Genomics 100 (2012) 102–109

Contents lists available at SciVerse ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Highly conserved influenza A virus epitope sequences as candidates of H3N2 flu vaccine targets

Ko-Wen Wu [a], Chih-Yi Chien [a], Shiao-Wen Li [a,b], Chwan-Chuen King [c,d], Chuan-Hsiung Chang [a,e,*]

[a] Institute of BioMedical Informatics, National Yang-Ming University, Taipei, Taiwan
[b] Bioinformatics Program, Taiwan International Graduate Program, Academia Sinica, Taipei, Taiwan
[c] Institute of Epidemiology, College of Public Health, National Taiwan University, Taipei, Taiwan
[d] Department of Public Health, College of Public Health, National Taiwan University, Taipei, Taiwan
[e] Center for Systems and Synthetic Biology, National Yang-Ming University, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 1 February 2012
Accepted 4 June 2012
Available online 12 June 2012

Keywords:
Influenza virus
Epitope
Broad-spectrum vaccine
Bioinformatics

ABSTRACT

This study focused on identifying the conserved epitopes in a single subtype A (H3N2)—as candidates for vaccine targets. We identified a total of 32 conserved epitopes in four viral proteins [22 HA, 4PB1, 3 NA, 3 NP]. Evaluation of conserved epitopes in coverage during 1968–2010 revealed that (1) 12 HA conserved epitopes were highly present in the circulating viruses; (2) the remaining 10 HA conserved epitopes appeared with lower percentage but a significantly increasing trend after 1989 [$p<0.001$]; and (3) the conserved epitopes in NA, NP and PB1 are also highly frequent in wild-type viruses. These conserved epitopes also covered an extremely high percentage of the 16 vaccine strains during the 42 year period. The identification of highly conserved epitopes using our approach can also be applied to develop broad-spectrum vaccines.

1. Introduction

Influenza A viruses are negative-sense RNA viruses that possess a segmented genome of eight single-stranded RNA segments and encode eleven proteins. These viruses undergo genetic drifts and shifts due to error-prone replication of viral genomes and reassortment of viral gene segments [1]. Consequently, influenza A viruses escape that immune system of their hosts resulting in influenza epidemics and pandemics.

Antigenic drift is the continuous process of genetic and antigenic changes that occur through point mutations. On average it occurs every two to eight years in response to selection pressure to evade human immunity [2]. The antigenic distance between flu strains is increasing with time by the drift. To address this, influenza vaccines are reviewed annually to ensure protection is maintained despite the emergence of drift variants [3]. Moreover, the reassortment of influenza A viruses leads to antigenetic and genetic variabilities and dynamic interactions of viral gene segments of influenza viruses derived from different species of hosts. In the past century, influenza viruses had caused several major pandemics, namely the 1918 Spanish pandemic (H1N1), the 1957 Asian pandemic (H2N2) and the 1968 Hong Kong pandemic (H3N2) [4]. These pandemic viruses obtained novel HA, NA and PB1 gene segments either through direct animal (poultry)-to-human transmission or through mixing of human influenza A and animal influenza A virus genes to create a new human influenza A subtype virus under a process called genetic reassortment, resulting in antigenic shift, phenotypic change and spread among human populations without prior immunity [5–7].

In facing such challenges and needs against the changing components of influenza vaccine, one possible approach of developing a universal influenza vaccine is to design a vaccine using the conserved immunogenic regions in the viral proteins to protect human population against influenza viruses. The universal influenza vaccines currently under development were designed mainly based on the highly conserved external domain of the influenza matrix 2 (M2) protein and conserved epitopes from the influenza NP, matrix 1 (M1), and HA proteins. In 2007, a computational prediction of HLA restricted T-cell epitope sequences in the conserved regions of influenza A virus proteins was performed [8]. However, those predicted epitopes still need to be verified by further experiments [9]. In 2010, one study demonstrated a hemagglutinin subunit 2 protein (HA2)-based synthetic peptide vaccine that provides protection in mice against influenza viruses of the structurally divergent subtypes H3N2, H1N1, and H5N1 [10].

Due to the antigenic variability of influenza A viruses among different subtypes, only few viral epitopes have been shown to induce host immune response. A highly conserved domain of the M2 protein (M2e) in almost all influenza A viruses was identified [11]. Recently, it has been reported that there is only one HA conserved epitope in influenza A viruses recognized by antibodies [12].

Practically, a universal vaccine might restrict the ability of protection against particular subtypes of influenza A viruses. A universal antibody

* Corresponding author at: Center for Systems and Synthetic Biology, Institute of Biomedical Informatics, National Yang-Ming University, No. 155, Sec. 2, Li-Nong St., Taipei 11221, Taiwan. Fax: +886 2 2820 6754.
E-mail address: cchang@ym.edu.tw (C.-H. Chang).

0888-7543/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2012.06.003

identified by previous studies should be against all influenza A viruses but the antibody failed to neutralize the viruses of H3 and H7 [13]. Therefore, Sui et al. argued that a cocktail vaccine comprising broad-spectrum antibodies of different subtypes could provide broad protection against all seasonal and pandemic influenza A viruses.

Several recent studies have used bioinformatics approach to identify conserved domains across influenza A viral strains. One study discovered 19 universal conserved functional motifs and accessible regions for functional mapping and annotation [14]. Another study reported motifs within the HA protein and correlated them with a high number of potential post-translational modification sites that overlap antigenic and receptor binding sites [15,16]. A third study performed phylogenic analysis and addressed the evolution of influenza A viral segments [16]. However, the analysis of protein and epitope sequences between influenza A/H3N2 viruses and vaccine strains remains to be elucidated.

Therefore, this study focused on identifying conserved epitopes in a single subtype rather than in all subtypes. We conducted a sequence-based analysis of available influenza B-cell and T-cell epitopes for human as well as measured sequence conservations of epitopes in influenza A/H3N2 viruses, the most rapid changing subtype, to provide the information with statistical methods for the candidates of broad-spectrum vaccine. The aims of this study were: (1) to investigate the viral diversity in influenza A/H3N2 viruses for finding representative viral clusters, (2) to identify the conserved epitopes in all viral proteins and (3) to evaluate the coverage of the conserved epitopes among various influenza strains.

2. Results

2.1. Distribution of clusters in different influenza proteins

Peptide sequences of each of all the non-identical and full-length amino acids in the 11 viral protein sequences of influenza A/H3N2 viruses were retrieved directly from the NCBI Influenza Resources Database [17]. A total of 7690 protein sequences were retrieved (Table 1).

The k-means clustering algorithm was used to group viral amino acid sequences into clusters. PB1-F2, NP and M1formed 15 clusters; NS1 formed 10 clusters; NA and M2 formed 6 clusters; PB2, PB1, PA, HA and NS2 datasets formed 5 clusters (Table 1). A consensus sequence was then generated as a representative sequence for each cluster.

2.2. Epitope sequences conserved in influenza A/H3N2 viruses

After performing the 5-fold cross-validation (with *t*-test) between the training and testing sequence data sets, we identified a total of 32 conserved epitopes (Table 2). These epitopes were located in only four of the proteins of influenza A/H3N2: HA (22 epitopes), NA (3 epitopes), PB1 (4 epitopes) and NP (3 epitopes). In general, more epitopes were found in the structural HA and NA proteins than in viral internal proteins, using currently available epitope data sources. From an immunological perspective, the conserved epitopes involved 31T-cell epitopes and only one B-cell epitope (see Supplementary data 1–4). In addition, most of the identified T-cell epitopes were CD4+ T-cell epitopes (26/32, 81.3%), including 21 HA (20.6%); 3 NA; 1PB1, and 1 NP. On the other hand, only 5 conserved CD8+ T-cell epitopes were identified [3 PB1 (30.00%), 2 NP, 1 NA].

Table 1
The number of influenza A/H3N2 virus protein sequences used in this study and the number of representative clusters identified by k-means clustering.

Protein	Number of sequences (%)	Number of clusters
NA	1789 (23.26)	6
HA	1619 (21.1)	5
PA	721 (9.38)	5
PB1	695 (9.0)	5
PB2	689 (9.0)	5
NS1	628 (8.2)	10
NP	432 (5.6)	15
PB1-F2	426 (5.5)	15
M2	346 (4.5)	6
M1	199 (2.6)	15
NS2	146 (1.9)	5
Total	7690 (100)	92

2.3. Conserved epitopes showed high identities among flu vaccine strains during 1968–2010

For practical concerns of vaccine development, we evaluated the coverage and sequence identities of our identified conserved epitopes against the conserved epitopes among various influenza vaccine strains. Figs. 1–2 show that these conserved epitopes had over 96% sequence identities to the sequence of flu vaccine strains during 1968–2010 [HA in 16 strains: 96–100% ± 0.00–4.13%, NA in 12 strains: 99–100% ± 0.00–2.34%, NP in 8 strains: 99–100% ± 0.00–2.21% and PB1 in 4 strains: 100% ± 0.00%]. These results revealed the feasibility of preparing a broad-spectrum vaccine, using the amino acid sequence information from the conserved epitopes.

2.4. Conserved epitopes existed in influenza A/H3N2 viruses during 1968–2010

To find out whether the identified conserved epitopes were present in circulated H3N2 strains over the past years, we calculated the percentages of conserved epitopes existing in all H3N2 viruses. The conserved epitopes during 1968–2010 could be classified into three types, including: (1) 12 frequently present HA conserved epitopes (8 in HA1 and 4 in HA2), with mean ± standard deviation (S.D.) of 93–98 ± 5.37–16.94%, (2) 10 rarely present the remaining 10 HA conserved epitopes (8 in HA1 and 2 in HA2) [mean ± S.D.: 47–86% ± 29.45%–47.90%], with a significant increase in percentage of the presence in wild-type viruses after 1989 than those before 1988 [after 1989 vs. before 1988: 94% ± 14.1 vs. 34% ± 45.0, $p<0.001$; HA2: 93 ± 18.4% vs. 63 ± 45.9%, $p<0.001$]; and (3) all the 4 NA, 3 NP and 3PB1 conserved T-cell epitopes also displayed with high percentages in all the available sequences of the wild-type viruses during the studied 42 years [95–97% ± 3.09–8.69%, 99–100% ± 0.54–2.09%, 100% ± 0.12–0.99%] (Figs. 3–4).

3. Discussion and conclusion

There is a need to develop a universal vaccine against all types of influenza virus to provide long-lasting and cross-strain protection. By identifying the conserved regions in the proteins of the influenza virus without antigenic variability by strain or over time, we provide candidates for the development of epitope-based universal influenza vaccine. In this study we carried out computational analysis with statistical inference to identify 32 conserved epitopes in influenza A/H3N2 viral proteins. The conserved epitopes have high sequence identities (>99%) in most of flu vaccine strains during 1957–2009 and high conservation in influenza viruses over the past 33 years (1968–2010). Further results indicated that conserved epitopes had over 96% sequence identities in flu vaccine strains, and that these epitopes existed frequently (96–100%) in wild-type vaccines during 1989–2010.

3.1. Feasibility of epitope-based vaccine developed in the future

Recent literature reported that the antibody CR8020 has broad neutralizing activity against most of the group 2 influenza viruses, including H3N2 and H7N7 [18]. The antibody recognizes the conserved epitope (GIFGAIAGFIENGWEGMVDGWYGFR, 346–371 aa) located in the HA2 domain (346–566 aa)—a conserved region in the HA protein that plays a role in the fusion activity of influenza infection [19]. Furthermore, previous studies indicated that HA2-based synthetic

Table 2
The number of influenza A/H3N2 epitopes identified in 11 viral proteins.

Protein	Epitope											
	B cell			CD4+ T cell			CD8+ T cell			Overall		
	Conserved	Total	Percentage	Conserved	Total	Percentage	Conserved	Total	Percentage	Conserved	Total	Percentage
HA	1	23	4.35	21	102	20.59	0	5	0.00	22	130	16.92
PB1	0	5	0.00	1	7	14.29	3	10	30.00	4	22	18.18
NP	0	3	0.00	1	62	1.61	2	33	6.06	3	98	3.06
NA	0	6	0.00	3	78	3.85	0	1	0.00	3	85	3.53
M1	0	4	0.00	0	70	0.00	0	15	0.00	0	89	0.00
PB2	0	1	0.00	0	2	0.00	0	3	0.00	0	6	0.00
PB1-F2	0	3	0.00	0	0	0.00	0	0	0.00	0	3	0.00
PA	0	2	0.00	0	0	0.00	0	1	0.00	0	3	0.00
M2	0	4	0.00	0	0	0.00	0	3	0.00	0	7	0.00
NS1	0	2	0.00	0	5	0.00	0	5	0.00	0	12	0.00
NS2	0	1	0.00	0	2	0.00	0	0	0.00	0	3	0.00
Total	1	54	1.85	26	328	7.93	5	76	6.58	32	458	6.99

peptide vaccine can provide broad protection against influenza viruses including H3N2 viruses [10,20]. HA2-specific antibodies can inhibit the fusion activity of influenza HA proteins to reduce the ability of viral replication [21].

As mentioned above, previous studies showed that the immune responses induced by HA2-based synthetic peptide vaccine can prevent the influenza infection. The HA conserved epitopes (IEDB ID: 97341, 346–363 aa) located in the HA2 domains and the conserved epitopes could be responsible for the induction of antigenicity against influenza A/H3N2 viruses (see Supplementary data 5). Furthermore, conservations of other HA conserved epitopes are greater than those of HA conserved epitopes. It suggested that there were more potential epitopes that could be candidates for the universal vaccine.

The NA molecule is another major protein and antigen on the viral surface and is involved in the release of the viral particles from infected host cells. The NA gene has a high nucleotide substitution rate, resulting in antigenic drift [22]. There are several ways to produce protection against influenza viruses, one of them is to raise antibodies against HA protein, hence block the binding of viruses to sialic acids on the host cell and preventing viral infection. Antibodies binding to NA stop the release of viruses from infected cells and limit the spread of the virus [23]. The list of NA conserved epitopes can be used to induce antibodies against influenza viruses to limit its spread. Furthermore, antibodies will only bind specifically to the functional regions to prevent viral release and allow patients to recover from infection.

Viral genomic RNAs, polymerase and NP protein are incorporated into virions as ribonucleoprotein complexes. Although NP is an internal protein located inside the virions, studies have shown that NP may contribute to protect immunity in animal model against diverse subtypes of influenza viruses through its ability to stimulate cellular immunity [24,25]. Many vaccination strategies have used the NP protein as an antigen to induce immune responses since it is well-conserved across influenza virus subtypes [26]. We identified three NP conserved epitopes including one CD4+ and two CD8+ T-cell epitopes with high sequence conservation. Therefore, an epitope-based vaccine developed from NP protein or viral antigens in combination with NP protein may provide higher protective efficacy against influenza viruses and be considered as a candidate for the universal vaccine design.

The PB1 protein is a subunit of the polymerase complex in influenza viruses. Sequence comparisons have revealed that the PB1 is the most conserved segment in human strains [14]. Due to its high sequence conservation and essentiality, PB1 has recently been thought as an ideal target for drug binding [27]. An earlier report suggested that PB1 can be considered for vaccine development since PB1 had contributed the highest number of epitopes for both class I- and class II-restricted responses in that study [28]. The PB1 conserved epitopes (IEDB ID: 97653) in our results are almost identical, we showed that the epitopes may be grouped into the same clusters to show similar properties. The PB1 conserved epitopes that showed less than 100% conservation may provide more choices for universal vaccine design. However, unlike HA or NA surface poteins, PB1 may not be the primary target recognized by the host immune system. Another application is to use the conserved epitopes for the design of drug target because of the biological importance of the conserved regions.

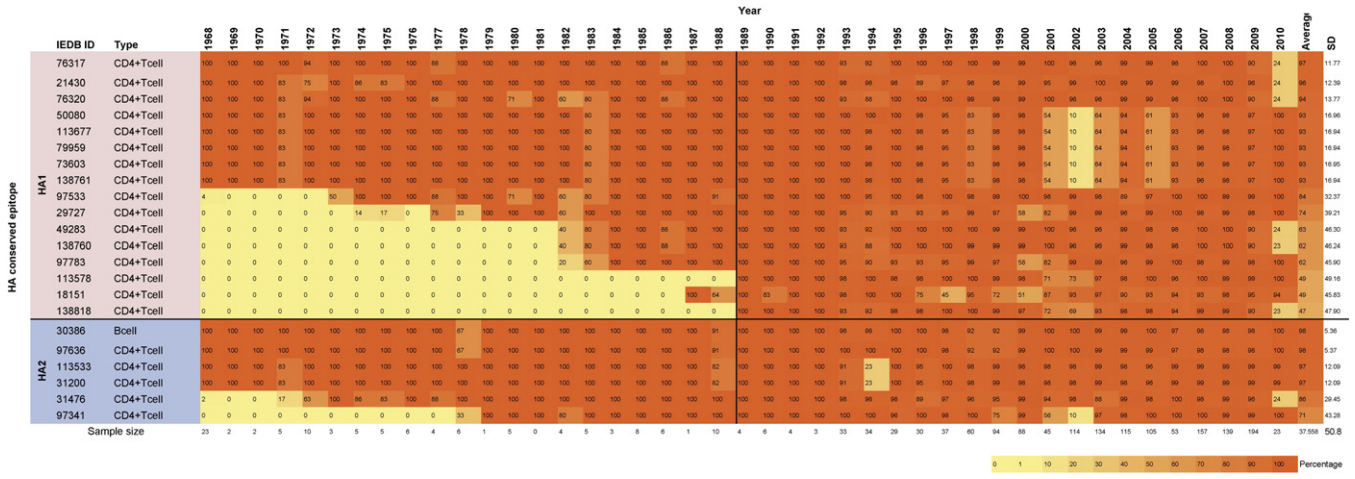

Fig. 1. The percentage of influenza A/H3N2 viruses containing HA conserved epitopes during the 1968–2010 period. The percentage was calculated as described in Materials and methods.

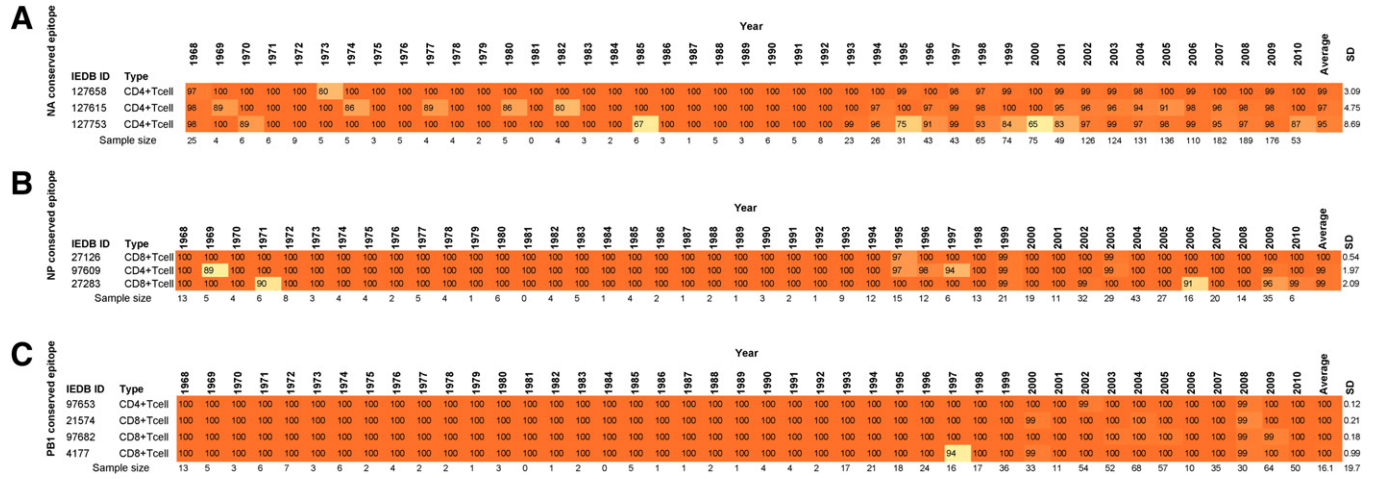

Fig. 2. The percentage of influenza A/H3N2 viruses containing NA (panel A), NP (panel B), and PB1 (panel C) conserved epitopes during the 1968–2010 period. The percentage was calculated as described in Materials and methods.

3.2. Problems of current flu vaccines and future directions of flu vaccine design

Antibodies induced by B cells play a key role in protection against influenza infection *in vivo* [29]. Therefore, B-cell epitopes are an important component in the design of traditional vaccines against influenza A/H3N2 viruses. Unlike the abundance of T-cell epitopes, there was only one B-cell epitope identified in our study, and it is located in the HA2 domain. This conserved epitope may also be considered as a candidate for the universal vaccine.

During influenza virus infection, cellular immunity induced by CD4+ and CD8+ T cells plays an important role to eliminate virus-infected cells. In this study, 26 CD4+ and 5 CD8+ T-cell conserved epitopes were found in HA, NA, NP and PB1 proteins. Previous studies showed that T-cell response can reduce the severity of viral infection in human [30,31], and the memory T-cell immunity can protect the human population against influenza viruses [32]. Therefore, the conserved T-cell epitopes may be used as vaccine candidates and protect humans from serious viral infection.

The M2e epitope identified in previous studies showed cross-reactivity against various strains and different subtypes of influenza viruses [11,33]. However, the M2e epitope is not a conserved epitope defined by this study. A possible explanation is that the different levels of genomic diversity of influenza A/H3N2 viruses were observed at different times over different geological regions [34,35]. For this reason, the heterosubtypic epitope might not provide maximum immunological protection against influenza viruses that diverged quickly. On the other hand, our current analysis did not consider the cross-reactivity

	Epitope ID	Type	Sequence	Vaccine strain: A/Victoria/3/1975	A/Texas/1/1977	A/Bangkok/1/1979	A/Christchurch/4/1985	A/Mississippi/1/1985	A/Leningrad/360/1986	A/Sichuan/2/1987	A/Shanghai/11/1987	A/Guizhou/54/89	A/Beijing/353/1989	A/Johannesburg/33/1994	A/Wuhan/359/1995	A/Sydney/5/1997	A/Moscow/10/1999	A/Fujian/411/2002	A/Perth/16/2009	Average	SD
HA1	50080	CD4+Tcell	PYDVPDYASLRSLVA	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0.00
	73603	CD4+Tcell	YDVPDYASLRSLVASS	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0.00
	76317	CD4+Tcell	YVKQNTLKL	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0.00
	76320	CD4+Tcell	YVKQNTLKLATGMRNV	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0.00
	79959	CD4+Tcell	PDYASLRSLVASS	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0.00
	113677	CD4+Tcell	PDYASLRSLVASSG	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0.00
	138761	CD4+Tcell	CYPYDVPDYASLRSLVASSG	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0.00
	21430	CD4+Tcell	GNGCFKIYHKCDNACI	100	NA	100	NA	NA	100	NA	NA	NA	100	NA	NA	100	100	NA	100	100	0.00
	97533	CD4+Tcell	NVPEKQTRGIFGAIAGFI	100	91	100	100	100	100	100	100	100	100	100	100	100	100	100	100	99	2.25
	138760	CD4+Tcell	CPRYVKQNTLKLATGMRNVP	95	95	95	100	100	100	100	100	100	100	100	100	100	100	100	100	99	2.02
	29727	CD4+Tcell	IYWTIVKPGDILLINS	93	100	100	93	100	100	100	100	100	100	100	100	93	100	100	100	99	2.82
	49283	CD4+Tcell	PRYVKQNTLKLAT	92	92	92	100	100	100	100	100	100	100	100	100	100	100	100	100	99	3.22
	97783	CD4+Tcell	YWTIVKPGDILLINSTGNL	89	94	94	94	94	100	100	94	100	100	100	100	94	100	100	100	97	3.64
	18151	CD4+Tcell	FVERSKAYSNCYPYDV	93	93	93	93	93	87	100	100	100	100	100	100	100	100	100	100	97	4.24
	138818	CD4+Tcell	VNRITYGACPRYVKQNTLKL	90	90	90	95	95	95	95	95	100	100	100	100	100	100	100	100	97	3.97
	113578	CD4+Tcell	KPFQNVNRITYGA	92	92	92	92	92	92	92	92	100	100	100	100	100	100	100	100	96	4.13
HA2	30386	Bcell	KEFSEVEGRIQDLEKYV	100	NA	100	NA	NA	100	NA	NA	NA	100	NA	NA	100	100	NA	100	100	0.00
	31200	CD4+Tcell	KIDLWSYNAELLVALE	100	NA	100	NA	NA	100	NA	NA	NA	100	NA	NA	100	100	NA	100	100	0.00
	31476	CD4+Tcell	KIYHKCDNACIGSIRN	100	NA	100	NA	NA	100	NA	NA	NA	100	NA	NA	100	100	NA	100	100	0.00
	97636	CD4+Tcell	SEVEGRIQDLEKYVEDTK	100	NA	100	NA	NA	100	NA	NA	NA	100	NA	NA	100	100	NA	100	100	0.00
	113533	CD4+Tcell	IDLWSYNAELLVAL	100	NA	100	NA	NA	100	NA	NA	NA	100	NA	NA	100	100	NA	100	100	0.00
	97341	CD4+Tcell	GIFGAIAGFIENGWEGMV	94	NA	100	NA	NA	100	NA	NA	NA	100	100	100	100	94	100	100	99	2.53

Fig. 3. Sequence identities of HA conserved epitopes among influenza vaccine strains. The values were calculated by using the BLAST tool. NA: The epitope is not found in the vaccine strains because sequences of those vaccine strains were not full-length. Average: A of epitope conservations in flu vaccines. SD: Standard deviation of epitope conservations in flu vaccines.

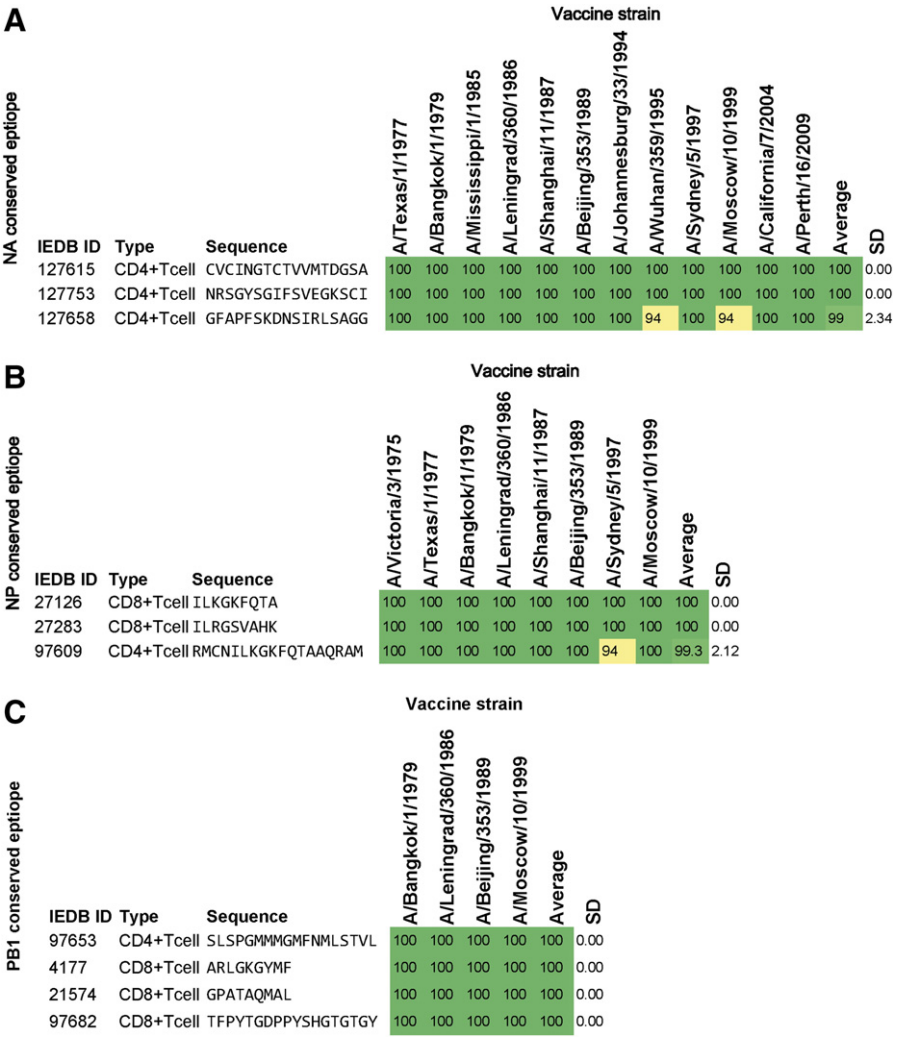

Fig. 4. Sequence identities of NA (panel A), NP (panel B), and PB1 (panel C) conserved epitopes among influenza vaccine strains. The values were calculated by using the BLAST tool. NA: The epitope is not found in the vaccine strains because sequences of those vaccine strains were not full-length. Average: A of epitope conservations in flu vaccines. SD: Standard deviation of epitope conservations in flu vaccines.

of the substitutions in epitopes, and further work is needed to extend such analysis.

3.3. High population coverage of influenza vaccines

A critical thing for T-cell epitope-based vaccine strategy is to identify and select T-cell epitopes that bind to several alleles of HLA supertypes to reach maximal population coverage [36]. Here, the HLA types of conserved epitopes in this study include A3, A28, B7, B8, B27, DR1, DRA, DRB1 and DRB5 (see Supplementary data 6–10). High population coverage can be achieved in most prominent ethnicities by focusing on only three major HLA class I types: A1, A3 and B7 [37,38]. In previous studies, HLA class I A2, A3 and B7 supertype-restricted epitopes conserved among different influenza subtypes were identified, and it can be of relevance for the development of a potential type-restricted, T-cell epitope-based vaccine [8,28,39,40].

3.4. Can these conserved epitopes also be found in avian and/or swine H3N2 isolates or in influenza A/H1N1?

There is one HA conserved epitope (IEDB ID: 97341, GIFGAIAGFIE-NGWEGMV, 346–363 aa) very similar to the avian epitope reported in IEDB database (IEDB ID: 20838, GLFGAIAGFIE, 345–355 aa). The epitope sequence of the hemagglutinin of avian influenza A virus H1N1 at amino acid residues 345–355 shares about 90% protein sequence identity with human influenza A/H3N2. This is the only epitope that was found to be common to avian and human influenza isolates, while there is no conserved epitope found between human and swine influenza isolates.

A total of 32 conserved epitopes were identified in HA, NA, PB1 and NP proteins of human influenza A/H3N2. Sequence comparison of the conserved epitopes against the strains H1N1, H5N1 and H9N2 showed that 22 HA and 3 NA conserved epitopes were unique to influenza A/H3N2. For NP epitopes, two of three were found in both A/H3N2 and A/H5N1 (IEDB ID: 27126, 225–233 aa; IEDB ID: 27283, 265–273 aa). For PB1 epitopes, two were found to be common in both A/H3N2 and A/H5N1 (IEDB ID: 97653, 402–419 aa; IEDB ID: 97682, 21–38 aa), and two epitopes were found to be common in A/H3N2, A/H1N1, and A/H9N2 (IEDB ID: 4177, 349–357aa; IEDB ID: 21574, 540–548 aa).

4. Conclusion

This study was designed to identify the candidates of highly conserved epitopes of influenza A/H3N2 viruses for the broad-spectrum epitope-based vaccine design. Conserved epitopes were identified in several proteins including: HA, NA, NP and PB1 proteins. We proposed that the conserved epitopes identified in this study be considered as

candidates for broad-spectrum epitope-based vaccine design. Our results have suggested that epitopes with robust conserved patterns in influenza A/H3N2 viral proteins are promising functional candidates for broad-spectrum epitope-based vaccine design.

4.1. Future efforts

We have computationally identified conserved epitopes of influenza A/H3N2 viruses with the support of statistical tests. Developing broad-spectrum vaccines against influenza viruses requires a fast and robust strategy and procedure. For preventing or controlling flu outbreaks, the technology of producing abundant epitope-based vaccines can be achieved by using synthetic epitope peptides. Finally, we need to understand and advance our knowledge on influenza viruses and provide a more efficient preventive method to protect humans from the threats of influenza viruses.

5. Materials and methods

5.1. Approach overview

The main objective of this study is to identify epitopes that are highly conserved over long periods of time within existing influenza A/H3N2 viruses. Fig. 5 shows an overall study design involving a six-step process employed in this study. Before calculating the conservations of epitope sequences, we considered that some highly similar sequences in influenza viruses could affect the calculation of epitope conservations. Therefore, we grouped the viruses that share similar viral protein sequences by clustering analysis and generated the consensus sequences of each cluster. Epitope conservations were then estimated using those consensus sequences. In addition, to evaluate the statistical significance of epitope conservations, we calculated the means and standard errors of conservations by random sampling for each subset. Epitopes conserved in up to 99% influenza A/H3N2 viruses were defined as conserved epitopes. Furthermore, to evaluate the coverage among various influenza strains, we investigated whether conserved epitopes existed in the past years and whether they showed high sequence identities in flu vaccine strains.

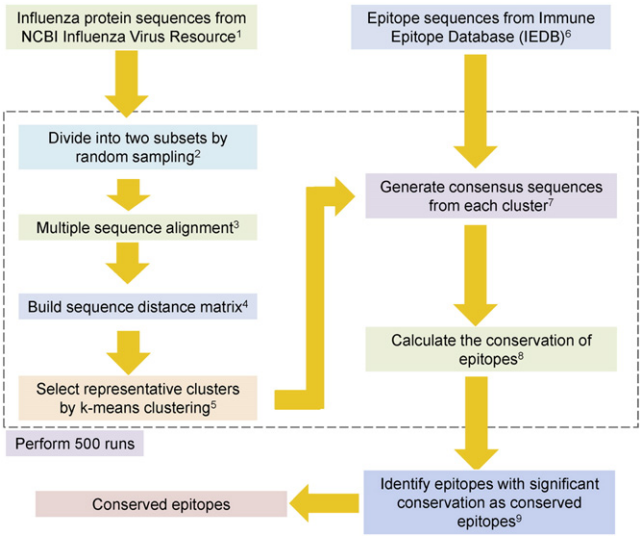

Fig. 5. The workflow for identifying conserved epitopes in influenza A/H3N2 viruses. [1]: The 11 non-redundant full-length protein segments of human influenza A/H3N2 viruses were retrieved. [2]: The process involves 500 runs of random sampling with a five-fold cross-validation to obtain two subsets in each process. [3]: Sequence alignment was performed by the MUSCLE software. [4]: Sequence distances were calculated by pairwise comparison of the sequence differences with Poisson correction model. [5]: The cluster size was obtained by optimum average silhouette width. [6]: The human-related T-cell and B-cell linear epitope sequences with positive responses from the IEDB were collected. [7]: The consensus sequence, or the representative sequence, in each cluster was determined by a simple majority rule consensus method. [8]: The percentage of epitope conservation was calculated by dividing the number of epitopes that are identical to consensus sequences by the total number of consensus sequences and then multiplying by 100. [9]: *T*-test was used to test the conservations of two subsets and to determine whether conservations of subsets ≥99%.

5.2. Data collection of influenza A/H3N2 protein sequences and epitope sequences

The available 11 protein sequences of human influenza A/H3N2 viruses, namely PB2, PB1, PB1-F2, PA, HA, NP, NA, M1, M2, NS1 and NS2, were downloaded from the National Center for Biotechnology Information (NCBI) Influenza Virus Resource [17] on March, 2011. Only the full-length protein sequences were selected and identical sequences were removed.

We selected influenza B- and T-cell epitope sequences from Immune Epitope Database (IEDB) [41]. These epitopes were able to induce immunity in experimental tests (query options in IEDB: –linear sequences of epitopes: exact matches, –source organism of epitopes: Influenza A virus, –host organism of immunization: human; –qualitative measurements of B- and T-cell assay: "positive," "positive-low," "positive-intermediate," or "positive-high").

To understand availability of conserved epitopes in the population coverage, information of HLA restriction of conserved epitopes was collected from the annotation of conserved epitopes in IEDB.

5.3. Alignment, sequence distances and clustering analysis of influenza A/H3N2 virus sequences

First, pairwise distances of the protein sequences were calculated using MEGA4 software [42] (using the Poisson correction model to measure sequence differences as sequence distances). The MUSCLE software [43] was used to align multiple sequences (using the following options: -distance1 = kbit20_3, -maxiters = 1, -diags, -sv, -stable). Sequence distances were normalized by dividing the number of differences by the length of the aligned sequences in an alignment.

Next, k-means clustering was performed on the sequence distance matrix of each viral protein to determine the viral clusters. K-means clustering is a partitioning method that will partition n input values/observations into k clusters. In order to determine the appropriate number of clusters, the function pamk() in the fpc package in R software was used to obtain the suggested number of clusters based on optimum average silhouette width.

Finally, consensus sequences as the representative sequences in each cluster were determined by the majority vote rule. We then calculated the sequence conservations of epitopes in these representative sequences.

5.4. Sequence analysis of conservation of epitopes in influenza A/H3N2 viruses

When performing epitope conservation calculation, we used the following formula to depict the conservation of the given epitope:

$$\frac{\text{Number of epitopes that are identical to consensus sequences}}{\text{Number of consensus sequences}} \times 100$$

5.5. Statistical evaluation of conserved epitopes by cross-validation with random sampling

Statistical tests were performed to evaluate the significance of epitope conservations in training sets and testing sets which were selected by using 5-fold cross-validation with random sampling. 80% of total influenza sequences were randomly assigned into training sets and the

remaining 20% were put into testing sets. The shuffle function from the PERL package List::Util was used for dividing the sequences into the training and testing sets.

The "conserved epitope" must meet each of the following requirements. The paired sample *t*-test was used (1) to test whether the mean of conservation in training set is less than and equal to the conservation in testing set ($p<0.05$) and (2) to determine whether mean of conservation in the training set is greater than 99% ($p<0.05$).

5.6. Calculation of percentages of influenza A/H3N2 viruses containing conserved epitopes during the 1968–2010 period

To understand whether conserved epitopes had existed in wild-type H3N2 strains, the percentages of the presence of conserved epitopes were estimated using the following formula:

$$\frac{\text{Number of viral sequences containing identical epitope sequences in a given year}}{\text{Number of total viral sequences in a given year}} \times 100$$

5.7. Evaluation of sequence identities of conserved epitopes among vaccine strains

To evaluate whether the sequences of conserved epitope match to sequences of various influenza strains, sequence identities of the conserved epitopes among influenza vaccine strains were calculated by using the BLAST sequence comparison tool [44]. The conserved epitopes identified in our results were used as query sequences and sequences of vaccine strains were available in NCBI (options of filters: influenza A/H3N2 viruses and only vaccine strains) as subject sequences. The length of the alignment must cover the total length of the given epitope.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2012.06.003.

Acknowledgments

The authors thank Dr. Chuan-Liang Kao (National Taiwan University), Dr. Yi-Ming Arthur Chen (National Yang-Ming University) and Dr. Yu-Jiun Chan (Taipei Veterans General Hospital) for providing constructive comments and suggestions that improved this paper. This study was supported by the National Science Council, Taiwan (NSC 100-2319-B-010-002) and the Ministry of Education, Taiwan, Aim for the Top University Plan.

Competing interests

The authors declare that they have no competing interests.

Author contributions

KW CH conceived and designed the study. KW CY CH performed the experiments. KW CY analyzed the data. KW CY SW CC CH contributed reagents/materials/analysis tools. KW wrote the original manuscript. SW and CH contributed to revisions of the manuscript. All authors read and approved the final version.

References

[1] E. Domingo, J.J. Holland, RNA virus mutations and fitness for survival, Annu. Rev. Microbiol. 51 (1997) 151–178.
[2] D.J. Smith, A.S. Lapedes, J.C. de Jong, T.M. Bestebroer, G.F. Rimmelzwaan, A.D. Osterhaus, R.A. Fouchier, Mapping the antigenic and genetic evolution of influenza virus, Science 305 (2004) 371–376.
[3] F. Carrat, A. Flahault, Influenza vaccine: the challenge of antigenic drift, Vaccine 25 (2007) 6852–6862.
[4] E.D. Kilbourne, Influenza pandemics of the 20th century, Emerg. Infect. Dis. 12 (2006) 9–14.
[5] R. Fang, W. Min Jou, D. Huylebroeck, R. Devos, W. Fiers, Complete structure of A/duck/Ukraine/63 influenza hemagglutinin gene: animal virus as progenitor of human H3 Hong Kong 1968 influenza hemagglutinin, Cell 25 (1981) 315–323.
[6] C. Scholtissek, W. Rohde, V. Von Hoyningen, R. Rott, On the origin of the human influenza virus subtypes H2N2 and H3N2, Virology 87 (1978) 13–20.
[7] J.K. Taubenberger, J.V. Hultin, D.M. Morens, Discovery and characterization of the 1918 pandemic influenza virus in historical context, Antivir. Ther. 12 (2007) 581–591.
[8] A.T. Heiny, O. Miotto, K.N. Srinivasan, A.M. Khan, G.L. Zhang, V. Brusic, T.W. Tan, J.T. August, Evolutionarily conserved protein sequences of influenza a viruses, avian and human, as vaccine targets, PLoS One 2 (2007) e1190.
[9] P.T. Tan, A.T. Heiny, O. Miotto, J. Salmon, E.T. Marques, F. Lemonnier, J.T. August, Conservation and diversity of influenza A H1N1 HLA-restricted T cell epitope candidates for epitope-based vaccines, PLoS One 5 (2010) e8754.
[10] T.T. Wang, G.S. Tan, R. Hai, N. Pica, L. Ngai, D.C. Ekiert, I.A. Wilson, A. Garcia-Sastre, T.M. Moran, P. Palese, Vaccination with a synthetic peptide from the influenza virus hemagglutinin provides protection against distinct viral subtypes, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 18979–18984.
[11] S. Neirynck, T. Deroo, X. Saelens, P. Vanlandschoot, W.M. Jou, W. Fiers, A universal influenza A vaccine based on the extracellular domain of the M2 protein, Nat. Med. 5 (1999) 1157–1163.
[12] D. Corti, J. Voss, S.J. Gamblin, G. Codoni, A. Macagno, D. Jarrossay, S.G. Vachieri, D. Pinna, A. Minola, F. Vanzetta, C. Silacci, B.M. Fernandez-Rodriguez, G. Agatic, S. Bianchi, I. Giacchetto-Sasselli, L. Calder, F. Sallusto, P. Collins, L.F. Haire, N. Temperton, J.P. Langedijk, J.J. Skehel, A. Lanzavecchia, A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins, Science 333 (2011) 850–856.
[13] J. Sui, W.C. Hwang, S. Perez, G. Wei, D. Aird, L.M. Chen, E. Santelli, B. Stec, G. Cadwell, M. Ali, H. Wan, A. Murakami, A. Yammanuru, T. Han, N.J. Cox, L.A. Bankston, R.O. Donis, R.C. Liddington, W.A. Marasco, Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses, Nat. Struct. Mol. Biol. 16 (2009) 265–273.
[14] M. ElHefnawi, O. Alaidi, N. Mohamed, M. Kamar, I. El-Azab, S. Zada, R. Siam, Identification of novel conserved functional motifs across most influenza A viral strains, Virol. J. 8 (2011) 44.
[15] D.M. Gendoo, M.M. El-Hefnawi, M. Werner, R. Siam, Correlating novel variable and conserved motifs in the hemagglutinin protein with significant biological functions, Virol. J. 5 (2008) 91.
[16] R.M. Bush, W.M. Fitch, C.A. Bender, N.J. Cox, Positive selection on the H3 hemagglutinin gene of human influenza virus A, Mol. Biol. Evol. 16 (1999) 1457–1465.
[17] Y. Bao, P. Bolotov, D. Dernovoy, B. Kiryutin, L. Zaslavsky, T. Tatusova, J. Ostell, D. Lipman, The influenza virus resource at the National Center for Biotechnology Information, J. Virol. 82 (2008) 596–601.
[18] D.C. Ekiert, R.H. Friesen, G. Bhabha, T. Kwaks, M. Jongeneelen, W. Yu, C. Ophorst, F. Cox, H.J. Korse, B. Brandenburg, R. Vogels, J.P. Brakenhoff, R. Kompier, M.H. Koldijk, L.A. Cornelissen, L.L. Poon, M. Peiris, W. Koudstaal, I.A. Wilson, J. Goudsmit, A highly conserved neutralizing epitope on group 2 influenza A viruses, Science 333 (2011) 843–850.
[19] J. Chen, J.J. Skehel, D.C. Wiley, N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 8967–8972.
[20] A. Horvath, G.K. Toth, P. Gogolak, Z. Nagy, I. Kurucz, I. Pecht, E. Rajnavolgyi, A hemagglutinin-based multipeptide construct elicits enhanced protective immune response in mice against influenza A virus infection, Immunol. Lett. 60 (1998) 127–136.
[21] E. Vareckova, V. Mucha, S.A. Wharton, F. Kostolansky, Inhibition of fusion activity of influenza A haemagglutinin mediated by HA2-specific monoclonal antibodies, Arch. Virol. 148 (2003) 469–486.
[22] T. Yano, E. Nobusawa, A. Nagy, S. Nakajima, K. Nakajima, Effects of single-point amino acid substitutions on the structure and function neuraminidase proteins in influenza A virus, Microbiol. Immunol. 52 (2008) 216–223.
[23] B.E. Johansson, B. Grajower, E.D. Kilbourne, Infection-permissive immunization with influenza virus neuraminidase prevents weight loss in infected mice, Vaccine 11 (1993) 1037–1039.
[24] S.L. Epstein, W.P. Kong, J.A. Misplon, C.Y. Lo, T.M. Tumpey, L. Xu, G.J. Nabel, Protection against multiple influenza A subtypes by vaccination with highly conserved nucleoprotein, Vaccine 23 (2005) 5404–5410.
[25] J.B. Ulmer, T.M. Fu, R.R. Deck, A. Friedman, L. Guan, C. DeWitt, X. Liu, S. Wang, M.A. Liu, J.J. Donnelly, M.J. Caulfield, Protective CD4+ and CD8+ T cells against influenza virus induced by vaccination with nucleoprotein DNA, J. Virol. 72 (1998) 5648–5653.
[26] A. Patel, K. Tran, M. Gray, Y. Li, Z. Ao, X. Yao, D. Kobasa, G.P. Kobinger, Evaluation of conserved and variable influenza antigens for immunization against different isolates of H5N1 viruses, Vaccine 27 (2009) 3083–3089.
[27] K. Das, J.M. Aramini, L.C. Ma, R.M. Krug, E. Arnold, Structures of influenza A proteins and insights into antiviral drug targets, Nat. Struct. Mol. Biol. 17 (2010) 530–538.
[28] E. Assarsson, H.H. Bui, J. Sidney, Q. Zhang, J. Glenn, C. Oseroff, I.N. Mbawuike, J. Alexander, M.J. Newman, H. Grey, A. Sette, Immunomic analysis of the repertoire of T-cell specificities for influenza A virus in humans, J. Virol. 82 (2008) 12241–12251.
[29] W. Gerhard, K. Mozdzanowska, M. Furchner, G. Washko, K. Maiese, Role of the B-cell response in recovery of mice from primary influenza virus infection, Immunol. Rev. 159 (1997) 95–103.
[30] R.J. Webby, S. Andreansky, J. Stambas, J.E. Rehg, R.G. Webster, P.C. Doherty, S.J. Turner, Protection and compensation in the influenza virus-specific CD8+ T cell response, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 7235–7240.
[31] J.E. McElhaney, D. Xie, W.D. Hager, M.B. Barry, Y. Wang, A. Kleppinger, C. Ewen, K.P. Kane, R.C. Bleackley, T cell responses are better correlates of vaccine protection in the elderly, J. Immunol. 176 (2006) 6333–6339.

[32] J.A. Greenbaum, M.F. Kotturi, Y. Kim, C. Oseroff, K. Vaughan, N. Salimi, R. Vita, J. Ponomarenko, R.H. Scheuermann, A. Sette, B. Peters, Pre-existing immunity against swine-origin H1N1 influenza viruses in the general human population, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 20365–20370.
[33] S.M. Tompkins, Z.S. Zhao, C.Y. Lo, J.A. Misplon, T. Liu, Z. Ye, R.J. Hogan, Z. Wu, K.A. Benton, T.M. Tumpey, S.L. Epstein, Matrix protein 2 vaccination and protection against influenza viruses, including subtype H5N1, Emerg. Infect. Dis. 13 (2007) 426–435.
[34] A. Rambaut, O.G. Pybus, M.I. Nelson, C. Viboud, J.K. Taubenberger, E.C. Holmes, The genomic and epidemiological dynamics of human influenza A virus, Nature 453 (2008) 615–619.
[35] N. Creanza, J.S. Schwarz, J.E. Cohen, Intraseasonal dynamics and dominant sequences in H3N2 influenza, PLoS One 5 (2010) e8544.
[36] Z. Stanekova, E. Vareckova, Conserved epitopes of influenza A virus inducing protective immunity and their prospects for universal vaccine development, Virol. J. 7 (2010) 351.
[37] A. Sette, M. Newman, B. Livingston, D. McKinney, J. Sidney, G. Ishioka, S. Tangri, J. Alexander, J. Fikes, R. Chesnut, Optimizing vaccine design for cellular processing, MHC binding and TCR recognition, Tissue Antigens 59 (2002) 443–451.
[38] J. Sidney, H.M. Grey, R.T. Kubo, A. Sette, Practical, biochemical and evolutionary implications of the discovery of HLA class I supermotifs, Immunol. Today 17 (1996) 261–266.
[39] C. Gianfrani, C. Oseroff, J. Sidney, R.W. Chesnut, A. Sette, Human memory CTL response specific for influenza A virus is broad and multispecific, Hum. Immunol. 61 (2000) 438–452.
[40] H.H. Bui, B. Peters, E. Assarsson, I. Mbawuike, A. Sette, Ab and T cell epitopes of influenza A virus, knowledge and opportunities, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 246–251.
[41] Q. Zhang, P. Wang, Y. Kim, P. Haste-Andersen, J. Beaver, P.E. Bourne, H.H. Bui, S. Buus, S. Frankild, J. Greenbaum, O. Lund, C. Lundegaard, M. Nielsen, J. Ponomarenko, A. Sette, Z. Zhu, B. Peters, Immune epitope database analysis resource (IEDB-AR), Nucleic Acids Res. 36 (2008) W513–518.
[42] K. Tamura, J. Dudley, M. Nei, S. Kumar, MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0, Mol. Biol. Evol. 24 (2007) 1596–1599.
[43] R.C. Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity, BMC Bioinforma. 5 (2004) 113.
[44] D.W. Mount, Using the Basic Local Alignment Search Tool (BLAST), Cold Spring Harbor Protocols (2007), http://dx.doi.org/10.1101/pdb.top17, http://cshprotocols.cshlp.org/content/2007/7/pdb.top17.long.