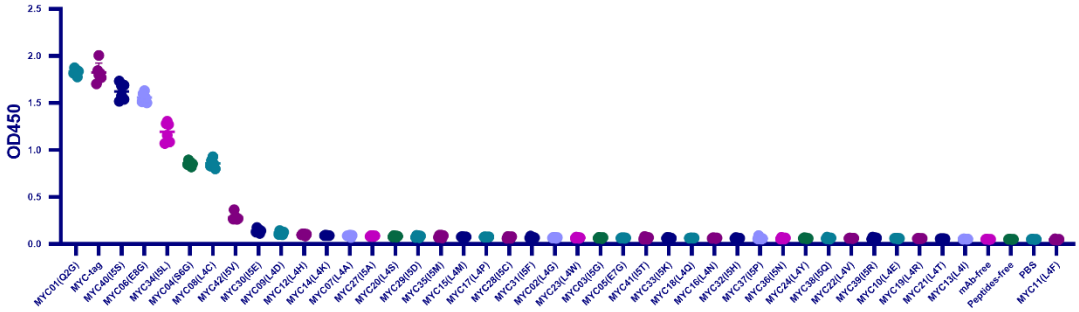


Figure 2. Scanning the binding affinity of MYC-tag mutants to mAb6003-2-lg.

MYC-tag mutants Q2X, L4X, I5X, S6X, E8X

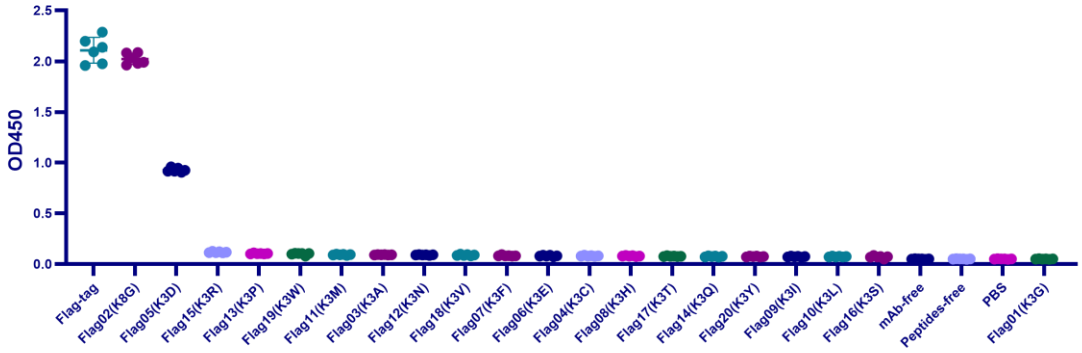


Figure 3. Scanning the binding affinity of Flag-tag mutants K3X to mAb 66008-3-lg.

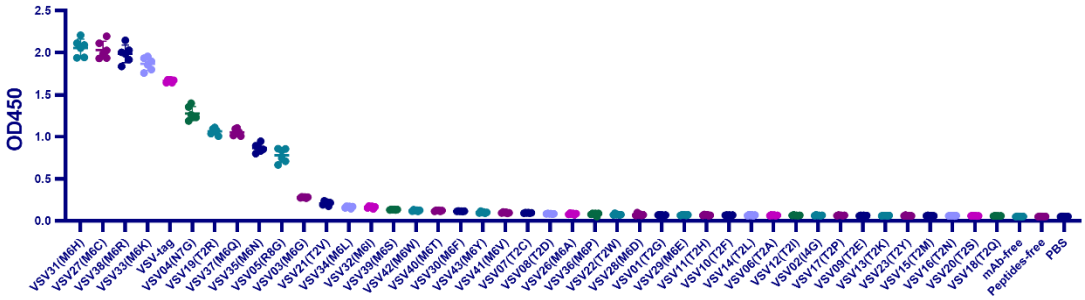


Figure 4. Scanning the binding affinity of VSV-tag mutants to mAb A02180.

VSV-tag mutants T2X, M6X, N7X, R8X, I4X.

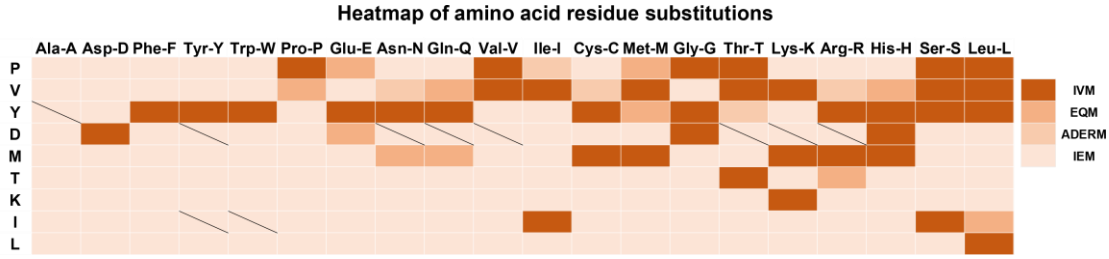


Figure 5. Heat map of amino acid residue substitution verified by ELISA.

IVM: Invalid mutation, EQM: Equivalent mutation, ADERM: antibody-dependent enhancement risk mutation; IEM: Immune escape mutation; P: HA-P6; V: HA-V5; Y: HA-Y3; D: HA-D4; M: VSV-M6; T: VSV-T2; K: Flag-K3; I: MYC-I5; L: MYC-L4； Slash line: No experimental data available.

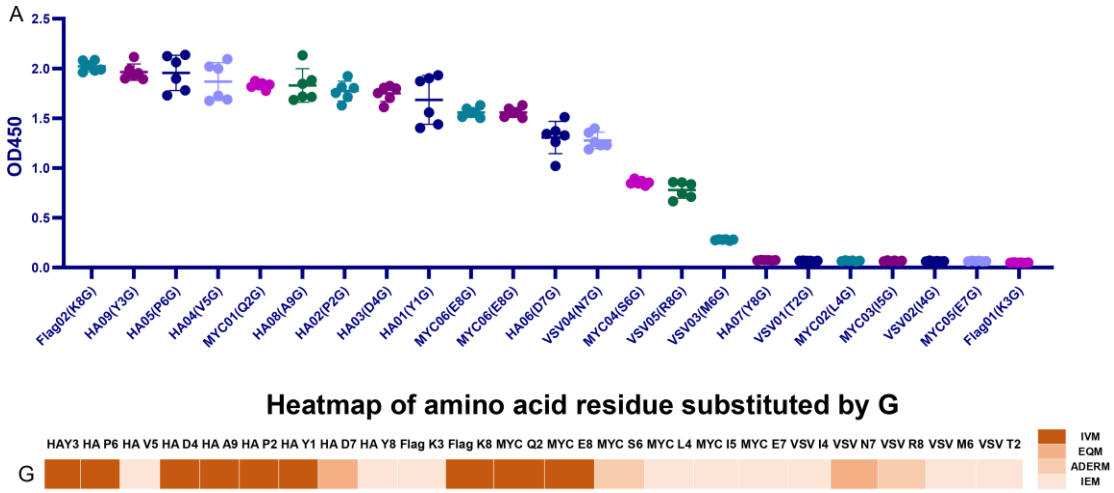


Figure 6. The binding affinity map of mutant tags replaced with G

(A)Map of ELISA results of mutant strains and monoclonal antibodies. (B) heat map of binding affinity of mutant tags and monoclonal antibodies.

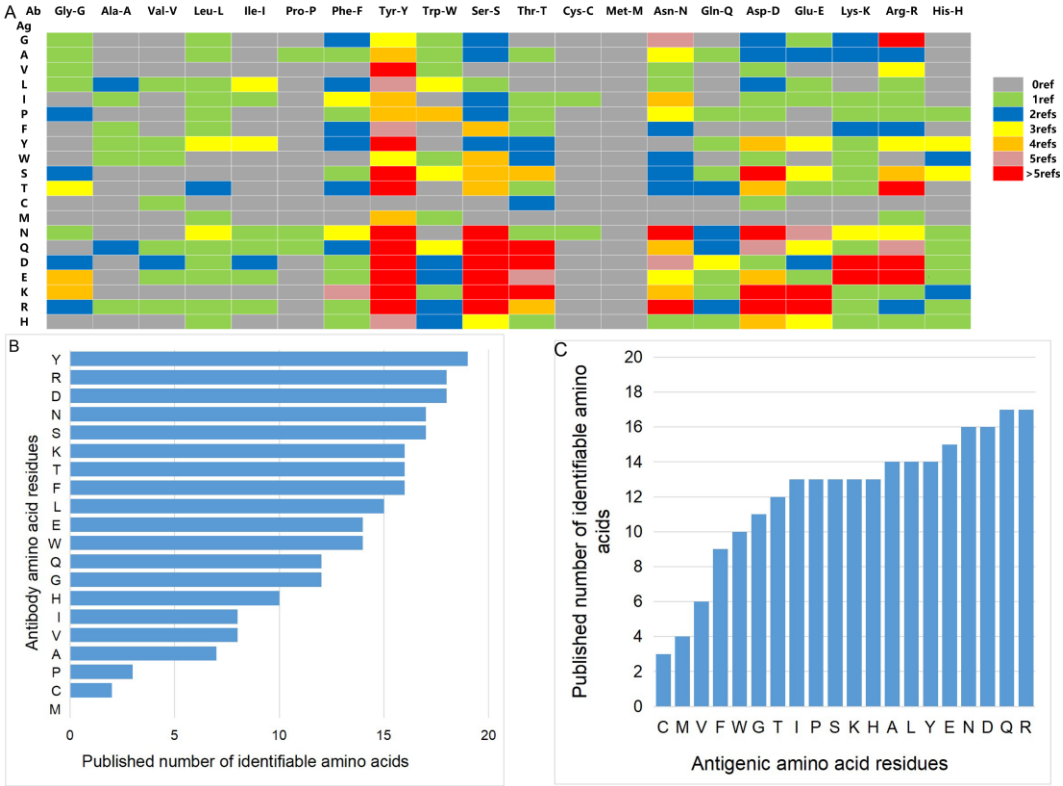


Figure 7. Summary of literature amino acid recognition rules.

(A) Amino acid recognition rule matrix collected by literature. (B) Histogram of identifiable amount of amino acid residues in antibodies. (C) Histogram of identifiable quantity of amino acid residues in antigen.

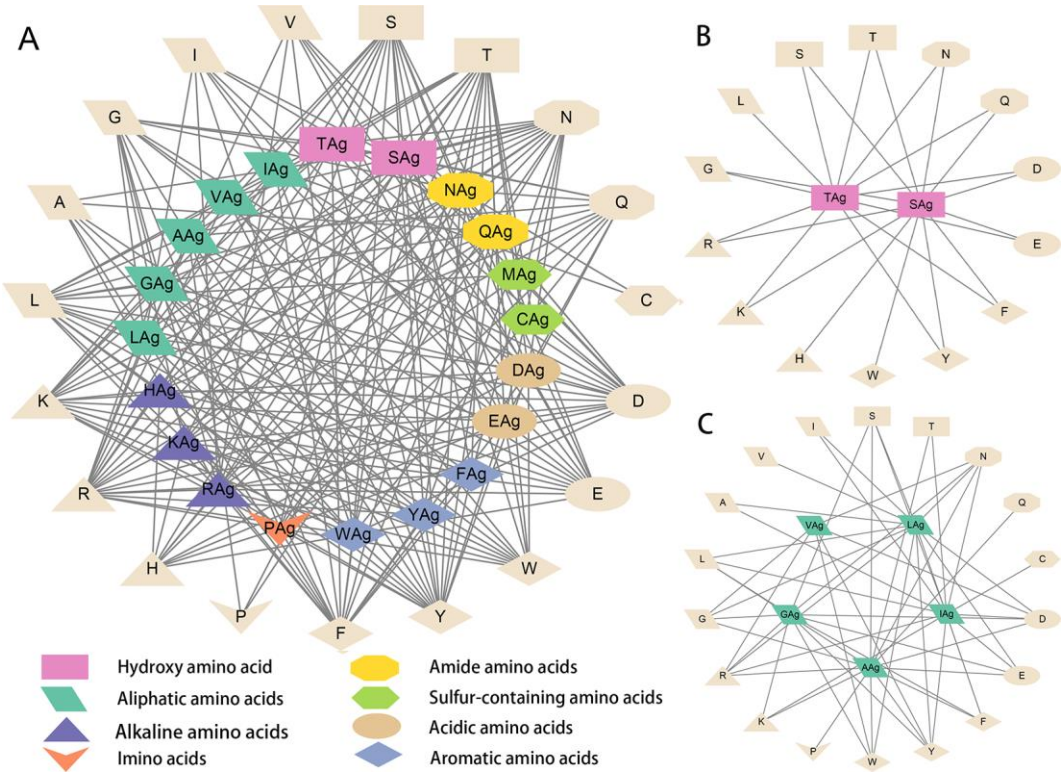


Figure 8. Literature antigen-antibody amino acid interaction diagram.

(A) Summary diagram of Ag-Ab amino acid interaction based on physical and chemical properties. (B) Diagram of interactions between hydroxy amino acids and 20 amino acids. (C) Diagram of interactions between aliphatic amino acids and 20 amino acids. The antigenic amino acid abbreviation was labeled with the subscript Ag to distinguish antibody amino acids.

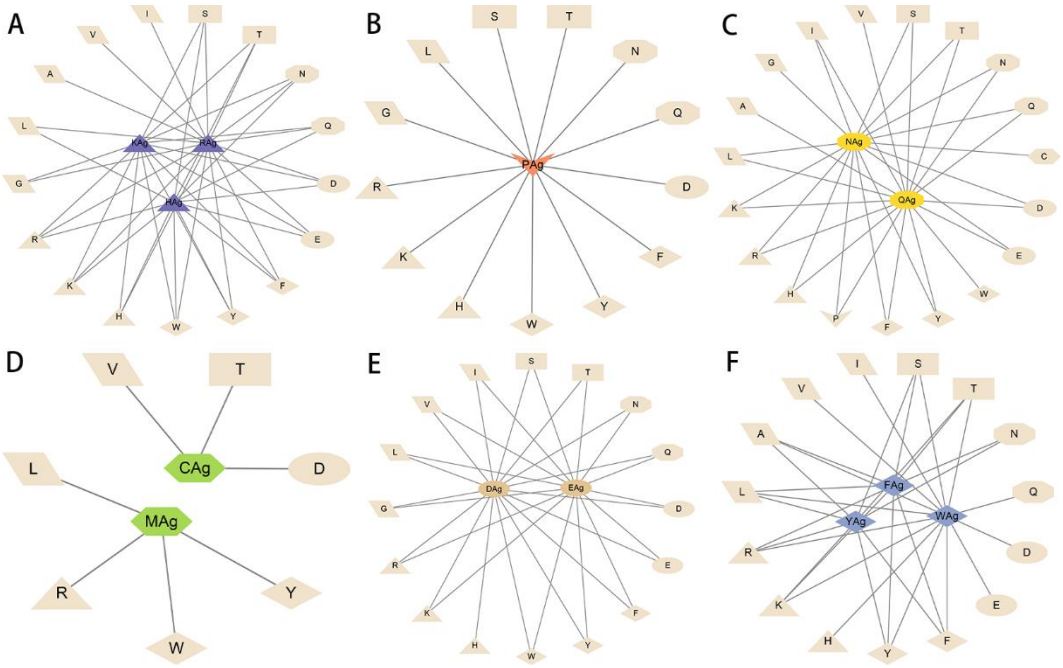


Figure 9. Literature antigen-antibody amino acid interaction diagram.

(A) Diagram of interactions between alkaline amino acids and 20 amino acids. (B) Imino acids. (C) Amide amino acids. (D) Sulfur-containing amino acids. (E) Acidic amino acids. (F) Aromatic amino acids. The antigenic amino acid abbreviation was labeled with the subscript Ag to distinguish antibody amino acids.

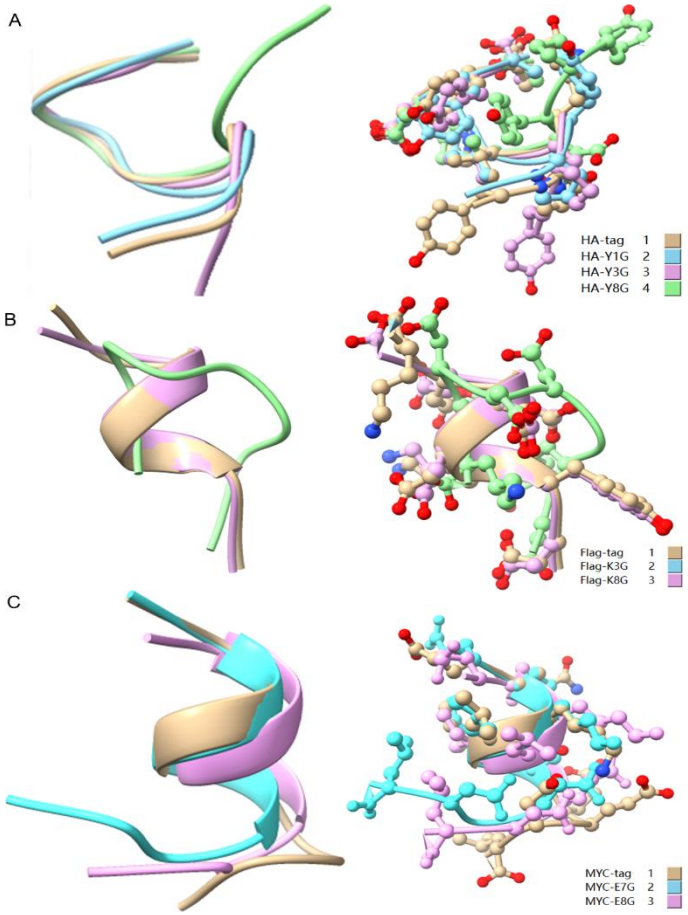


Figure10. Comparison of the 3D structure of mutant tags replaced with G

(A) HA-tag with HA-Y1G, HA-Y3G and HA-Y8G (B) Flag-tag with Flag-K8G and Flag-K3G (C) MYC-tag with MYC-E8G and MYC-E7G

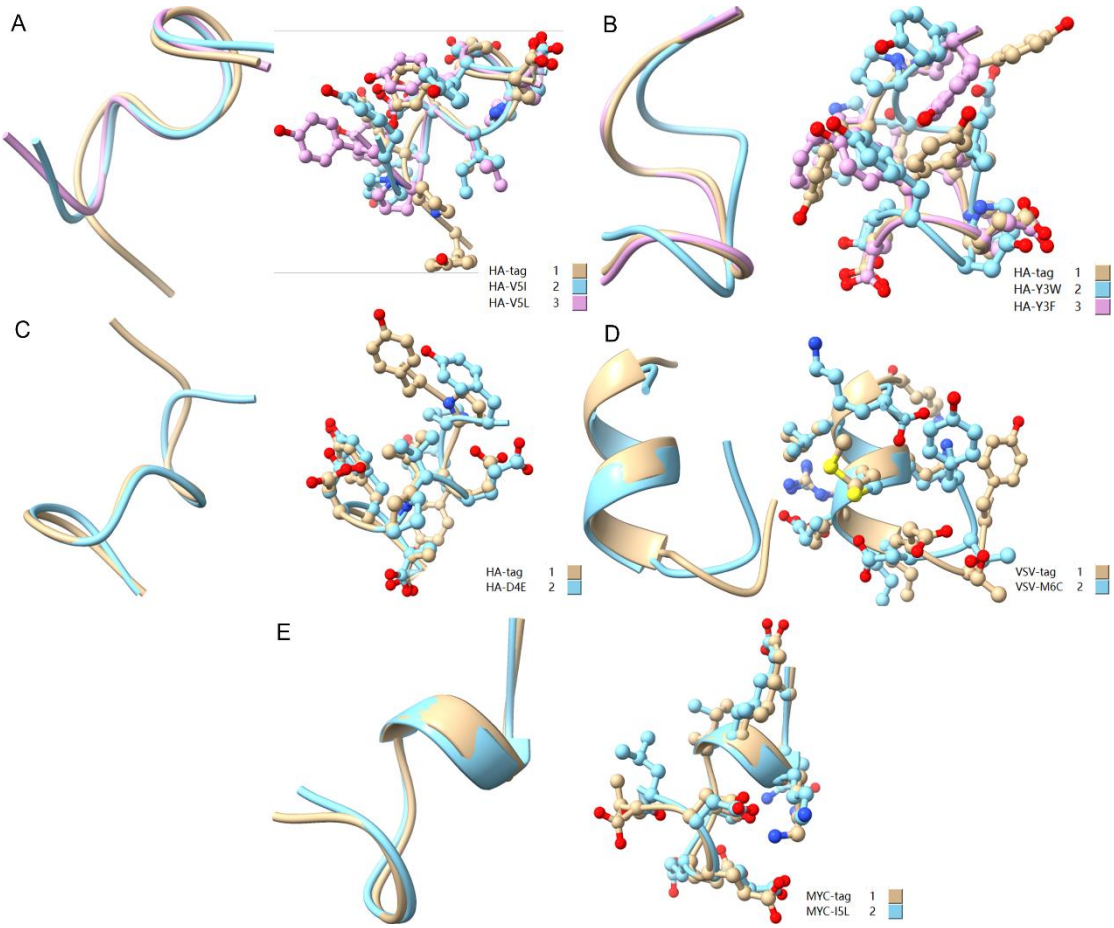


Figure 11. Comparison of amino acid mutant with similar physicochemical property residues.

(A) HA - tag, HA - V5I, and HA - V5L; (B) HA-tag, HA-Y3W, and HA-Y3F; (C) HA-tag and HA-D4E; (D) VSV-tag and VSV-M6C; (E) MYC-tag and MYC-I5L;

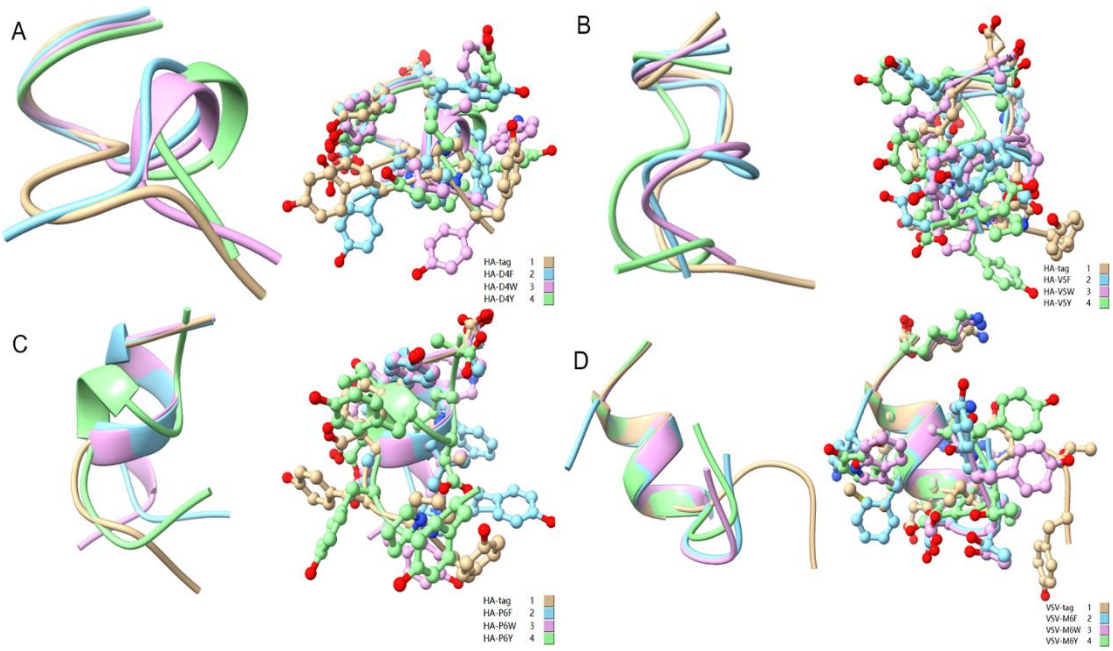


Figure 12. Comparison of aromatic amino acid substitution mutants

(A) HA-tag, HA-D4F, HA-D4Y, and HA-D4W; (B) HA-tag, HA-V5F, HA-V5Y, and HA-V5W; (C) HA-tag, HA-P6F, HA-P6Y, and HA-P6W((D) VSV-tag, VSV-M6F, VSV-M6Y, and VSV-M6W;

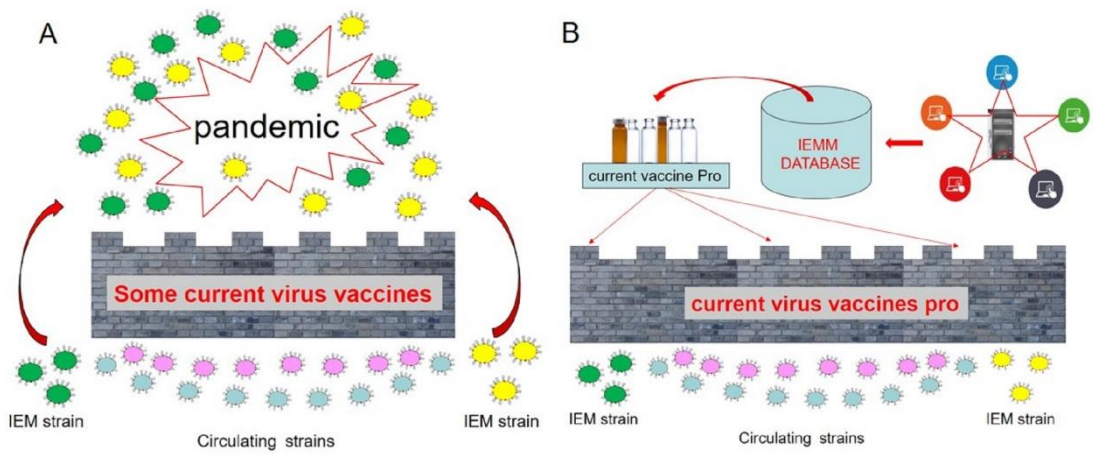


Figure 13. IEMM vision and future vaccine improvement.

(A) Antigenic drift caused circulating virus strains to mismatch with current vaccine strains, leading to IEM and pandemic. (B) Based on IEMM database, current vaccine strains were improved to match with novel circulating virus strains (antigenic drift strains).

SUPPLEMENTARY MATERIALS

Supplementary Table S1 Sequences of HA-tag and its mutants

Number	sequences
HA-tag	**YPYDVPDYA**
HA01(Y1G)	**GPYDVPDYA**
HA02(P2G)	Y**GYDVPDYA**
HA03(D4G)	YPY**GVPDYA**
HA04(V5G)	YPYD**GPDYA**
HA05(P6G)	YPYDV**GDYA**
HA06(D7G)	YPYDVP**GYA**
HA07(Y8G)	YPYDVPD**GA**
HA08(A9G)	YPYDVPDY**G**
HA09(Y3G)	YP**GDVPDYA**
HA10(Y1E)	**EPYDVPDYA**
HA11(P2E)	Y**EYDVPDYA**
HA12(Y3E)	YP**EDVPDYA**
HA13(D4E)	YPY**EVPDYA**
HA14(V5E)	YPYD**EPDYA**
HA15(P6E)	YPYDV**EDYA**
HA16(D7E)	YPYDVP**EYA**
HA17(Y8E)	YPYDVPD**EA**
HA18(A9E)	YPYDVPDY**E**
HA21(Y1H)	**HPYDVPDYA**
HA22(P2H)	Y**HYDVPDYA**
HA23(Y3H)	YP**HDVPDYA**
HA24(D4H)	YPY**HVPDYA**
HA25(V5H)	YPYD**HPDYA**
HA26(P6H)	YPYDV**HDYA**
HA27(D7H)	YPYDVP**HYA**
HA28(Y8H)	YPYDVPD**HA**
HA29(A9H)	YPYDVPDY**H**
HA31(V5A)	YPYD**APDYA**

HA32(V5T)	YPYD**T**PDYA
HA33(V5K)	YPYD**K**PDYA
HA34(V5R)	YPYD**R**PDYA
HA35(V5Y)	YPYD**Y**PDYA
HA36(V5S)	YPYD**S**PDYA
HA37(V5C)	YPYD**C**PDYA
HA38(V5L)	YPYD**L**PDYA
HA39(V5M)	YPYD**M**PDYA
HA40(V5I)	YPYD**I**PDYA
HA41(V5N)	YPYD**N**PDYA
HA42(V5F)	YPYD**F**PDYA
HA43(V5Q)	YPYD**Q**PDYA
HA44(V5D)	YPYD**D**PDYA
HA45(V5W)	YPYD**W**PDYA
HA46(V5P)	YPYD**P**PDYA
HA51(P6A)	YPYDV**A**DYA
HA52(P6T)	YPYDV**T**DYA
HA53(P6K)	YPYDV**K**DYA
HA54(P6Y)	YPYDV**Y**DYA
HA55(P6S)	YPYDV**S**DYA
HA56(P6C)	YPYDV**C**DYA
HA57(P6V)	YPYDV**V**DYA
HA58(P6L)	YPYDV**L**DYA
HA59(P6M)	YPYDV**M**DYA
HA60(P6I)	YPYDV**I**DYA
HA61(P6W)	YPYDV**W**DYA
HA62(P6F)	YPYDV**F**DYA
HA63(P6Q)	YPYDV**Q**DYA
HA64(P6D)	YPYDV**D**DYA
HA65(P6R)	YPYDV**R**DYA
HA66(P6N)	YPYDV**N**DYA
HA82(D4A)	YPY**A**VPDYA
HA83(D4C)	YPY**C**VPDYA

HA84(D4F)	YPY**F**VPDYA
HA85(D4I)	YPY**I**VPDYA
HA87(D4L)	YPY**L**VPDYA
HA88(D4M)	YPY**M**VPDYA
HA90(D4P)	YPY**P**VPDYA
HA93(D4S)	YPY**S**VPDYA
HA96(D4W)	YPY**W**VPDYA
HA99(Y3C)	YP**C**DVPDYA
HA100(Y3D)	YP**D**DVPDYA
HA101(Y3F)	YP**F**DVPDYA
HA102(Y3I)	YP**I**DVPDYA
HA103(Y3K)	YP**K**DVPDYA
HA104(Y3L)	YP**L**DVPDYA
HA105(Y3M)	YP**M**DVPDYA
HA106(Y3N)	YP**N**DVPDYA
HA107(Y3P)	YP**P**DVPDYA
HA108(Y3Q)	YP**Q**DVPDYA
HA109(Y3R)	YP**R**DVPDYA
HA110(Y3S)	YP**S**DVPDYA
HA111(Y3T)	YP**T**DVPDYA
HA112(Y3V)	YP**V**DVPDYA
HA113(Y3W)	YP**W**DVPDYA

 # the replaced residues were highlighted in red color.

Supplementary Table S2 Sequences of MYC-tag and its mutants

Number	sequences
MYC-tag	**EQKLISEEDL**
MYC01(Q2G)	E**G**KLISEEDL
MYC02(L4G)	EQK**G**ISEEDL
MYC03(I5G)	EQKL**G**SEEDL
MYC04(S6G)	EQKLI**G**EEDL
MYC05(E7G)	EQKLIS**G**EDL
MYC06(E8G)	EQKLISE**G**DL

MYC07(L4A)	EQK**A**ISEEDL
MYC08(L4C)	EQK**C**ISEEDL
MYC09(L4D)	EQK**D**ISEEDL
MYC10(L4E)	EQK**E**ISEEDL
MYC12(L4H)	EQK**H**ISEEDL
MYC13(L4I)	EQK**I**ISEEDL
MYC14(L4K)	EQK**K**ISEEDL
MYC15(L4M)	EQK**M**ISEEDL
MYC16(L4N)	EQK**N**ISEEDL
MYC17(L4P)	EQK**P**ISEEDL
MYC18(L4Q)	EQK**Q**ISEEDL
MYC19(L4R)	EQK**R**ISEEDL
MYC20(L4S)	EQK**S**ISEEDL
MYC21(L4T)	EQK**T**ISEEDL
MYC22(L4V)	EQK**V**ISEEDL
MYC23(L4W)	EQK**W**ISEEDL
MYC24(L4Y)	EQK**Y**ISEEDL
MYC27(I5A)	EQKL**A**SEEDL
MYC28(I5C)	EQKL**C**SEEDL
MYC29(I5D)	EQKL**D**SEEDL
MYC30(I5E)	EQKL**E**SEEDL
MYC31(I5F)	EQKL**F**SEEDL
MYC32(I5H)	EQKL**H**SEEDL
MYC33(I5K)	EQKL**K**SEEDL
MYC34(I5L)	EQKL**L**SEEDL
MYC35(I5M)	EQKL**M**SEEDL
MYC36(I5N)	EQKL**N**SEEDL
MYC37(I5P)	EQKL**P**SEEDL
MYC38(I5Q)	EQKL**Q**SEEDL
MYC39(I5R)	EQKL**R**SEEDL
MYC40(I5S)	EQKL**S**SEEDL
MYC41(I5T)	EQKL**T**SEEDL
MYC42(I5V)	EQKL**V**SEEDL

 # the replaced residues were highlighted in red color.

Supplementary Table S3 Sequences of Flag-tag and its mutants

Number	Sequences
Flag-tag	**DYKDDDDK**
Flag01(K3G)	DY**G**DDDDK
Flag02(K8G)	DYKDDDD**G**
Flag03(K3A)	DY**A**DDDDK
Flag04(K3C)	DY**C**DDDDK
Flag05(K3D)	DY**D**DDDDK
Flag06(K3E)	DY**E**DDDDK
Flag07(K3F)	DY**F**DDDDK
Flag08(K3H)	DY**H**DDDDK
Flag09(K3I)	DY**I**DDDDK
Flag10(K3L)	DY**L**DDDDK
Flag11(K3M)	DY**M**DDDDK
Flag12(K3N)	DY**N**DDDDK
Flag13(K3P)	DY**P**DDDDK
Flag14(K3Q)	DY**Q**DDDDK
Flag15(K3R)	DY**R**DDDDK
Flag16(K3S)	DY**S**DDDDK
Flag17(K3T)	DY**T**DDDDK
Flag18(K3V)	DY**V**DDDDK
Flag19(K3W)	DY**W**DDDDK
Flag20(K3Y)	DY**Y**DDDDK

 # the replaced residues were highlighted in red color.

Supplementary Table S4 Sequences of VSV-tag and its mutants

Number	Sequences
VSV-tag	**YTDIEMNRLGK**
VSV01(T2G)	Y**G**DIEMNRLGK
VSV02(I4G)	YTD**G**EMNRLGK
VSV03(M6G)	YTDIE**G**NRLGK

VSV04(N7G)	YTDIEM**G**RLGK
VSV05(R8G)	YTDIEMN**G**LGK
VSV06(T2A)	Y**A**DIEMNRLGK
VSV07(T2C)	Y**C**DIEMNRLGK
VSV08(T2D)	Y**D**DIEMNRLGK
VSV09(T2E)	Y**E**DIEMNRLGK
VSV10(T2F)	Y**F**DIEMNRLGK
VSV11(T2H)	Y**H**DIEMNRLGK
VSV12(T2I)	Y**I**DIEMNRLGK
VSV13(T2K)	Y**K**DIEMNRLGK
VSV14(T2L)	Y**L**DIEMNRLGK
VSV15(T2M)	Y**M**DIEMNRLGK
VSV16(T2N)	Y**N**DIEMNRLGK
VSV17(T2P)	Y**P**DIEMNRLGK
VSV18(T2Q)	Y**Q**DIEMNRLGK
VSV19(T2R)	Y**R**DIEMNRLGK
VSV20(T2S)	Y**S**DIEMNRLGK
VSV21(T2V)	Y**V**DIEMNRLGK
VSV22(T2W)	Y**W**DIEMNRLGK
VSV23(T2Y)	Y**Y**DIEMNRLGK
VSV26(M6A)	YTDIE**A**NRLGK
VSV27(M6C)	YTDIE**C**NRLGK
VSV28(M6D)	YTDIE**D**NRLGK
VSV29(M6E)	YTDIE**E**NRLGK
VSV30(M6F)	YTDIE**F**NRLGK
VSV31(M6H)	YTDIE**H**NRLGK
VSV32(M6I)	YTDIE**I**NRLGK
VSV33(M6K)	YTDIE**K**NRLGK
VSV34(M6L)	YTDIE**L**NRLGK
VSV35(M6N)	YTDIE**N**NRLGK
VSV36(M6P)	YTDIE**P**NRLGK
VSV37(M6Q)	YTDIE**Q**NRLGK
VSV38(M6R)	YTDIE**R**NRLGK

VSV39(M6S)	YTDIE**S**NRLGK
VSV40(M6T)	YTDIE**T**NRLGK
VSV41(M6V)	YTDIE**V**NRLGK
VSV42(M6W)	YTDIE**W**NRLGK
VSV43(M6Y)	YTDIE**Y**NRLGK
VSV35(M6N)	YTDIE**N**NRLGK
VSV36(M6P)	YTDIE**P**NRLGK
VSV37(M6Q)	YTDIE**Q**NRLGK
VSV38(M6R)	YTDIE**R**NRLGK
VSV39(M6S)	YTDIE**S**NRLGK
VSV40(M6T)	YTDIE**T**NRLGK
VSV41(M6V)	YTDIE**V**NRLGK
VSV42(M6W)	YTDIE**W**NRLGK
VSV43(M6Y)	YTDIE**Y**NRLGK

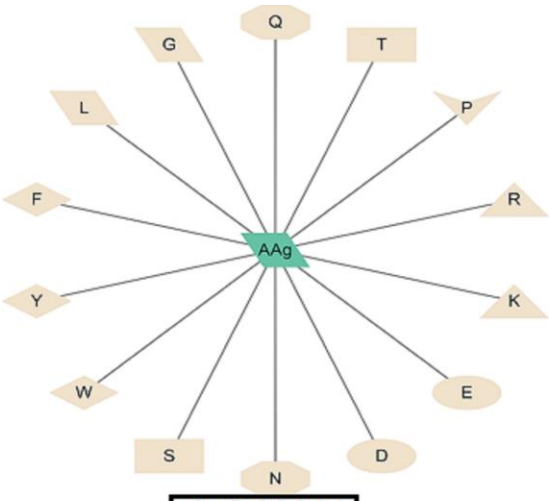
 # the replaced residues were highlighted in red color.

Supplementary Table S5 RMSD of the fitted structures of mutant tags and original protein-tags

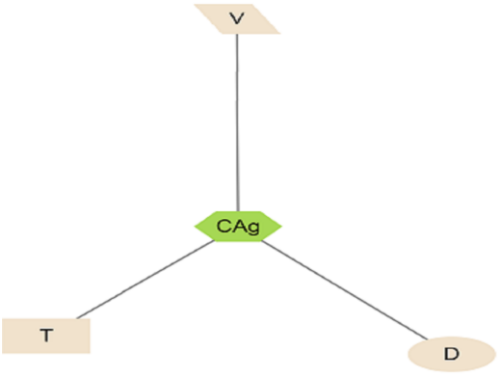
Mutant	RMSD(Å)
HA-Y1G	0.776
HA-Y3G	1.594
HA-Y8G	**4.496**
MYC-E8G	2.656
MYC-E7G	**3.835**
Flag-K3G	**4.682**
Flag-K8G	0.408
HA-V5L	1.825
HA-V5I	1.885

HA-Y3F	0.332
HA-Y3W	2.457
HA-D4E	1.845
VSV-M6C	2.799
MYC-I5L	0.557
HA-D4F	**4.139**
HA-D4W	**3.401**
HA-D4Y	**4.301**
HA-P6F	**3.792**
HA-P6W	**4.189**
HA-P6Y	**4.214**
HA-V5F	**4.760**
HA-V5W	**3.509**
HA-V5Y	**3.650**
VSV-M6F	**4.550**
VSV-M6W	**4.212**
VSV-M6Y	**4.175**

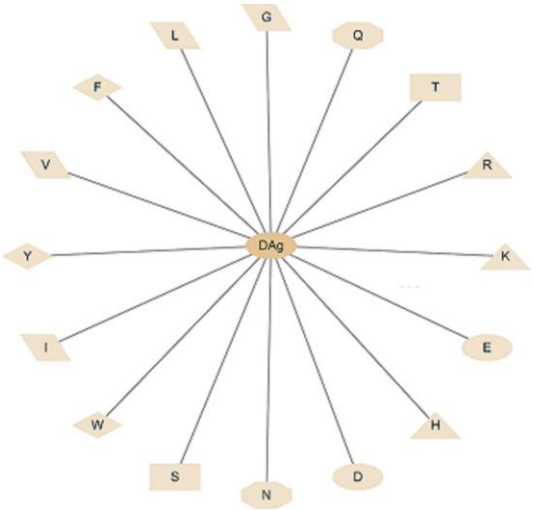
RMSD < 3.0 referred higher goodness of fit.



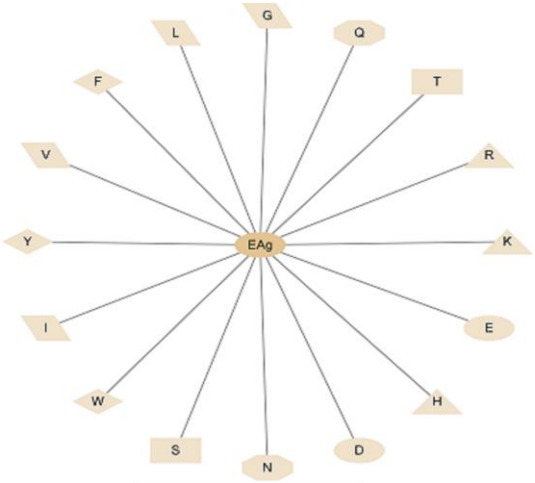
A C H I M



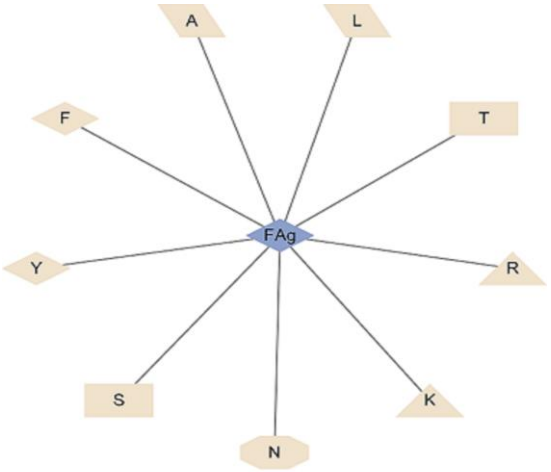
A C E F G H I K L
M N P Q R S W Y



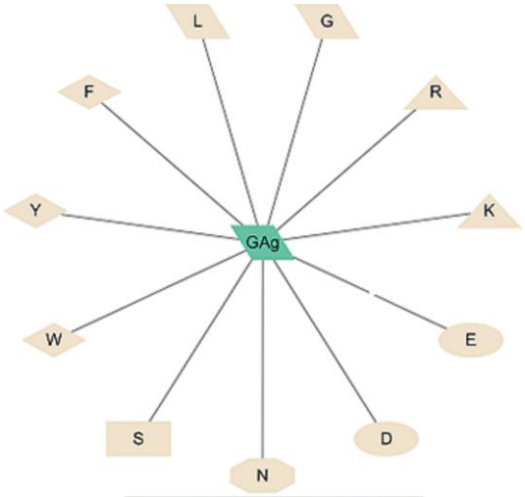
A C M P



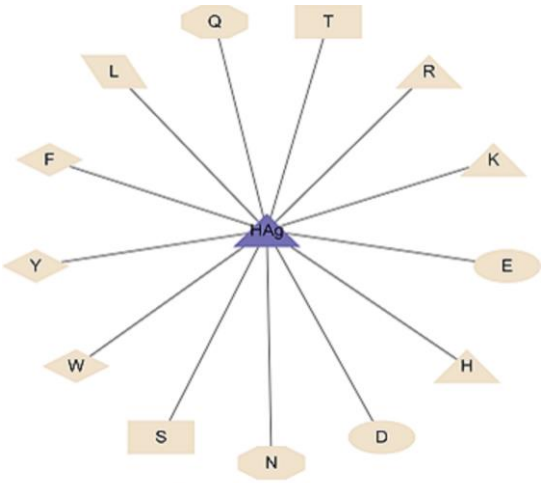
A C M P



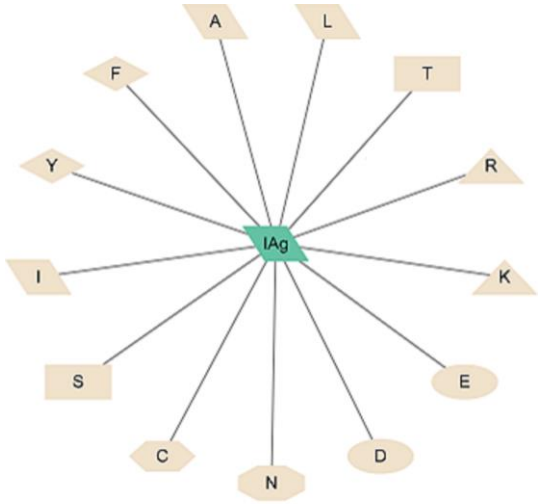
C D E G H I M P Q W



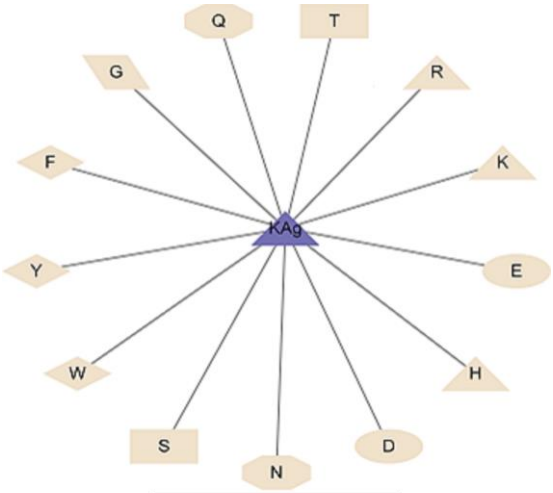
A C H I M P Q T V



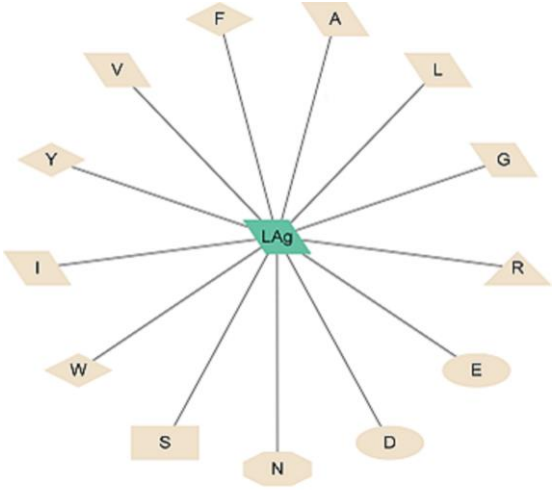
A C G I M P V



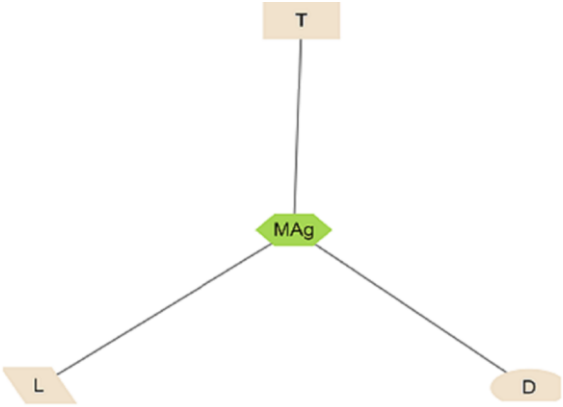
G H M P Q V W



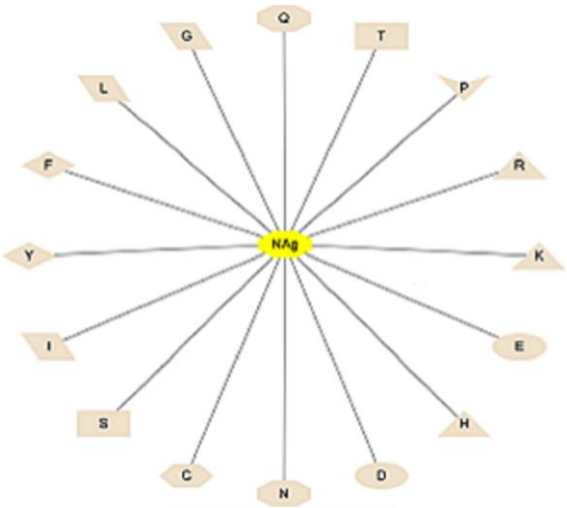
A C I L M P V



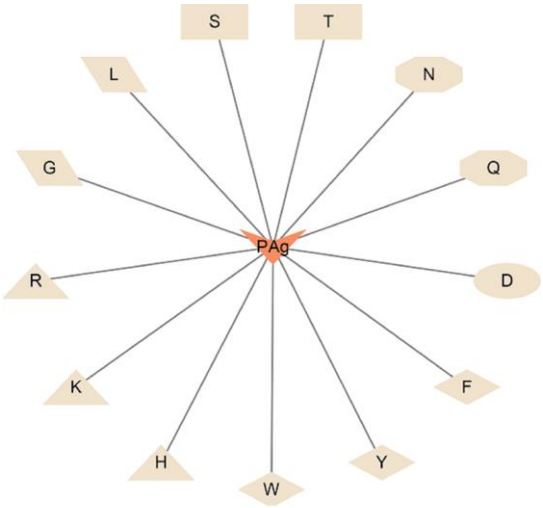
C H K M P Q T



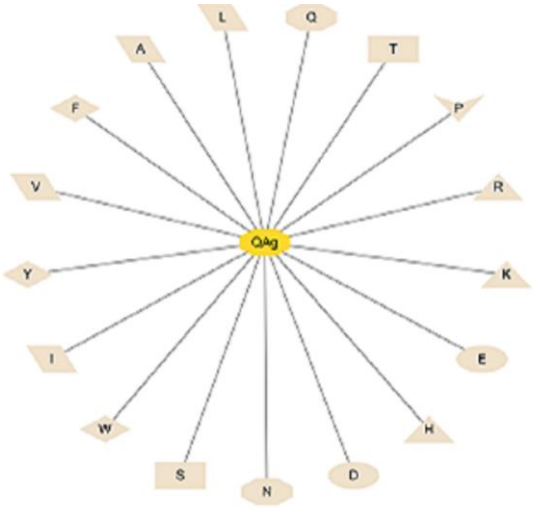
A C E F G H I K M
N P Q R S V W Y



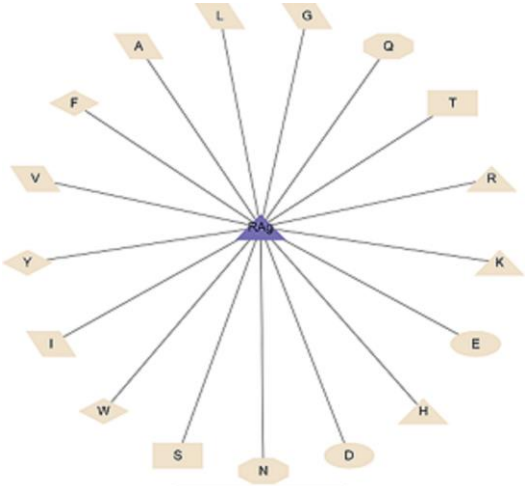
A M V W



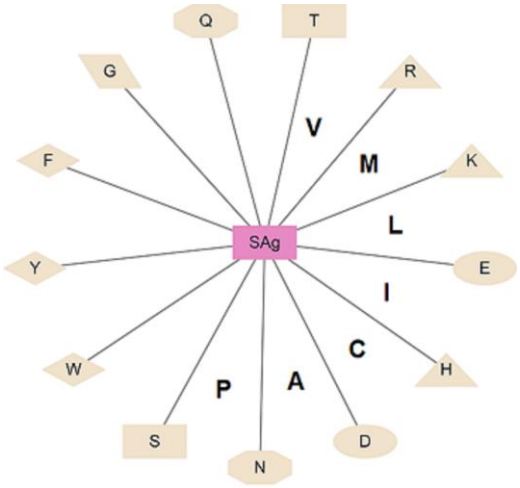
A C E I M P V



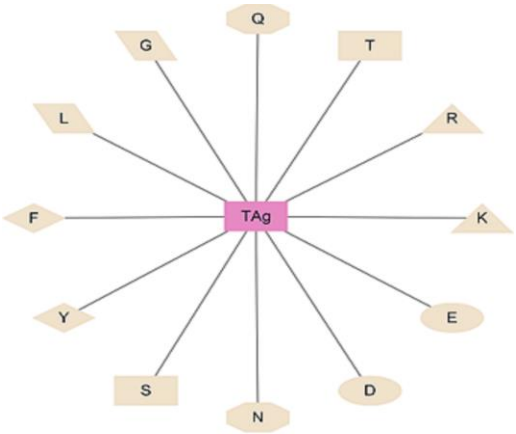
C G M



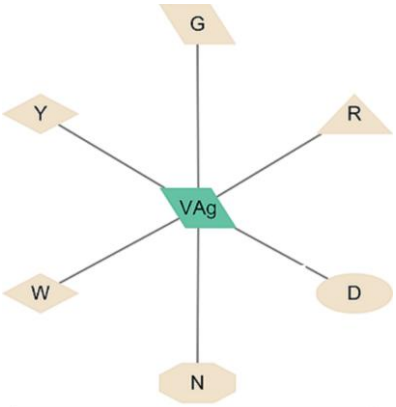
C M P



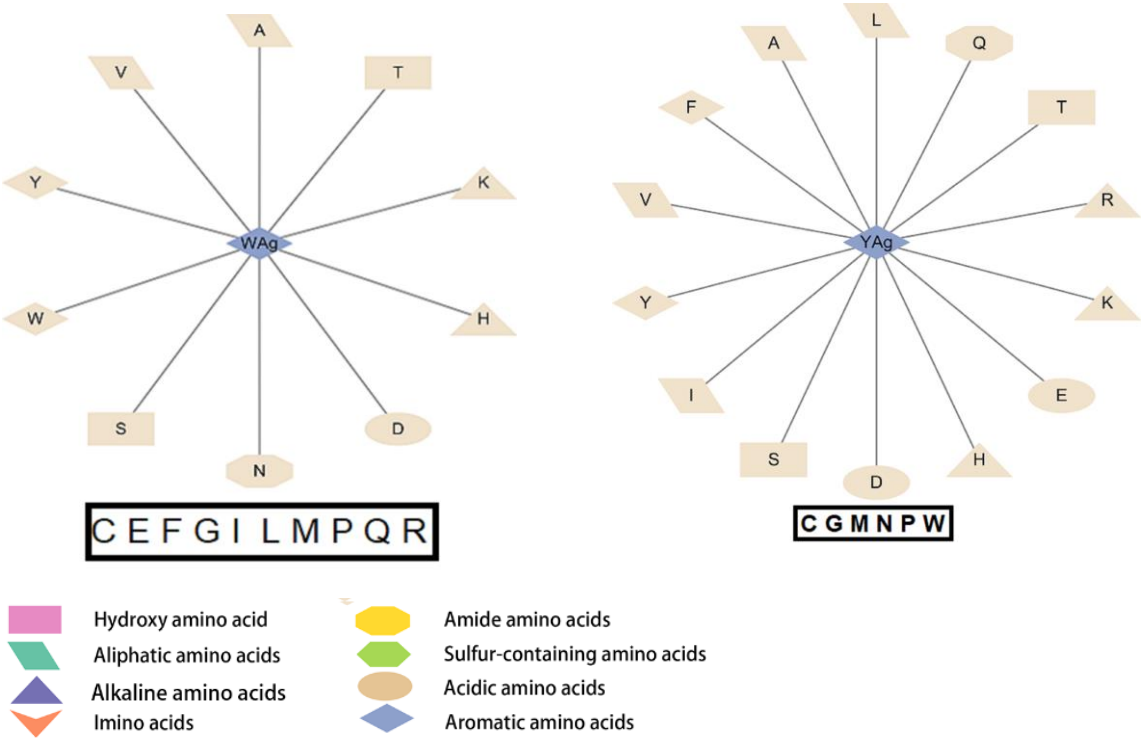
A C I L M P V



A C H I M P V W



A C E F H I K L M P Q S T V



Supplementary Figure S1 20 antigenic amino acids interaction plotted individually. Antibody amino acids without interaction were listed in boxes below the main figure in bold type. The antigenic amino acid abbreviation was labeled with the subscript Ag to distinguish antibody amino acids.