Ab and T cell epitopes of influenza A virus, knowledge and opportunities

Huynh-Hoa Bui*, Bjoern Peters*, Erika Assarsson*, Innocent Mbawuike†, and Alessandro Sette*‡

*Division of Vaccine Discovery, La Jolla Institute for Allergy and Immunology, 9420 Athena Circle, La Jolla, CA 92037; and †Department of Molecular Virology, Baylor College of Medicine, Houston, TX 77030

Communicated by Howard M. Grey, La Jolla Institute for Allergy and Immunology, San Diego, CA, November 1, 2006 (received for review October 1, 2006)

The Immune Epitope Database and Analysis Resources (IEDB) (www.immuneepitope.org) was recently developed to capture epitope related data. IEDB also hosts various bioinformatics tools that can be used to identify novel epitopes as well as to analyze and visualize existing epitope data. Herein, a comprehensive analysis was undertaken (*i*) to compile and inventory existing knowledge regarding influenza A epitopes and (*ii*) to determine possible cross-reactivities of identified epitopes among avian H5N1 and human influenza strains. At present, IEDB contains >600 different epitopes derived from 58 different strains and 10 influenza A proteins. By using the IEDB analysis resources, conservancy analyses were performed, and several conserved and possibly cross-reactive epitopes were identified. Significant gaps in the current knowledge were also revealed, including paucity of Ab epitopes in comparison with T cell epitopes, limited number of epitopes reported for avian influenza strains/subtypes, and limited number of epitopes reported from proteins other than hemagglutinin and nucleoprotein. This analysis provides a resource for researchers to access existing influenza epitope data. At the same time, the analysis illustrates gaps in our collective knowledge that should inspire directions for further study of immunity against the influenza A virus.

B lymphocytes | T lymphocytes | conservancy | pandemic cross-reactivity

Influenza A viruses are widely distributed in nature and can infect a variety of birds and mammals. Their genomes consists of eight single-stranded RNA segments that code for 10 different proteins, one nucleoprotein (NP), three polymerase proteins (PA, PB1, and PB2), two matrix proteins (M1 and M2), two nonstructural proteins (NS1 and NS2), and two external glycoproteins [hemagglutinin (HA) and neuraminidase (NA)]. The viruses are classified on the basis of differences in the antigenic structure of HA and NA proteins, with their different combinations representing unique virus subtypes that are further classified into specific strains. Although all known subtypes can be found in birds, currently circulating human influenza A subtypes are H1N1 and H3N2, with intermittent circulation of H1N2 reassortants. Seasonal outbreaks are caused by subtypes already circulating among people, whereas pandemics are caused by either an emerging novel subtype derived by reassortment with avian viruses (1957 A/H2N2 pandemic and 1968 A/H3N2 pandemic), or all-avian (1918 A/H1N1 pandemic). It is possible that some of these pandemics are "recycled" subtypes that had not circulated in human populations for many years. This was, for example, the case for the 1977 "pseudopandemic" of A/H1N1 viruses that resurfaced after 20 years of absence after the 1957 A/H2N2 pandemic.

"Avian" influenza refers to subtypes found chiefly in birds, but infections with these viruses can also occur in humans. Confirmed cases of human disease caused by several subtypes of avian influenza, including H7N7, H9N2, and other emerging avian viruses such as low-pathogenic H5N1 and H5N2, have been reported since 1997 (1–5). However, of the few avian influenza viruses that have crossed the species barrier to infect humans, the emerging high-pathogenic H5N1 virus in Asia has caused the largest number of detected cases of severe disease and death in humans (6). Because these viruses do not commonly infect humans, little or no immunity may be present in the general human population (with the exception of potential cross-reactive immunity originating from exposure to the other strains commonly infecting humans). Therefore, if the high-pathogenic H5N1 virus were to gain the capacity to spread easily from person to person, an influenza pandemic could ensue (7–10).

Results and Discussion

Why Analyze Influenza A-Derived Epitopes? Because of recent events, there has been resurgent interest in the study of influenza A virus in general and avian influenza H5N1 in particular. Further studies must be completed, ranging from basic studies of immune responses and interactions of influenza virus with its hosts, to the evaluation of new vaccine candidates (11–14). Epitopes can be used to accurately monitor immune responses as well as to tease out which influenza responses are specific for a given virus strain or subtype or are cross-reactive with several or most strains.

Immune responses to influenza A virus have been studied for decades, not only as a model system, but also because of their medical importance. However, the vast amount of resulting epitope information available in the literature has not been globally analyzed and made accessible to the scientific community. Herein, we perform such an analysis (*i*) to compile and inventory existing knowledge regarding influenza A epitopes and (*ii*) to determine possible cross-reactivities of identified epitopes among avian H5N1 and human influenza strains. The data source and results of our analysis are available in the Immune Epitope Database and Analysis Resources (IEDB), which was recently developed to capture epitope related data and is publicly available at www.immuneepitope.org (15, 16). Besides the efforts of compiling and making comprehensive epitope information available to the public domain (17), the IEDB also hosts various bioinformatics tools to analyze epitope data (including, for example, population coverage (18) and epitope conservancy) as well as tools to predict epitope cellular processing (19), binding to MHC (20–22), and recognition by T cell receptors and Ab molecules. In the context of this analysis, the conservancy tool provided by the IEDB was used to identify conserved epitopes that might be cross-reactive among avian H5N1 and human influenza strains. Finally, as an outcome of this analysis, important gaps in the global knowledge relating to

Author contributions: A.S. designed research; H.-H.B. and B.P. contributed new reagents/analytic tools; H.-H.B., B.P., E.A., and I.M. analyzed data; and H.-H.B. and E.A. wrote the paper.

The authors declare no conflict of interest.

Abbreviations: HA, hemagglutinin; IEDB, Immune Epitope Database and Analysis Resource; NA, neuraminidase; NP, nucleoprotein.

‡To whom correspondence should be addressed. E-mail: alex@liai.org.

This article contains supporting information online at www.pnas.org/cgi/content/full/0609330104/DC1.

immunity directed against the influenza A virus were also identified, pointing a way forward in immune epitope research.

Ab and T cell epitopes are defined as the molecular structures interacting with Abs and T cell receptor (TCR) molecules, respectively (23). In our analysis of the existing scientific literature relating to influenza A derived epitopes, we considered only epitopes shown to be recognized by Abs or TCR in the context of the whole influenza virus or proteins. We excluded epitopes that were defined solely by their use as immunogens (to induce the responses) and as antigens (to measure the response), because it is not possible to evaluate the relevance of such data with respect to antiviral immune responses.

Historically, a variety of different assays (ranging from T cell proliferation, cytokine production, and ELISAs, to neutralization and protection from live virus challenge) have been used to evaluate the recognition of influenza epitopes. Challenge with live virus and neutralization assays are used to define protective Ab and T cell epitopes. We make no attempt to enforce a common set of criteria for defining immunogenicity and protective efficacy, because widely divergent methodologies were used by different laboratories to measure immune responses. Rather, we record, for each epitope the specific assay category and conditions used, and conform to the criteria for defining positive and negative measurements as reported by the authors themselves in each published article.

We believe that the definition of the structural and functional determinants of influenza-derived epitopes could be useful in detecting and monitoring infections as well as being crucial to project potential cross-reactive immunity and efficacy against new strains by existing vaccines and diagnostics (24, 25), because once the structure of an epitope is known, databases of influenza genomic information such as Influenza Sequence Database (26), Influenza Virus Resources (27) and BioHealthBase (28) can be searched to project whether the same structure is also conserved in all or most influenza strains, or is specific to a particular influenza strain or subtype. The Influenza Sequence Database (www.flu.lanl.gov) contains all published influenza viral sequences that have been curated by domain experts to ensure high standards of accuracy and completeness (26). The Influenza Virus Resource (www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html) presents data obtained from the NIAID Influenza Genome Sequencing Project as well as from GenBank, combined with tools for flu sequence analysis and annotation (27). Finally, the BioHealthBase system (www.biohealthbase.org) focuses on six priority pathogens, including influenza, to help fill in gaps in genomic and other data critical to scientific researchers (28).

In a diagnostic and disease-monitoring setting, epitopes that are specific to a given strain or subtype can be used to monitor responses to that particular strain or subset, removing the confounding influence of immune responses derived from previous exposures to partially cross-reactive strains or subtypes (11, 29). One of the shortcomings of the currently available influenza vaccines is the induction of a strain-specific immunity, which requires a new vaccine to be produced each year and for each different strain. In this context, if conserved epitopes can be defined, different immunization regimens and vaccine candidates could be evaluated for their capacity to induce immune responses to those specific conserved determinants.

Conversely, samples from individuals vaccinated and/or naturally infected with viral strains commonly infectious for humans, such as H1N1 and H3N2, could be screened for the presence of cross-reactive immunity. Such cross-reactive immune recognition may represent a minor component of the total response, but its precise mapping would nevertheless be of significant interest. Several groups have analyzed the potential for cross-reactive epitopes, both at the Ab level (between different types of N1) and in the highly conserved internal gene segments. This work constitutes the basis for the recent suggestion that one of the potential strategies to develop universal influenza vaccines relies on the identification of protective and cross-reactive antibodies, followed by the mapping of the epitopes recognized by such antibodies (30).

How Many Influenza A Epitopes Have Been Reported in the Literature?

As mentioned above, immune responses against influenza A virus have been intensely characterized over the course of several decades. However, this knowledge is dispersed over a large number of scientific references, and a simple search in PubMed using the keywords "epitope" and "influenza" reveals >2,000 different scientific reports. It is unclear how many of these reports contain data relating to new epitopes or new information relating to old ones. Furthermore there is no simple way to extract from these references answers to simple questions. For example, how many epitopes are known from strain "X"?; In which host have they been characterized?; Which epitopes are unique, and which are conserved in other strains, and so on. To address these issues, we perform a comprehensive analysis of all epitope data relating to influenza A virus. The analysis consists of two separate tasks: (*i*) data-compilation efforts that involve identification and curation of influenza A epitope literature into IEDB and (*ii*) data-analysis efforts that involve the use of the IEDB-provided conservancy tool to analyze and identify epitopes that are conserved among various avian H5N1 and human influenza strains.

As a first task of the analysis, the current state of knowledge of influenza A-derived Ab and T cell epitopes was determined (Fig. 1). To accomplish this task, a query [see supporting information (SI) Fig. 2] was constructed to identify potentially relevant influenza epitope-related articles from the entirety of published literature available in PubMed. As of May 22, 2006, the PubMed contained >16 million references, of which 2,063 were identified as influenza epitope-related. Running a similar query without any specific constraints on the source of the epitope yielded ≈100,000 references. Thus, a significant fraction (≈2%) of the worldwide epitope literature is related to the flu virus, likely reflecting the extended period that this pathogen has been studied, its biomedical importance, and its use as a model for basic studies in virology, immunology, and vaccinology. By comparison, a similar search in the case of HIV (AIDS) yielded 4,442 references (4.4% of the total). In the case of lymphocytic choriomeningitis virus (LCMV), *Mycobacterium tuberculosis* (tuberculosis) and *Plasmodium* (malaria), the corresponding figures were 472, 856, and 1,397 (0.5%, 0.9%, and 1.4%), respectively. After manual inspection of all abstracts and full-text review of potentially relevant influenza A epitope articles, a total of 429 references were curated in detail (17). Of these references, 103 contained Ab epitope information. In addition, a total of 114, 13, and 291 references, respectively, contained data relating to MHC binding, elution of MHC ligands, and T cell assays.

To determine how many influenza A epitopes have been described in the literature, a query was performed to search data contained in the IEDB. A total of 412 T cell epitopes (175 CD4, 148 CD8, and 89 undefined) and 190 Ab epitopes (75 linear and 115 conformational) were retrieved. These data provide an indication of the wealth of information already available in the scientific literature relating to influenza A epitopes and should constitute a useful resource for researchers worldwide. Given the well-established importance of Ab responses in vaccine efficacy and in prevention of influenza infection, the relatively small number of published Ab epitopes is unexpected. Although the structure and technological means for identifying Ab and T cell epitopes are radically different, given the fact that Ab titers are the only accepted correlate of protection from influenza and of vaccine efficacy, the paucity of Ab epitopes in comparison with T cell epitopes is indeed surprising. The >2:1 ratio of T cell vs. Ab influenza epitopes is likely because of the fact that Ab

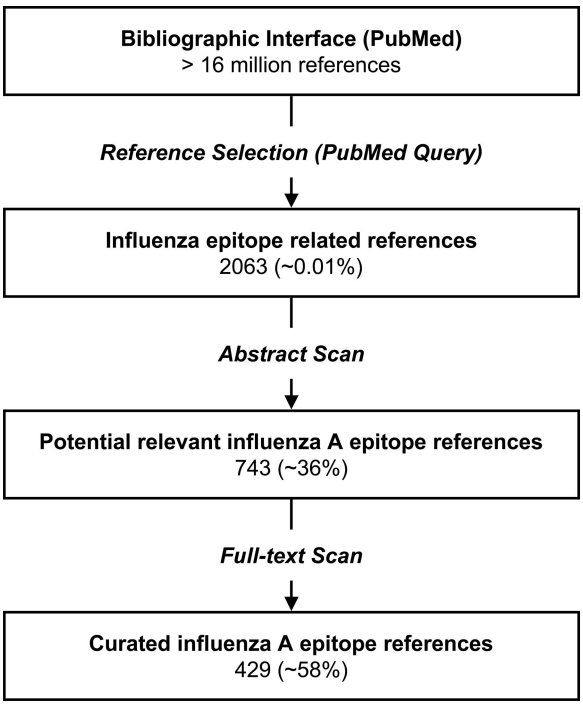

Fig. 1. Process for selection and curation of relevant influenza A epitope literature references.

epitopes are inherently more difficult to characterize than T cell epitopes.

Of 190 identified Ab epitopes, ≈40% are linear sequences. The knowledge of epitope 3D structure can offer important insights into understanding virus neutralization, predicting epitope conservancy across different strains, and rationally designing new vaccine candidates. However, we note that the 3D structures of only 22 epitope/receptor complexes, which represent an average of 4% of all reported epitopes, were determined (an additional 12 epitope/MHC structures have also been described).

The issue of which strain of influenza A was used to define the various epitopes is of obvious importance, in light of the potential use of the epitopes to monitor immune responses to influenza vaccination and infection. Knowledge relating to a diverse set of strains is also desirable to ensure a general biological and immunological relevancy of the results. A lesson learned from HIV research is that excessive reliance on long-term maintained laboratory strains can lead to difficulties in extrapolating results to fresh patient isolates. Our influenza A analysis identifies epitopes from 13 different subtypes and 58 different strains (SI Table 4). The vast majority are from the human influenza H1N1 and H3N2 subtypes, and a relatively large proportion of these epitopes are derived from prototype strains used for model studies, such as A/Puerto Rico/8/34(H1N1) (≈24%) and A/X-31(H3N2) (≈32%), with fewer epitopes having been characterized from fresh isolates of human pathogenic strains (≈1.2%, on average, for a given strain). Only two epitopes from the H5N1 avian influenza A/Viet Nam/1194/2004 are included in this database. These results suggest that more studies need to be focused on the identification of epitopes from the strains responsible for human infections and also point to the urgent need to identify epitopes recognized by responses directed against avian influenza strains. It is, of course, not surprising that the number of epitopes that have been described in either humans or animal models for avian influenza infections are going to be comparatively few compared with the circulating human strains. Some of the original work defining the very

Table 1. Total number of published influenza A epitopes by protein

Protein	Antibody	T cell		Total
		CD4	CD8	
HA	150	113	35	298
NP	3	44	49	96
PA	0	1	11	12
NA	24	7	8	39
M1	4	9	15	28
PB2	0	0	9	9
M2	9	0	3	12
PB1	0	0	10	10
NS1	0	1	7	8
NS2	0	0	1	1

nature of Ab and T cell epitopes used influenza as a model and, as such, these data have been generated for >30 years. The emergence of the avian strains in 1997 (and their reemergence in 2003) has provided far less time for their study; in addition, the increased pathogenicity of fresh isolates has led to their being classified as select agents, making immunological analysis more difficult because of the special containment facilities required. Our analysis demonstrates and underlines this fundamental weakness and gap in our collective knowledge.

Another issue of obvious relevance is the distribution of epitopes by the source proteins from which they are derived (Table 1). It is generally anticipated that Ab responses to vaccination or infection are directed mostly toward epitopes from viral surface-exposed proteins, whereas epitopes recognized by cellular immunity may be broadly derived from both internal and surface proteins. Because internal proteins are far more conserved among different influenza strains and thereby potentially offer the best choice for vaccines aimed at eliciting the broadest possible strain coverage, knowledge of the source proteins from which the epitopes are derived is particularly relevant. Ab epitopes have been identified from only 5 of the 10 viral proteins, and the majority are derived from the virus surface proteins HA, NA, and M2. Compared with HA, fewer Ab epitopes were derived from NA and M2 proteins. T cell epitopes have been identified from all 10 influenza proteins; the highest number of epitopes being derived from HA and NP. Indeed, most published CD4 T cell epitopes are derived from the HA protein, whereas most CD8 T cell epitopes are derived from the NP protein. It should be emphasized that this analysis cannot determine whether the uneven distribution of epitopes as a function of the protein of origin is reflective of poor immunogenicity of those proteins in certain contexts or, perhaps more likely, reflects a bias in the number of studies addressing the immunogenicity of different proteins.

The host species in which the epitopes are identified is shown in Table 2. The majority of Ab and T cell epitopes were identified in mouse, human, or rabbit hosts. Few epitopes are described in birds, which are relevant hosts to study virus evolution. Studies using ferrets, a commonly used experimental model, and non-human primates are also underrepresented. Furthermore, rather astonishingly, only one Ab epitope, compared with 160 T cell epitopes, has been identified by using human samples. Compared with other animal hosts, such as rodents, relatively few human host data available in the literature is probably a reflection of the inherent complexity in characterizing and interpreting immune epitope data from human models, because the repertoire of epitopes recognized in rodents and rabbits is almost invariably measured after a single exposure to influenza. By contrast, in adult humans, the immune response is the final product of a long series of repeated exposures to different viral influenza strains.

Table 2. Total number of published influenza A epitopes by host species

Protein	Ab	T cell	Total
Mouse	71	290	361
Rabbitt	35	0	35
Chicken	3	0	3
Human	1	160	161
Ferret	1	0	1
Goat	1	0	1
Rhesus monkey	1	0	1
Cotton-top tamarin	0	2	2

Responses induced in humans by previous influenza infections or vaccinations might significantly skew the repertoire of epitopes recognized upon infection or vaccination with a different influenza strain, a phenomenon termed "original antigenic sin" (31). This situation highlights the need for more studies defining the Ab epitopes recognized in humans, and the degree to which they overlap with those recognized in animal model systems.

Conservancy of Ab and T Cell Influenza A Epitopes. For the second part of the analysis, we are interested in evaluating conservancy of epitopes among various influenza strains in general and with H5N1 in particular, using the conservancy tool provided by the IEDB. As mentioned above, identification of conserved epitopes is of interest in terms of the prospect for development of broader-spectrum influenza vaccines. Conservancy analysis could also identify epitopes detected from previously vaccinated or infected individuals and associated with cross-reactivity and potential protection from the avian H5N1 strains. Conversely, epitopes that are specific for a given subtype can be used for monitoring responses, removing the confounding influence of immune responses derived from previous exposures to partially cross-reactive strains or subtypes.

To analyze epitope conservancy, we first assembled a collection of representative human and avian H5N1 influenza strains for the analysis, because inclusion of all available sequences would generate a biased conservancy picture reflective of relative abundance of available sequences from a given strain or subtype. For the human influenza strains, our strategy was to select viral strains that had been used for vaccination or were known to cause infection in the human population. A total of 17 influenza strains including 7 H1N1, 8 H3N2, and 2 H5N1 strains were selected for the analysis. Of these, 5 H1N1 and 6 H3N2 strains had been used in annual influenza vaccinations from 1968 to 2004 (SI Table 5). In addition, other pathogenic H1N1 and H3N2 human influenza strains of potential interest, such as A/Brevig Mission/1/18, which circulated in the 1918 pandemic, were also included. The two H5N1 strains that circulated in the 1997 and 2003–2004 H5N1 outbreaks, respectively, were also selected.

Next, using the epitope conservancy analysis tool provided in the analysis resources of the IEDB, we find that, overall, T cell epitopes are more conserved than Ab epitopes (SI Tables 6–8). For T cell epitopes, ≈50% and 30% are conserved at 80% and 90% identity levels, respectively, in both human (H1N1 and H3N2) and avian (H5N1) strains (SI Table 8). At the 100% identity level, 15.0% of T cell epitopes are conserved in the human strains, and 11.4% are also conserved in the avian H5N1 strains. In contrast, only 2.7% of Ab epitopes are conserved at 100% identity level, and <11% were conserved at 80% identity level. A possible reason for this difference is that ≈80% of the linear Ab epitopes, compared with only 40% of the T cell epitopes, are derived from the two most variable influenza proteins, HA and NA. In general, the results suggest that significant levels of interstrain cross-reactivity are likely for T cell epitopes, but much less so for Ab epitopes. Several highly conserved discontinuous conformational Ab epitopes are also identified (SI Table 9). However, their degree of conservation should be interpreted with caution, because pattern-wise conserved discontinuous sequences may not be cross-reactive because of the influence of unknown neighboring and interdispersed amino acids on protein 3D structures. Finally, it should be emphasized that the fact that an epitope is conserved does not necessarily imply that it is also cross-protective.

In this analysis we have organized the data around the subtypes in which the epitopes are found (e.g., H3N2 and H1N1). This is relevant for Ab epitopes for obvious reasons. But it is also relevant for T cell epitopes because, even if an epitope sequence is conserved in different subtypes, flanking regions and differences in the viral genome might affect whether the epitope is recognized as dominant in the context of a different subtype. Furthermore, our analysis will help to determine whether or not a given epitope could be used as a marker for a given subtype. In this context, whereas responses to conserved epitopes might be most useful with respect to vaccine development, subtype-specific epitopes might be most useful for diagnostic purposes and the study of viral evolution.

It should be noted here that the main purpose of the current study is to provide a resource analyzing and making accessible influenza information with potential implications in terms of future research in areas relevant to vaccine research, understanding the role of T cell immunity in influenza, and to highlight multiple pandemic influenza issues surrounding H1N1 and H5N1 viruses in particular. Our analysis is purely bioinformatics and can address only experiments that have been performed and published in peer-reviewed journals. However, the analysis also suggests possible experiments that could be conducted to further validate the epitopes and improve our understanding of the immune response to influenza or ability to combat influenza. For example, several linear and MAb-defined epitopes (SI Table 6) were shown to be highly conserved within H1 or H3 subtypes. Data like these maybe useful for identification of new vaccine targets, and experiments to demonstrate that one of these MAbs indeed neutralized virus *in vitro* or provided passive protection *in vivo* would be neither time-consuming nor technically difficult. Similarly, for the MAb-defined cross-reactive conformational epitopes (SI Table 9), it should be possible to test one or several of these MAbs for cross-reaction with intact viruses. As a result, the effort that has gone into the collection and assembly of influenza-specific information/reagents are justified by its utility and conceivable experimental applications.

Identification of Protective Ab and T Cell Influenza A Epitopes. It is well appreciated that not all Ab and T cell responses are protective. Indeed, responses directed against certain influenza-derived epitopes have been reported in a murine animal model to actually exacerbate disease (32, 33). To address this issue, we focus specifically on epitopes for which protective data are available. Protective epitopes are defined herein as those that tested positive in virus challenge or neutralization assays, even though we are aware that caution needs to be exercised in directly equating *in vitro* neutralization assays with *in vivo* protection. Only nine Ab and nine T cell epitopes are identified to meet this criterion (SI Table 10). As a result, these data emphasize the need for more studies that evaluate the protective and neutralizing efficacy of immune responses directed against different epitopes. In particular, focusing the immune response on relatively conserved epitopes is considered as an avenue to develop influenza vaccines, but their prophylactic efficacy as compared with nonconserved ones must be established.

All data presently available are derived from animal models in

Table 3. Proposed research agenda toward a more systematic and comprehensive collection of influenza immune epitopes

Knowledge gap	Proposed research agenda
Only a few protective Ab and T cell epitopes were reported in the literature	Focus on determining protective Ab and T cell epitopes
Paucity of Ab epitopes in comparison with T cell epitopes	Promote and increase Ab epitope identification studies
Limited spectrum of animal hosts (currently predominantly mouse) used for epitope identification	Expand and balance the repertoire of tested host species, especially avian, nonhuman primates, and human
Limited number of epitopes reported for avian influenza strains/subtypes	Focus on identifying epitopes derived from avian influenza viruses
Limited number of epitopes reported from proteins other than HA and NP	Identify epitopes derived from all 10 influenza proteins

hosts such as mice, rabbits, and macaques. To the best of our knowledge, no study defining human protective epitopes has been conducted, most likely because of ethical reasons. The degree of conservation of protective epitopes across different avian H5N1 and human influenza viral strains is also calculated. In general, protective T cell epitopes are highly conserved between human and avian influenza strains. Protective Ab epitopes are, as expected, less conserved. However, one protective Ab epitope from the M2 protein shows appreciable conservation among the selected human influenza strains and H5N1. Because M2 is a relatively conserved protein, identification of protective Ab epitopes derived from this protein, as has been pointed out, holds promise for the future development of a universal influenza epitope-based vaccine (34). However, it has been shown that even the limited degree of sequence variation between this epitope and the homologous H5N1 sequences might result in lack of cross-reactivity (35). Nevertheless, whether these epitopes could be used to induce cross-reactive responses and also confer protection in humans needs to be addressed experimentally.

An important issue that influenza epitope research must address is which epitopes are likely to confer greatest protection. Cross-protective cytotoxic T lymphoctes (CTL) have been the focus of many studies over the last decade, but their impact on influenza infection in human *in vivo* still needs to be conclusively established. Influenza virus appears to be most sensitive to neutralizing Abs, and Abs to HA are more effective than those specific for NA and M2, perhaps the reason why the virus has evolved to evade such responses, just like herpes viruses have evolved strategies to evade CTL responses. In that respect, it could be difficult to find broadly cross-reactive epitopes.

Conclusions

In summary, a comprehensive analysis of influenza A Ab and T cell epitopes indicates that a large set of influenza epitope data exists for researchers to use in their studies. To the best of our knowledge, all characterized epitopes, defined as presented above, were included in the analysis. If however, inadvertently omitted data were brought to our attention, we would be grateful to update IEDB accordingly. Nevertheless, given the present focus of the scientific community on influenza viruses, the amount of data are likely to increase in the near future. Therefore, we are continually updating the IEDB with new epitope information as it becomes available in the literature. These results are publicly accessible to the scientific community, and we are working to integrate our efforts with other bioinformatics resources such as BioHealthBase (28). Several different protective epitopes are found to be conserved, highlighting how the collation of relevant data from disparate sources, and the integration of immunological data with sequence variability information can yield results of great potential impact.

From our perspective, significant knowledge gaps and opportunities for future research in influenza A epitope identification also became apparent, including (*i*) Determination of protective Ab and T cell epitopes (only a few were reported in the literature), (*ii*) paucity of Ab epitopes in comparison with T cell epitopes, (*iii*) limited spectrum of animal hosts used for epitope identification, (*iv*) a limited number of epitopes reported for avian influenza strains/subtypes, and (*v*) a limited number of epitopes reported from proteins other than HA and NP. Based on these gaps, a proposed research agenda toward a more systematic and comprehensive collection of influenza immune epitopes is tabulated in Table 3.

This is a comprehensive analysis of the world-wide knowledge in a given research area, with the specific intent of not only making curated information accessible to the scientific community, but also with the specific goal of revealing gaps and consequent potential vulnerabilities in the available aggregated knowledge. Some of the results are unexpected and illustrate the power of the approach. Future similar analyses may encompass different disease targets of immunological relevance. The results could assist in correlating the amount of knowledge available with the actual importance of a particular disease, analyzing the impact of funding initiatives and other related topics, and transcending basic research and impacting global research and scientific policies.

In conclusion, influenza research is currently of high general interest. This analysis provides researchers with information that can be used to evaluate different vaccine concepts and design new basic studies. The study also provides the general scientific audience with an objective evaluation of which information is well represented within our current literature, which gaps exist, and what might be addressed by future investigations. In addition, the availability of databases relating to influenza A, as recently pointed out, is an important component of our strategy to combat seasonal outbreaks and a potential pandemic (36). Therefore, the influenza A epitope data analysis reported herein represents an important step in this direction. Specifically, the revealed gaps in our collective knowledge might inspire and guide directions for future research in the study of immunity against the influenza A virus.

Materials and Methods

Selection of IEDB-Curated Influenza A Epitopes. To maximize the immunological relevance of the study, IEDB-curated records were filtered to exclude data in which only the epitope was used as both immunogen and test antigen, because such information does not provide data on recognition of the epitope in the context of the whole virus or protein. T cell epitopes identified by MHC binding alone were also excluded from further analyses, because peptide MHC binding implies only that there exists a potential for immunogenicity but does not prove that this potential has or will be realized. This ability to select records

based on relevant assays is a key example of the IEDB flexibility. It should be noted that because an epitope is defined as a distinct molecular structure that interacts with specific immune receptors, largely overlapping or nearly identical structures were counted as separate entries. Similarly, for Ab conformational epitopes, single residues identified by mutant studies were also considered as separate entries. We have recently developed a computer algorithm to specifically cluster similar and related entries, thus mapping to a single structure or "antigenic site" largely overlapping or homologous (>80%) structures. The results obtained after clustering were qualitatively the same, even though the total number of epitopes was reduced by approximately a third. Development of such a filter was important because the inclusion of the extra sequence changed the "conservation score" between subtypes where the actual epitope is conserved. These duplicate entries could be a hindrance to effective use of the database, especially because these longer sequences for Class I "epitopes" include a sequence that is not actually part of the epitope, that is, actually being trimmed off before presentation. However, the database needs to record the original data as reported to avoid bias and data corruption.

Epitope Conservancy Analysis. To determine the conservation of continuous linear Ab and T cell influenza epitopes, we used the epitope conservancy-analysis tool provided in the analysis resources of IEDB. Using an epitope sequence and a set of protein sequences of a given influenza strain, this tool computes the maximum identity level at which the epitope can be found in the given protein sequence set or the influenza strain. For each epitope, the highest epitope identity level in each influenza strain was calculated. For discontinuous Ab epitopes, the algorithm was implemented to identify a matching epitope discontinuous-sequence pattern in a given protein sequence or set. For example, given the epitope discontinuous sequence "A1,B3,C6", its matching sequence pattern is **AXBXXC**, where X is any amino acid residue, and the number of Xs between two nearest known amino acid residues is equal to the gap distance between them. If an epitope's pattern is found within a protein sequence/set, the epitope is considered to be conserved within that protein sequence/set. In addition, the identity level was also calculated based on the known epitope residues. For patternwise matching sequences, the identity level is 100%. To obtain meaningful results, only discontinuous sequences consisting of at least three identified residues were used in the analysis. We emphasize that the algorithm developed here does not predict cross-reactivity but merely detects whether the residues involved in a conformational epitope are conserved in different sequences. Whether this conservancy would translate in Ab cross-reactivity should be experimentally determined. It should also be noted that, in this analysis, conservancy was calculated and reported for all epitope entries even though similar entries may be related to a single "epitope" (the difference being whether flanking residues were included). Because conservancy is not calculated based on the shared epitope subsequence, different conservancy values are expected in the context of different flanking regions for a single epitope. The differences are due to the variations in sizes and amino acid compositions of flanking regions considered by the algorithm in its calculation.

We thank the IEDB curation team (John Mokili, Jong de Castro, Julia Ponomarenko, Laura Zarebski, Michael Alexander, Muthuraman Sathiamurthy, Nima Salimi, Randi Vita, Russell Chan, Leora Zalman, Huda Makhluf, Michael Lyman, and Kerrie Vaughan) for their excellent work in curating influenza A epitope information and Alison Deckhut Augustine, Howard Gray, Ward Fleri, Muthuraman Sathiamurthy, and Randi Vita for helpful suggestions and critical review of the manuscript. This work was supported by the National Institutes of Health's (National Institute of Allergy and Infectious Disease) Contract HHSN26620040006C (Immune Epitope Database and Analysis Program), Contract N01 AI30039 (Epitope-Based Multipeptide Vaccines for Expanded Coverage Against Inter-Pandemic Influenza), and Kirin pharmaceutical division. This is LIAI publication number 768. E.A. was supported by the Wenner-Gren Foundations.

1. Choi YK, Ozaki H, Webby RJ, Webster RG, Peiris JS, Poon L, Butt C, Leung YH, Guan Y (2004) *J Virol* 78:8609–8614.
2. Fouchier RA, Schneeberger PM, Rozendaal FW, Broekman JM, Kemink SA, Munster V, Kuiken T, Rimmelzwaan GF, Schutten M, Van Doornum GJ, *et al.* (2004) *Proc Natl Acad Sci USA* 101:1356–1361.
3. Guo Y, Li J, Cheng X (1999) *Zhonghua Shi Yan He Lin Chuang Bing Du Xue Za Zhi* 13:105–108.
4. Koopmans M, Wilbrink B, Conyn M, Natrop G, van der Nat H, Vennema H, Meijer A, van Steenbergen J, Fouchier R, Osterhaus A, Bosman A (2004) *Lancet* 363:587–593.
5. Peiris M, Yuen KY, Leung CW, Chan KH, Ip PL, Lai RW, Orr WK, Shortridge KF (1999) *Lancet* 354:916–917.
6. World Health Organization (WHO) (2006) www.who.int/csr/disease/avian_influenza/country/cases_table_2006_10_03/en/index.html.
7. Enserink M (2006) *Science* 311:932.
8. Fauci AS (2006) *Cell* 124:665–670.
9. Fauci AS (2006) *Emerg Infect Dis* 12:73–77.
10. U.S. Department of Health and Human Services (HHS) (2005) www.hhs.gov/pandemicflu/plan/.
11. Doherty PC, Turner SJ, Webby RG, Thomas PG (2006) *Nat Immunol* 7:449–455.
12. Enserink M (2005) *Science* 309:996.
13. Wadman M (2005) *Nature* 438:23.
14. Sambhara S, Poland GA (2006) *Lancet* 367:1636–1638.
15. Peters B, Sidney J, Bourne P, Bui HH, Buus S, Doh G, Fleri W, Kronenberg M, Kubo R, Lund O, *et al.* (2005) *PLoS Biol* 3:e91.
16. Peters B, Sidney J, Bourne P, Bui HH, Buus S, Doh G, Fleri W, Kronenberg M, Kubo R, Lund O, *et al.* (2005) *Immunogenetics* 57:326–336.
17. Vita R, Vaughan K, Zarebski L, Salimi N, Fleri W, Grey H, Sathiamurthy M, Mokili J, Bui HH, Bourne PE, *et al.* (2006) *BMC Bioinformatics* 7:341.
18. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A (2006) *BMC Bioinformatics* 7:153.
19. Peters B, Bulik S, Tampe R, Van Endert PM, Holzhutter HG (2003) *J Immunol* 171:1741–1749.
20. Nielsen M, Lundegaard C, Worning P, Lauemoller SL, Lamberth K, Buus S, Brunak S, Lund O (2003) *Protein Sci* 12:1007–1017.
21. Bui HH, Sidney J, Peters B, Sathiamurthy M, Sinichi A, Purton KA, Mothe BR, Chisari FV, Watkins DI, Sette A (2005) *Immunogenetics* 57:304–314.
22. Peters B, Sette A (2005) *BMC Bioinformatics* 6:132.
23. Janeway CA, Travers P, Walport M, Shlomchik MJ (2005) *Immunobiology* (Garland Science, New York).
24. Gupta V, Earl DJ, Deem MW (2006) *Vaccine* 24:3881–3888.
25. Munoz ET, Deem MW (2005) *Vaccine* 23:1144–1148.
26. Macken C, Lu H, Goodman J, Boykin L (2001) in *Options for the Control of Influenza*, eds Osterhaus ADME, Cox N, Hampson AW (Elsevier Science, Amsterdam), pp 103–106.
27. National Center for Biotechnology Information (NCBI) (2004) www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html.
28. BioHealthBase (2006) http://www.biohealthbase.org.
29. Johnson PR, Feldman S, Thompson JM, Mahoney JD, Wright PF (1986) *J Infect Dis* 154:121–127.
30. Palese P, Tumpey TM, Garcia-Sastre A (2006) *Immunity* 24:121–124.
31. Fazekas de St. Groth S, Webster RG (1966) *J Exp Med* 124:331–345.
32. Crowe SR, Miller SC, Woodland DL (2006) *Vaccine* 24:452–456.
33. Crowe SR, Miller SC, Shenyo RM, Woodland DL (2005) *J Immunol* 174:696–701.
34. Kaiser J (2006) *Science* 312:380–382.
35. Fan J, Liang X, Horton MS, Perry HC, Citron MP, Heidecker GJ, Fu TM, Joyce J, Przysiecki CT, Keller PM, *et al.* (2004) *Vaccine* 22:2993–3003.
36. Editorial (2006) *Nature* 440:255–256.

Table 4. Distribution of epitope molecular structures by influenza strains

Subtype	Strains	Antibody			T Cell				Grand Total
		Linear Sequence	Conformational Sequence	Total	CD4	CD8	Undefined	Total	
H1N1	A/PUERTO RICO/8/34(H1N1))	10	5	**15**	22	85	8	**115**	**130**
	A/PUERTO RICO/8/34/MOUNT SINAI(H1N1))	1	6	**7**	6			**6**	**13**
	A/USSR/90/77(H1N1))	2	8	**10**	1			**1**	**11**
	A/WSN/33(H1N1))	4	2	**6**					**6**
	A/SWINE/NEW JERSEY/11/76(H1N1))		1	**1**					**1**
H1N9	A/NWS-G70C(H1N9))		1	**1**					**1**
H2N2	A/ANN ARBOR/6/60(H2N2))					1		**1**	**1**
	A/JAPAN/305/57(H2N2))				1	15		**16**	**16**
	A/OKUDA/57(H2N2))				7		2	**9**	**9**
H3N2	A/AICHI/2/68(H3N2))	2	2	**4**	1			**1**	**5**
	A/BANGKOK/1/79(H3N2))		1	**1**					**1**
	A/BEIJING/32/92(H3N2))				29			**29**	**29**
	A/HONG KONG(H3N2))				15			**15**	**15**
	A/HONG KONG/1/68(H3N2))					1		**1**	**1**
	A/HONG KONG/2/68(H3N2))					2		**2**	**2**
	A/MEMPHIS/1/71(H3N2))	6		**6**	9	1	1	**11**	**17**
	A/MEMPHIS/102/72(H3N2))				3	3		**6**	**6**
	A/MEMPHIS/31/98(H3N2))		6	**6**					**6**
	A/MEMPHIS/6/86(H3N2))	2		**2**					**2**
	A/NETHERLANDS/785E/90 (H3N2))					1		**1**	**1**
	A/NETHERLANDS/889/91 (H3N2))					1		**1**	**1**
	A/NT/60/68/(H3N2))				32	16	1	**49**	**49**
	A/PORT CHALMERS/1/73(H3N2))		7	**7**		1		**1**	**8**
	A/SHANGHAI/16/89(H3N2))					1		**1**	**1**
	A/TEXAS/1/77(H3N2))	2		**2**	2	4	4	**10**	**12**
	A/UDORN/307/72(H3N2))				1	3		**4**	**4**
	A/VICTORIA/3/75(H3N2))	32	8	**40**	1	1		**2**	**42**
	A/WUHAN/359/95(H3N2))	5		**5**					**5**
	A/X-31(H3N2))	11	35	**46**	38	27	78	**143**	**189**
	A/ARGENTINA/3779/94(H3N2))						1	**1**	**1**
	A/CHRIST_CHURCH/2/88(H3N2))						1	**1**	**1**
	A/CORDOBA/3278/96(H3N2))						1	**1**	**1**
	A/FRANCE/75/97(H3N2))						1	**1**	**1**
	A/NANCHANG/58/93(H3N2))						1	**1**	**1**
	A/NEW_YORK/15/94(H3N2))						1	**1**	**1**
	A/NEW_YORK/17/94(H3N2))						1	**1**	**1**
	A/OHIO/3/95(H3N2))						1	**1**	**1**
	A/SHANGDONG/5/94(H3N2))						1	**1**	**1**
	A/SWINE/HONG KONG/126/82(H3N2))						1	**1**	**1**
	A/SYDNEY/05/97-LIKE(H3N2))						1	**1**	**1**
	A/USSR/26/(H3N2))						1	**1**	**1**
	A/LOS_ANGELES/(H3N2))		5	**5**					**5**
	A/PHILIPPINES/2/82(H3N2))		1	**1**					**1**
	INFLUENZA A VIRUS H3N2	1		**1**	2	11	1	**14**	**15**
H3N8	A/DUCK/UKRAINE/1/63(H3N8))		3	**3**					**3**
H5N1	A/VIET NAM/1194/2004(H5N1))					2		**2**	**2**
H5N2	A/CHICKEN/PENNSYLVANIA/1370/83(H5N2))		1	**1**					**1**
	A/MALLARD DUCK/PA/10218/84(H5N2))		5	**5**					**5**
H5N9	A/TURKEY/ONTARIO/7732/66(H5N9))		5	**5**		1		**1**	**6**
H7N1	A/FPV/ROSTOCK/34(H7N1))		1	**1**	3			**3**	**4**
H7N7	A/SEAL/MASS/1/80(H7N7))		1	**1**					**1**
H9N2	A/SWINE/HONG KONG/9/98(H9N2))		2	**2**					**2**
H11N9	A/TERN/AUSTRALIA/G70C/75(H11N9))		10	**10**					**10**
H13N9	A/WHALE/MAINE/1/84(H13N9))		3	**3**					**3**
	A/NWS/33HA-A/TERN/AUSTRALIA/G70C/75NA)		1	**1**					**1**
	A/MEMPHIS/1/71H-A/BELLAMY/42N)					1	1	**2**	**2**
	STRAIN A/EQUINE/NEW MARKET/76)						1	**1**	**1**
	INFLUENZA A VIRUS	7	1	**8**	17	32	1	**50**	**58**

Table 5. Collection of influenza strains for conservancy analysis

Subtype	Strain	Vaccine Coverage
H1N1	A/PR/8/34	1968-75
	A/USSR/90/77	1976-86
	A/Taiwan/1/86	1987-89
	A/Texas/36/91	1989-98
	A/New Caledonia/20/99	1999-2004
	A/Brevig Mission/1/18	
	A/WS/33	
H3N2	A/England/42/72	1968-75
	A/Hong Kong/1/68	1968-75
	A/Bangkok/1/79	1976-86
	A/Leningrad/360/86	1987-89
	A/Beijing/353/89	1989-98
	A/Panama/2007/99	1999-2004
	A/New York/5/2004	
	A/UDORN/307/72	
H5N1	A/Viet Nam/1194/2004	
	A/Hong Kong/156/97	

Table 6. Conservancy analysis of antibody linear epitope sequences

No.	Sequence	Influenza Source Subtype	Source Protein	Antibody Type(s)	Host Species	H1N1							H3N2								H5N1	
						A/Brevig Mission/1/18	A/New Caledonia/20/99	A/PR/8/34	A/Taiwan/1/86	A/Texas/36/91	A/USSR/90/77	A/WS/33	A/Bangkok/1/79	A/Beijing/353/89	A/England/42/72	A/Hong Kong/1/68	A/Leningrad/360/86	A/New York/5/2004	A/Panama/2007/99	A/JDORN/307/72	A/Hong Kong/156/97	A/Viet Nam/1194/2004
1	AIYHTENAYVSVVSSHYNR		HA	MONOCLONAL		58	89	74	100	95	79	63	32	32	32	32	32	32	32	32	32	32
2	AMEQMAGSSEQAAEAMEVASQARQMVQA MRTIGTHPSSS	H1N1	M1	MONOCLONAL	MOUSE	100	97	100	97	97	97	95	97	95	97	97	97	95	95	97	100	92
3	CKRGPDSGFFSRLNWLY	H3N2	HA	POLYCLONAL	RABBIT	35	35	35	35	35	35	35	82	71	94	88	76	65	65	100	35	35
4	CKRGPDSGFFSRLNWLYKSGSTYPVQNVT MPNNDNS	H3N2	HA	POLYCLONAL	RABBIT	36	28	25	28	31	28	33	75	61	89	89	64	56	56	94	31	33
5	CLGHHAVPNGTLVKTITNDQIEVTNATELVQ SSSTGKIC	H3N2	HA	POLYCLONAL	RABBIT	33	33	33	33	33	33	33	97	97	100	97	97	95	95	100	33	33
6	CNNPHRIL	H3N2	HA	POLYCLONAL	RABBIT	38	38	38	38	38	50	38	75	75	100	100	75	63	63	100	38	50
7	CNNPHRILDGINC	H3N2	HA	POLYCLONAL	RABBIT	38	38	31	38	38	38	38	77	77	85	92	77	69	69	92	38	38
8	CNNPHRILDGINCTLIDALLGDPHCDGFQNE KWDL	H3N2	HA	POLYCLONAL	RABBIT	23	23	23	23	23	20	26	91	86	89	91	91	83	83	94	31	29
9	CPKYVKQNTLKLATGMRNVPEKQT	H3N2	HA	MONOCLONAL; POLYCLONAL	MOUSE;RABBIT	54	50	50	50	50	50	50	100	96	100	100	96	96	96	100	63	63
10	CPKYVKQNTLKLATGMRNVPEKQTR	H3N2	HA	POLYCLONAL	RABBIT	56	52	52	52	52	52	52	100	96	100	100	96	96	96	100	64	64
11	DCTLIDALLGDPH	H3N2	HA	POLYCLONAL	MOUSE	38	38	38	38	38	38	38	92	92	100	100	92	85	92	100	46	46
12	DPNNMDKAVKLYRKLKREITFHGAKEIALSY	H1N1	M1	MONOCLONAL		94	97	97	97	97	97	100	97	97	97	97	94	100	100	97	87	90
13	DVPDYAS	H3N2	HA	MONOCLONAL	MOUSE	43	43	43	43	43	43	43	100	100	86	100	100	100	100	100	43	43
14	DVPDYASL	H3N2	HA	MONOCLONAL	MOUSE	50	50	50	50	50	50	50	100	100	88	100	100	100	100	100	50	50
15	EGSYPKLKNSYENK		HA	MONOCLONAL	MOUSE	57	50	93	57	50	57	64	43	43	43	43	43	43	43	43	43	43
16	EGSYPKLKNSYVNK	H1N1	HA	MONOCLONAL	MOUSE	64	57	100	64	57	64	71	36	43	36	36	43	43	43	36	43	43
17	EKQT	H3N2	HA	POLYCLONAL	MOUSE	75	75	75	75	75	75	75	100	100	100	100	100	100	100	100	75	75
18	ETPIRNEWGCR	H3N2	M2	POLYCLONAL	RABBIT	91	82	100	82	82	100	100	82	82	100	100	82	100	82	82	73	82
19	EVETPIRN	H1N1	M2	MONOCLONAL	MOUSE	88	63	100	63	63	100	100	63	63	100	100	63	100	63	63	75	88
20	FQNEKWDL	H3N2	HA	MONOCLONAL	MOUSE	38	50	50	50	50	38	38	100	75	75	88	100	88	75	88	38	38
21	GFFSRLNWLTKS	H3N2	HA	POLYCLONAL	RABBIT	50	42	42	42	50	42	50	75	75	83	100	75	75	67	92	42	42
22	GKICNNPHRILDGIDCTLID	H3N2	HA	MONOCLONAL	MOUSE	30	30	30	30	30	30	30	75	75	95	100	75	70	70	100	30	30
23	GKVTVSTKRSQQTIIPNVGSRPWVRGL	H3N2	HA	POLYCLONAL	RABBIT	26	26	30	30	30	30	30	93	89	89	85	85	74	81	93	30	26
24	GLFGAIAGFIE	H1N1	HA	MONOCLONAL; POLYCLONAL	GOAT;MOUSE;RABBIT	100	100	100	100	100	100	100	91	91	82	100	91	91	91	100	100	100
25	GLFGAIAGFIENGWEGMIDGWYGFRHQNS EGTGQA	H3N2	HA	MONOCLONAL	MOUSE	77	74	77	77	77	77	77	94	94	23	100	94	94	94	100	66	71
26	GLIYNRMGAVTTEVAFGLVCATCEQIADSQ HRSHRQ	H1N1	M1	MONOCLONAL	MOUSE	97	92	100	92	92	94	100	100	100	100	100	100	100	100	100	92	97
27	GVTQNGGSSACKRGPDSGFFSR	H3N2	HA	POLYCLONAL	RABBIT	32	32	27	32	32	32	32	77	64	91	91	68	73	68	95	36	36
28	HCDGFQNEKWDL	H3N2	HA	MONOCLONAL	MOUSE	33	33	42	33	33	33	33	100	83	83	83	100	83	83	92	33	33
29	HCDGFQNEKWDLFVE	H3N2	HA	POLYCLONAL	RABBIT	33	33	33	33	33	33	33	100	87	87	87	93	87	87	93	33	33
30	HCDGFQNEKWDLFVERSKAFSNCYPYDVP DYASLRS	H3N2	HA	MONOCLONAL; POLYCLONAL	MOUSE;RABBIT	25	22	25	25	25	25	25	100	92	92	94	97	92	92	97	25	25
31	HHPITDSDQTRLY	H3N2	HA	POLYCLONAL	RABBIT	62	54	46	46	46	46	46	69	69	62	62	62	77	69	69	54	54
32	HHPSTDKEQTNLY	H3N2	HA	POLYCLONAL	RABBIT	54	46	62	46	46	46	62	92	85	69	77	85	62	62	85	62	62
33	KAYSNCYPYDVPDY	H3N2	HA	POLYCLONAL	RABBIT	43	36	43	36	36	36	43	93	100	93	93	93	100	100	93	43	43
34	KWDLFVERSK	H3N2	HA	MONOCLONAL	MOUSE	50	50	50	50	50	50	50	100	90	90	90	90	100	90	90	50	50
35	LKLAT	H3N2	HA	MONOCLONAL; POLYCLONAL	MOUSE	60	60	60	60	60	60	60	100	100	100	100	100	100	100	100	80	80
36	LKTRPILSPLTKGILGFVFTLTVPSERGLQRR RFVQNALNGNGD	H1N1	M1	MONOCLONAL	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
37	MRNVPEKQT	H3N2	HA	POLYCLONAL	MOUSE	44	44	44	44	44	44	44	100	100	100	100	100	100	100	100	56	44
38	MSLLTEVETPIRNEWGCRCNDSSD	H1N1	M2	MONOCLONAL	MOUSE	96	63	96	63	63	100	100	63	63	100	100	63	96	63	63	83	88
39	MSLLTEVETPIRNEWGCRCNGSSD	H1N1	M2	POLYCLONAL	MOUSE	92	58	100	58	58	96	96	58	58	96	96	58	92	58	58	79	83
40	NATELVQSSSTGKICNNPHRILDGINC	H3N2	HA	POLYCLONAL	RABBIT	26	26	26	26	26	26	26	85	85	93	96	85	81	81	96	26	26
41	NEWGCRCNDSSD	H3N2	M2	POLYCLONAL	RABBIT	100	100	92	100	100	100	100	100	100	100	100	100	92	100	100	83	83
42	NSDKLYIWGVHHPSTDKEQTNLY	H3N2	HA	POLYCLONAL	RABBIT	48	43	48	43	43	43	52	91	83	78	83	83	70	70	87	57	52
43	NVPEKQT	H3N2	HA	MONOCLONAL	MOUSE	43	43	43	43	43	43	43	100	100	100	100	100	100	100	100	57	57
44	NVPEKQTRGIFGAIAGFIE		HA		MOUSE	74	74	74	74	74	74	74	100	100	95	95	100	100	100	95	58	58
45	QDLPGNDNNSTATLC	H3N2	HA	POLYCLONAL	RABBIT	33	33	33	33	33	33	33	47	47	53	53	47	47	47	47	33	33
46	QDLPGNDNNSTATLCLGHHAVPNGTLVKTI TNDQIE	H3N2	HA	POLYCLONAL	RABBIT	22	22	22	22	22	22	25	78	78	78	75	78	75	75	78	22	22
47	SKAFSNCYPYDVPDYASL	H3N2	HA	POLYCLONAL	MOUSE;RABBIT	39	33	39	33	33	33	39	100	94	94	100	100	94	94	100	44	39
48	SLLTEVETPIR		M2	POLYCLONAL	RABBIT	91	73	100	73	73	100	100	73	73	100	100	73	100	73	73	82	91
49	SLLTEVETPIRNEWGCRCNDSSD	H1N1;H3N2	M2	MONOCLONAL; POLYCLONAL	MOUSE;RHESUS MONKEY	96	65	96	65	65	100	100	65	65	100	100	65	96	65	65	83	87
50	SLLTEVETPIRNEWGCRCNDSSDP	H1N1	M2	MONOCLONAL	MOUSE	96	67	96	67	67	100	100	67	67	100	100	67	96	67	67	83	88
51	TQNGGSSACKRGPDS	H3N2	HA	POLYCLONAL	RABBIT	33	33	33	33	33	33	33	73	53	93	87	60	67	60	93	40	40
52	VERSKAFSNCYPYDVPDYASLRS	H3N2	HA	MONOCLONAL	MOUSE	35	30	35	30	30	30	35	100	96	96	100	96	96	96	100	35	30
53	VTGLRNIPSIQSR	H1N1	HA	POLYCLONAL	MOUSE	92	100	100	100	100	100	92	54	54	54	54	54	54	54	54	54	54
54	VTGLRNIPSIQSRGLFGAIAGFIEG	H1N1	HA		MOUSE	96	100	100	100	100	100	96	68	68	64	72	68	68	68	72	52	52
55	WTGVAQD	H3N2	HA	POLYCLONAL	RABBIT	43	43	43	43	43	43	43	71	86	71	71	71	71	86	71	43	43
56	WTGVTQN	H3N2	HA	POLYCLONAL	MOUSE	43	43	43	43	43	43	43	86	71	100	100	86	100	86	100	43	43
57	YDVPDYAS	H3N2	HA	MONOCLONAL	MOUSE	38	50	50	50	50	50	50	100	100	88	100	100	100	100	100	50	50
58	YPYDVPDYA	H3N2	HA	MONOCLONAL	MOUSE	56	44	56	44	44	44	56	100	100	100	100	100	100	100	100	44	56
59	YPYDVPDYAS	H3N2	HA	MONOCLONAL		50	40	50	40	40	40	50	100	100	90	100	100	100	100	100	40	50
60	CYPYDVPDY	H3N2	HA	POLYCLONAL	RABBIT	67	56	67	56	56	56	67	100	100	100	100	100	100	100	100	56	67
61	DYASLRSLVASSGTLEFINEGFNWTGVTQN GGSSAC	H3N2	HA	POLYCLONAL	RABBIT	22	25	22	22	22	22	25	94	86	92	92	92	89	89	92	22	22
62	LCLGHHAVPNGTLVKTITNDQIEVTNATELV QSSSTGKI	H3N2	HA	POLYCLONAL	RABBIT	31	31	31	31	31	31	31	97	97	100	97	97	95	95	100	31	31
63	NSDKLYIWGVHHPSTDKEQTNLYV	H3N2	HA	POLYCLONAL	RABBIT	46	42	46	42	42	42	50	92	83	79	83	83	67	67	88	54	50
64	QDLPGNDNNSTATLCLGHHAVPNGTLVKTI TNDQIEVTN	H3N2	HA	POLYCLONAL	RABBIT	26	26	26	26	26	26	28	79	79	79	77	79	77	77	79	26	23
65	SSIMRSDAPIGTCSSECITPNGSIPNDKPFQ NVNKIT	H3N2	HA	POLYCLONAL	RABBIT	46	46	41	43	43	41	43	100	97	97	95	100	92	92	97	41	41
66	TNATELVQSSSTGKICNNPHRILDGIN	H3N2	HA	POLYCLONAL	RABBIT	26	26	30	26	26	26	26	85	85	93	96	85	81	81	96	26	26
67	FESTGNLI		HA	MONOCLONAL	MOUSE	88	75	75	75	75	75	88	63	75	63	63	75	75	75	63	75	75
68	GFRHQNSEGTGQAADL	H3N2	HA	MONOCLONAL	MOUSE	56	56	56	56	56	56	56	100	100	31	100	100	100	100	100	38	50
69	KRGPGSG	H3N2	HA	MONOCLONAL	MOUSE	71	71	71	71	71	71	57	57	57	71	100	57	57	57	86	57	57
70	LTEVETPIRNEWG	H1N1	M2	MONOCLONAL	MOUSE	92	54	100	54	54	100	100	54	54	100	100	54	100	54	54	77	85
71	RSQQTII	H3N2	HA	MONOCLONAL	MOUSE	57	57	57	57	57	57	57	100	86	86	100	86	86	86	100	57	57
72	SVSSFERFEIFPK		HA	MONOCLONAL	MOUSE	92	100	100	100	100	100	92	38	38	38	38	38	38	38	38	38	38
73	TNQEQTSLYV	H3N2	HA	MONOCLONAL	MOUSE	50	40	40	40	40	40	50	70	70	90	100	70	50	40	90	50	50
74	TYQRTRALV	H1N1	NP	POLYCLONAL	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
75	WLTEKEGSYP	H1N1	HA	MONOCLONAL	MOUSE	70	70	100	90	70	90	70	60	50	50	50	60	50	40	40	50	50

Identity level color code:
Yellow: 100%
Magenta: ≥90%
Green: ≥80%

Table 7. Conservancy analysis of T cell linear epitope sequences

No.	Sequence	Influenza Source Subtype	Source Protein	MHC Restriction Allele(s)	Host Species	H1N1							H3N2								H5N1	
						A/Brevig Mission/1/18	A/New Caledonia/20/99	A/PR/8/34	A/Taiwan/1/86	A/Texas/36/91	A/USSR/90/77	A/WS/33	A/Bangkok/1/79	A/Beijing/353/89	A/England/42/72	A/Hong Kong/1/68	A/Leningrad/360/86	A/New York/5/2004	A/Panama/2007/99	A/UDORN/307/72	A/Hong Kong/156/97	A/Viet Nam/1194/2004
1	AAFEDLRVLSFIRG	H3N2	NP	HLA-B37	HUMAN	93	93	93	93	93	93	93	93	93	93	100	93	93	93	93	93	93
2	ADLKSTQAAIDQING	H3N2	HA		MOUSE	67	60	60	60	60	60	60	100	100	33	100	100	93	93	100	53	53
3	AELLVALEN	H3N2	HA	H-2-IAD	MOUSE	56	89	89	89	89	89	89	100	100	44	100	100	89	100	100	78	78
4	AELLVALENQHTIDL	H3N2	HA	H-2-B CLASS I	MOUSE	33	67	67	67	67	67	67	100	100	33	100	93	93	100	100	60	60
5	AGFIENGWEGMVDGWYGFRHQNSEGTGQAADLKS	H3N2	HA		HUMAN	71	74	71	71	71	71	71	100	100	29	97	100	100	100	97	62	68
6	AHKSCLPACVYGPAV	H3N2	NP		MOUSE	100	100	100	93	93	100	93	100	100	100	100	100	100	100	100	93	93
7	AIMDKNIIL	H1N1;H3N2	NS1	HLA-A*0201;HLA-A*020101;HLA-A2.1	HUMAN;MOUSE	100	100	100	100	100	100	100	78	78	89	89	78	78	78	89	89	89
8	AIMDKNIML	H3N2	NS1	HLA-A2.1	HUMAN	89	89	89	89	89	89	89	89	89	100	100	89	89	89	100	78	78
9	AKNMEYDA	H1N1	PB1		MOUSE	100	100	100	100	100	100	100	88	88	88	88	88	88	88	88	88	88
10	ALENQHTIDLTDSEM	H3N2	HA	H-2-B CLASS I	MOUSE	33	47	47	47	47	47	40	100	100	33	100	93	100	100	100	40	40
11	ALNNRFQIKGVEL	H3N2	HA	H-2-IED	MOUSE	38	46	38	46	46	46	46	100	100	38	100	100	100	100	100	54	46
12	ALNNRFQIKGVELKS	H3N2	HA		MOUSE	33	47	40	47	47	47	47	100	100	33	100	100	100	100	100	53	47
13	ARLGKGYMF	H1N1	PB1	H-2-DK;HLA-B*2705	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
14	ARSALILRGSVAHK	H2N2	NP	H-2-S CLASS II	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
15	ARSALILRGSVAHKSCLPACVYGP	H2N2;H3N2	NP		MOUSE	100	100	100	100	96	100	96	100	100	100	100	100	100	100	100	96	96
16	ASAGQISVQPAFSVQRNLP	H3N2	NP	H-2-D CLASS II	HUMAN;MOUSE	95	89	89	89	89	89	84	95	95	100	100	95	95	95	100	95	95
17	ASCMGLIY	H3N2	M1	HLA-B*35;HLA-B*3501	HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
18	ASGRVTVSTKRSQQTV	H3N2	HA		HUMAN	31	31	31	31	31	31	31	94	100	88	88	100	94	100	94	31	31
19	ASMHECNTKCQT	H1N1	HA	H-2-IAD	MOUSE	75	67	100	75	75	83	100	42	42	42	42	42	42	42	42	58	58
20	ASNENMDAM	H3N2	NP	H-2-D CLASS I;H-2-DB	MOUSE	78	89	78	89	78	89	78	89	89	89	100	89	89	89	89	78	89
21	ASNENMDAMESSTL	H3N2	NP	H-2-DB	MOUSE	79	86	86	86	79	86	86	93	93	93	100	93	86	86	93	79	79
22	ASNENMDTM	H3N2	NP	H-2-DB	MOUSE	89	78	89	100	78	100	89	100	89	100	89	89	89	89	100	67	78
23	ASNENMETM	H1N1;H2N3;H3N2	NP	H-2-B CLASS I;H-2-DB	MOUSE	100	67	100	89	67	89	100	89	78	89	78	78	78	78	89	78	89
24	ASNENMETMESSTLE	H3N2	NP	H-2-DB	MOUSE	93	73	100	87	73	87	100	93	87	93	87	87	80	80	93	80	80
25	ASQGTKRSYEQMETDGERQNATE	H3N2	NP		HUMAN	100	100	100	96	100	100	91	100	100	100	100	96	96	96	100	96	96
26	ATGLRNVPQIESR	H2N2	HA	H-2-IED	MOUSE	77	69	69	69	69	69	62	62	62	62	62	62	62	62	62	77	77
27	ATGMRNVPEKQTR	H3N2	HA	H-2-IED	MOUSE	62	54	54	54	54	54	54	100	100	100	100	100	100	100	100	54	54
28	ATGMRNVPEKQTRGIFGAIAGFIENGWEGMVD	H3N2	HA		HUMAN	72	72	69	69	69	69	69	100	100	25	94	100	100	100	94	53	53
29	ATYQRTRALVRTGMD	H3N2	NP		MOUSE	100	93	100	93	93	93	100	93	93	93	93	93	100	93	93	100	100
30	AVKGVGTMVMELIRMIKRGINDRN	H3N2	NP	H-2-D CLASS II	HUMAN;MOUSE	100	96	96	96	96	96	100	100	96	100	100	96	96	92	100	96	100
31	AYERMCNIL	H1N1;H2N3	NP	H-2-D CLASS I;H-2-KD	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
32	AYERMCNILKGK	H1N1	NP	H-2-IAD	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
33	AYQKRMGVQMQR	H7N1	M1	HLA-DQ	HUMAN	100	100	100	100	100	100	100	92	92	100	100	92	92	92	100	100	100
34	CAAMDDFQLIPMISK	H3N2	PA		MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
35	CKISPLMVAYMLERE	H3N2	PB2	H-2-B CLASS I	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	93	93
36	CPIRGWAI	H1N1	NA	H-2-DD	MOUSE	88	75	100	88	100	100	100	50	50	50	50	50	50	50	50	88	75
37	CPKYVKQNTLKLAT	H3N2	HA	HLA-DR53	HUMAN	57	50	50	50	50	50	50	100	93	100	100	93	93	93	100	79	79
38	CPKYVKQNTLKLATG	H3N2	HA	H-2-B CLASS II;HLA-DRB1*1101;HLA-DRB1*1301	HUMAN;MOUSE	60	53	53	53	53	53	53	100	93	100	100	93	93	93	100	80	80
39	CPKYVKQNTLKLATGMRNV		HA	HLA-DRB1*0101	HUMAN	58	53	53	53	53	53	53	100	95	100	100	95	95	95	100	74	74
40	CPKYVKQNTLKLATGMRNVPEKQT	H3N2	HA	H-2-D CLASS II	MOUSE	54	50	50	50	50	50	50	100	96	100	100	96	96	96	100	63	63
41	CPKYVKQNTLKLATGMRNVPEKQTR	H3N2	HA		HUMAN	56	52	52	52	52	52	52	100	96	100	100	96	96	96	100	64	64
42	CPKYVRSAKLRM	H1N1	HA	H-2-IED	MOUSE	92	100	100	92	92	92	92	50	50	50	50	50	50	50	50	58	58
43	CSQRSKFLLMDALKL	H3N2	PA	H-2-B CLASS I	MOUSE	100	87	100	93	87	100	100	93	93	93	93	93	93	93	93	100	100
44	CTELKLSDY	H3N2	NP	HLA-A*0101;HLA-A1	HUMAN	100	89	100	89	89	100	100	100	100	100	100	100	100	100	100	89	100
45	CVNGSCFTV	H1N1	NA	HLA-A*0201	HUMAN;MOUSE	89	89	89	89	89	89	89	56	56	56	56	56	56	56	56	100	100
46	CYPYDVPDYASLRSLV	H3N2	HA		HUMAN	50	44	50	44	44	44	50	100	100	94	100	100	100	100	100	44	50
47	CYPYDVPDYASLRSLVASS	H3N2	HA	HLA-DR	HUMAN	47	42	47	42	42	42	47	100	100	95	100	100	100	100	100	37	42
48	CYPYDVPDYASLRSLVASSGTLEFINEDFNWT	H3N2	HA		HUMAN	31	28	31	28	28	28	34	97	100	91	91	97	91	94	91	28	31
49	DALLGDPHCDGFQNET	H3N2	HA	H-2-IAK	MOUSE	38	38	38	38	38	38	38	94	88	94	94	94	81	88	100	50	50
50	DALLGDPHCDVFQNET	H3N2	HA	H-2-IAK	MOUSE	38	31	38	31	31	31	38	88	81	88	100	88	75	81	94	50	50
51	DCTLIDALLGDPH	H3N2	HA	H-2-IAD	MOUSE	38	38	38	38	38	38	38	92	92	100	100	92	85	92	100	46	46
52	DDATAGLTHMMIWHS	H3N2	NP		MOUSE	93	93	100	93	93	100	100	100	100	100	100	100	93	93	100	93	87
53	DFQLIPMISKCRTKE	H3N2	PA		MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
54	DPRMCSLMQGSTLP	H3N2	NP		HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
55	DQSLPPNFSSLENFR	H3N2	PA		MOUSE	100	87	100	93	93	93	100	93	93	93	93	93	93	93	93	100	100
56	DRLRRDQKS	H1N1;H3N2	NS1	HLA-DR3	HUMAN	100	100	100	100	100	100	100	89	89	100	100	89	89	89	100	100	100
57	DYASLRSLVASSGTLEFINEGFNWTGVTQNGGSSAC	H3N2	HA		HUMAN	22	25	22	22	22	22	25	94	86	92	92	92	89	89	92	22	22
58	DYQGRLCNPLNPFVS	H3N2	PB1		MOUSE	100	100	100	100	100	100	93	100	100	100	100	100	93	87	100	100	100
59	EALNNRFQIKGVELKS	H3N2	HA		MOUSE	38	50	44	50	50	50	50	100	100	38	100	100	100	100	100	56	50
60	EAMEVASQARQMVQA	H3N2	M1		MOUSE	100	100	100	100	100	100	87	100	93	100	100	100	93	93	100	100	87
61	EDLTFLARSAL	H1N1	NP		MOUSE	91	100	100	100	100	91	91	82	82	91	91	82	91	91	91	91	91
62	EEGAIVGEI	H1N1;H3N2	NS1	H-2-KK	MOUSE	100	100	100	100	100	100	89	100	100	100	100	100	100	89	100	89	100
63	EGIPLYDA	H1N1	PA		MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
64	EIAERPKVRDQAG	H1N1	HA	H-2-IAD	MOUSE	92	85	100	77	85	92	85	46	46	46	46	46	46	46	46	69	54
65	EITGTMRKLADQSLP	H3N2	PA		MOUSE	93	93	100	93	93	93	100	93	87	93	93	80	87	87	93	93	93
66	EKYVEDTKIDLWSYN	H3N2	HA		MOUSE	33	47	47	47	47	47	47	100	100	33	100	100	100	100	100	47	47
67	ELRSRYWAI	H1N1;H3N2	NP	HLA-B8	HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	89	89	100	100	100
68	ELRSRYWAIRTSG	H1N1	NP	HLA-B27		100	100	100	100	100	100	100	100	100	100	100	100	93	93	100	100	100
69	ENQHTIDLTDSEMNKLFEKTRKQLRENAEDMGNGCF	H3N2	HA		HUMAN	25	50	50	50	50	50	47	97	100	25	97	97	94	94	97	53	53
70	ENSFEQITFMQALHL	H3N2	NS2		MOUSE	93	93	100	93	93	93	93	93	93	93	93	93	93	93	93	93	93
71	EQTSLYVQASGRVTV	H3N2	HA	H-2-B CLASS II	MOUSE	47	33	33	33	33	33	47	87	87	93	100	87	73	73	100	40	40
72	ERELVRKTR	H1N1	PB2	H-2-D CLASS I	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
73	ERRNKYLEEHPSAGKDPKKT	H3N2	NP	HLA-DR1	HUMAN	100	100	100	100	100	100	100	95	95	95	100	95	95	95	95	95	95
74	EVHIYYLEKANKIKS	H3N2	PA		MOUSE	100	100	100	100	100	100	100	100	100	100	93	100	100	100	100	100	93
75	EVSHCRATEYIMKGV	H3N2	PA		MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
76	FEANGNLI	H1N1;H1N2	HA	H-2-KK;MAMU-A*11	MOUSE	88	100	100	100	100	100	88	63	50	63	63	50	50	50	63	75	75

Table 7. Conservancy analysis of T cell linear epitope sequences

No.	Sequence	Influenza Source Subtype	Source Protein	MHC Restriction Allele(s)	Host Species	H1N1							H3N2								H5N1	
						A/Brevig Mission/1/18	A/New Caledonia/20/99	A/PR/8/34	A/Taiwan/1/86	A/Texas/36/91	A/USSR/90/77	A/WS/33	A/Bangkok/1/79	A/Beijing/353/89	A/England/42/72	A/Hong Kong/1/68	A/Leningrad/360/86	A/New York/5/2004	A/Panama/2007/99	A/UDORN/307/72	A/Hong Kong/156/97	A/Viet Nam/1194/2004
77	FEDLRVLS	H3N2	NP	HLA-B37	HUMAN	88	88	100	88	88	88	88	88	88	88	100	88	88	88	88	88	88
78	FEDLRVLSF	H11N6	NP	H-2-KK;HLA-B44		89	89	100	89	89	89	89	89	89	89	100	89	89	89	89	89	89
79	FERFEIFPK	H1N1	HA	H-2-IED	MOUSE	89	100	100	100	100	100	89	44	44	56	56	44	44	44	56	56	56
80	FERFEIFPKE	H1N1	HA	H-2-D CLASS II	MOUSE	80	100	100	100	100	100	90	40	40	50	50	40	40	40	50	50	50
81	FESTGNLI	H2N2	HA	H-2-KK;MAMU-A*11	MOUSE	88	75	75	75	75	75	88	63	75	63	63	75	75	75	63	75	75
82	FFPSSSYRRPVGISS	H3N2	PB1	H-2-KB	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
83	FITEGFTWTGVTQNGGSNAC	H3N2	HA	H-2-IAK	MOUSE	30	35	30	30	35	30	30	80	65	95	100	75	70	65	95	30	30
84	FNKACELTDSSWIEL	H3N2	PA		MOUSE	100	93	100	93	93	93	100	93	93	93	93	93	93	93	93	100	100
85	FSVIFDRL	H1N1;H3N2	NS1	H-2-KB	MOUSE	100	100	100	100	100	100	100	100	100	88	100	100	100	88	100	88	100
86	FVERSKAYSNCYPYDV	H3N2	HA	HLA-DRB1*0701	HUMAN	31	31	31	31	31	31	31	94	100	94	94	88	100	100	94	31	31
87	FVFTLTVPSER	H1N1	M1	HLA-DR4	HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
88	FWRGENGRKTRIAYE	H3N2	NP		MOUSE	93	100	100	100	100	100	93	93	93	93	93	93	93	93	93	93	93
89	FWRGENGRKTRSAYERMCNILKGK	H3N2	NP	H-2-D CLASS II;HLA-DR1	HUMAN;MOUSE	92	96	96	96	96	96	92	100	100	96	100	100	100	100	96	92	92
90	FYIQMCTEL	H1N1;H2N3	NP	H-2-D CLASS I;H2-KD	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
91	GACPRYVKQNTLKLAT	H3N2	HA	HLA-DRB1*0401;HLA-DRB1*0701	HUMAN	50	44	44	44	44	44	44	94	100	94	94	100	100	100	94	69	69
92	GACPRYVKQNTLKLATGMRNV	H3N2	HA	HLA-DR	HUMAN	52	48	48	48	48	48	48	95	100	95	95	100	100	100	95	67	67
93	GAKPEEMSFQGRGVFELSDEKAANP	H3N2	NP		HUMAN	76	84	80	84	80	84	76	88	88	84	100	88	88	88	88	80	80
94	GDDATAGLTHMMIWH	H3N2	NP		MOUSE	93	93	100	93	93	100	100	100	100	100	100	100	93	93	100	93	87
95	GDPHCDGFQCKEWDLFVERSKAYSNCY PYDVPDYAS		HA		HUMAN	22	22	22	19	19	22	22	89	97	83	86	86	92	97	89	22	22
96	GEISPLPSL	H3N2	NS1	HLA-B44;MAMU-A*11		100	100	100	100	100	100	100	89	89	100	100	89	89	89	89	100	100
97	GENMAPEKVDFDDCK	H3N2	PA		MOUSE	100	100	100	100	100	100	100	87	87	87	87	87	87	87	87	93	93
98	GERQNATEI	H1N1	NP	H-2-DB;HLA-B44	MOUSE	100	100	100	89	100	100	100	100	100	100	100	89	89	89	100	100	100
99	GGLPFSLL	H1N1		H-2-KB	MOUSE	63	63	63	63	63	63	63	63	63	63	63	63	63	63	63	63	63
100	GILGFVFTL	H1N1;H3N2;H5N1	M1	HLA-A*0201;HLA-A*020101;HLA-A*0203;HLA-A*0206;HLA-A2;HLA-A2.1	HUMAN;MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
101	GILGFVFTLT	H1N1	M1	HLA-A*0201;HLA-A2	HUMAN;MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
102	GILGFVFTLTV	H3N2	M1	HLA-A2;HLA-AW69	HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
103	GIPLYDAI	H1N1	PA	H-2-KB		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
104	GKICNNPHRILDGIDCTLID	H3N2	HA	H-2-K CLASS II	MOUSE	30	30	30	30	30	30	30	75	75	95	100	75	70	70	100	30	30
105	GKLSTRGVQIASNEN	H3N2	NP		MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	93	93
106	GKNTDLEVLMEWLKTRPILS	H1N1	M1	HLA-A2	HUMAN	95	95	100	95	95	95	100	95	95	95	95	95	95	95	95	95	95
107	GLRISSSFSFGGFTF	H3N2	PB2		MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
108	GNGCFKIYHKCDNACI	H3N2	HA	HLA-DRB1*0101;HLA-DRB1*0701	HUMAN	31	69	75	69	69	69	75	100	100	31	100	100	100	100	100	75	75
109	GPAVASGYDFEKEGYSLVG	H3N2	NP		HUMAN	95	100	95	95	100	95	89	95	95	100	100	95	95	89	100	89	89
110	GPLKAEIAQRLE		M1	HLA-DR1	HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	92
111	GRICDSPHRILDGKNCTLIDALLGDPHCD GFQCK		HA		HUMAN	24	24	24	24	24	21	24	94	97	76	76	94	85	91	79	26	26
112	GRIQDLEKYVEDTKIDLWS	H3N2	HA	HLA-DR	HUMAN	32	37	37	37	37	37	37	100	100	32	100	100	100	100	100	42	42
113	GSNACKRGPGSGFFS	H3N2	HA		MOUSE	40	40	40	40	40	40	33	67	60	87	100	60	53	53	93	40	40
114	GTGQAADLKSTQAAI	H3N2	HA		MOUSE	73	73	73	73	73	73	73	100	100	33	100	100	100	100	100	53	67
115	GTMVMELIRMIKRG	H2N2	NP		MOUSE	100	93	93	93	93	93	100	100	100	100	100	100	100	93	100	100	100
116	GWKEPNVVKPHEKGI	H3N2	PA		MOUSE	100	87	100	87	87	93	100	87	87	87	87	87	87	87	87	80	93
117	GYEEFTMV	H1N1	PB2	H-2-KB		100	100	100	88	100	100	100	100	100	100	100	100	100	100	100	100	100
118	GYKDWILWI	H3N2	HA	H-2-D CLASS I	MOUSE	56	56	56	56	56	56	56	100	100	56	100	100	100	100	100	56	56
119	HEKGINPNYLLSWKQ	H3N2	PA		MOUSE	93	93	100	93	93	100	100	93	93	100	100	93	93	93	100	93	93
120	HHPSTDRDQTSLYVRASGRVTVSTKRS QQTVTPNI	H3N2	HA		HUMAN	43	34	37	34	34	34	37	86	89	80	80	83	80	83	86	34	34
121	HNTNGVTAACSHE	H1N1	HA	H-2-IAD	MOUSE	62	54	69	69	54	62	69	38	38	38	38	38	38	38	38	38	38
122	HRILDGIDCTLIDALLGDPHC	H3N2	HA	H-2-IAD	MOUSE	33	29	33	29	29	29	33	90	90	95	100	90	81	86	100	38	38
123	HTIDLTDSEMNKLFE	H3N2	HA		MOUSE	40	40	40	40	40	40	40	100	100	40	100	93	100	100	100	40	40
124	IAPRGYFKIRNGKSSIMRSDAPIGTCSSE CIT	H3N2	HA		HUMAN	31	28	25	25	25	25	25	97	97	91	88	97	91	91	91	31	31
125	IASNENMDAMESSTLE	H2N2	NP		MOUSE	81	88	88	88	81	88	88	94	94	94	100	94	88	88	94	81	81
126	IASNENMETMESSTLE	H1N1	NP	H-2-DB	MOUSE	94	75	100	88	75	88	100	94	88	94	88	88	81	81	94	81	81
127	IEGGWTGMI	H1N1	HA	H-2-KK;MAMU-A*11	MOUSE	100	89	100	100	100	100	100	67	67	67	78	67	67	67	78	78	78
128	IGSIRNGTYDHDVYRDEALNNRFQIKGVE LKSGYKD	H3N2	HA		HUMAN	22	42	42	42	42	42	44	100	100	22	94	100	100	100	100	44	44
129	IHHPSTNQEQTSLYVQAS	H3N2	HA	H-2-IAD	MOUSE	50	33	39	33	33	33	50	72	67	89	94	67	61	61	89	44	50
130	ILAIYSTVASSL	H1N1	HA	H-2-KD	MOUSE	42	100	100	100	100	100	100	42	42	42	42	42	42	42	42	92	92
131	ILRGSVAHK	H3N2	NP	HLA-A3	HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
132	ILTGNSSLCPIRGWAIYSKDN	H1N1	NA	H-2-IED	MOUSE	95	76	100	86	86	90	95	38	38	38	38	38	38	38	38	86	81
133	INDRNFWRGENGRKT	H3N2	NP	H-2-B CLASS II	MOUSE	93	100	100	100	100	100	93	100	100	100	100	100	100	100	100	93	93
134	INSNGNLIAPRGYFK	H3N2	HA	H-2-B CLASS II	MOUSE	47	53	60	53	53	53	47	100	93	100	100	93	93	93	100	53	53
135	INSNGNLIAPRGYFKMRTGKSS	H3N2	HA	H-2-IEK	MOUSE	41	45	50	45	45	45	41	95	91	100	100	91	86	86	100	41	41
136	IRGWAIYSKDNSIRI	H1N1	NA	H-2-IED	MOUSE	87	87	100	93	93	100	87	67	67	67	67	67	67	67	67	93	87
137	IRPNENPAHKSQLVW	H3N2	NP	H-2-B CLASS II	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	93	100
138	ISPLMVAYM	H2N2;H3N2	PB2	H-2-D CLASS I;H2-DB;H2-KB;MAMU-A*01	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	89
139	ITYGACPKYVKQNTL	H3N2	HA		MOUSE	53	53	53	53	53	53	53	100	93	100	100	93	93	93	100	67	67
140	ITYSSSMMWEINGPE	H3N2	PB2		MOUSE	93	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
141	IWHSNLNDTTYQRT	H3N2	NP		MOUSE	93	100	93	100	100	100	93	100	100	100	100	100	93	100	100	93	93
142	IYATVAGSL	H2N2	HA	H-2-KD	MOUSE	56	78	78	78	78	78	78	44	44	44	44	44	44	44	44	78	78
143	IYSTVASSL	H1N1	HA	H-2-D CLASS I;H2-KD	MOUSE	44	100	100	100	100	100	100	56	56	56	56	56	56	56	56	100	100

Table 7. Conservancy analysis of T cell linear epitope sequences

No.	Sequence	Influenza Source Subtype	Source Protein	MHC Restriction Allele(s)	Host Species	H1N1							H3N2								H5N1	
						A/Brevig Mission/1/18	A/New Caledonia/20/99	A/PR/8/34	A/Taiwan/1/86	A/Texas/36/91	A/USSR/90/77	A/WS/33	A/Bangkok/1/79	A/Beijing/353/89	A/England/42/72	A/Hong Kong/1/68	A/Leningrad/360/86	A/New York/5/2004	A/Panama/2007/99	A/UDORN/307/72	A/Hong Kong/156/97	A/Viet Nam/1194/2004
144	IYSTVASSLVL	H1N1	HA	H-2-KD;HLA-B37	MOUSE	36	100	100	100	100	100	100	55	55	55	55	55	55	55	55	91	91
145	IYWTIVKPGDILLINS	H3N2	HA	HLA-DRB1*0101;HLA-DRB1*0701	HUMAN	38	38	44	38	38	38	38	100	100	94	88	100	100	100	94	44	44
146	IYWTIVKPGDILLINSTGNLIAPRGYFKIRN	H3N2	HA		HUMAN	45	42	48	42	42	42	45	94	97	87	84	97	97	97	87	45	45
147	KEIGNGCFEF	H1N1	HA	H-2-DB;MAMU-A*11		40	100	100	100	100	100	100	50	50	40	50	50	50	50	50	90	90
148	KEVNARIEPFLKTTP	H3N2	PA		MOUSE	100	100	100	100	100	100	93	93	93	100	93	93	87	93	100	100	100
149	KFLLMDALKLSIEDP	H3N2	PA		MOUSE	100	93	100	100	93	100	100	100	100	100	100	100	100	100	100	100	100
150	KGEIRRIWRQANNG	H3N2	NP		HUMAN	93	93	93	93	93	93	93	93	93	93	93	93	93	93	93	93	93
151	KGILGFVFTLTV	H1N1;H2N2	M1	HLA-A*02;HLA-A*0201;HLA-A2	HUMAN;MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
152	KGVELKSGYKDWILWISFAISCFLLCVVLLGFIM	H3N2	HA		HUMAN	24	35	35	35	35	35	35	100	100	26	100	100	97	100	100	24	24
153	KIDLWSYNAELLVALE	H3N2	HA	HLA-DRB1*0101;HLA-DRB1*0701;HLA-DRB1*1302;HLA-DRB1*1501	HUMAN	31	69	69	69	69	69	69	100	100	31	100	100	94	100	100	63	63
154	KIDLWSYNAELLVALENQHTI	H3N2	HA	HLA-DR	HUMAN	24	62	62	62	62	62	62	100	100	29	100	95	95	100	100	57	57
155	KIYHKCDNACIGSIRN	H3N2	HA	HLA-DRB1*0701	HUMAN	38	50	63	50	50	50	63	100	100	31	94	100	100	100	100	56	63
156	KLKNSYVNKKGK	H1N1	HA	H-2-IED	MOUSE	75	58	100	58	58	58	83	42	50	42	42	50	50	50	42	42	42
157	KLSTRGVQIASNEN	H1N1	NP		HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	93	93
158	KQNTLKLATGMRNVP	H3N2	HA		MOUSE	47	40	40	40	40	40	40	100	100	100	100	100	100	100	100	67	67
159	KQYDSDEPELRSLAS	H3N2	PA		MOUSE	100	87	100	93	93	93	100	93	93	93	93	93	93	93	93	93	87
160	KRMGVQMQR	H7N1	M1	HLA-DP	HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
161	KRYGPALSI	H3N2	PB2	HLA-B*27052/KB	MOUSE	89	89	89	89	89	89	89	89	89	89	89	89	89	89	89	100	100
162	KTGGPIYKR	H3N2	NP	HLA-A68;HLA-AW68	HUMAN	89	100	89	100	100	100	89	89	100	100	100	100	100	89	100	89	89
163	KTGGPIYRRVNGKWM	H3N2	NP		MOUSE	87	80	100	87	80	87	87	93	87	87	87	87	80	93	87	80	80
164	KYVKQNTLKL	H3N2	HA	H-2-IED	MOUSE	50	50	50	50	50	50	50	100	90	100	100	90	90	90	100	70	70
165	LEFINEDFNWTGVAQDGGSYACKRGSVNSF	H3N2	HA		HUMAN	23	23	23	23	23	23	23	87	90	73	67	87	67	73	67	27	27
166	LEFITEGFTWTGVTQNGGSNA	H3N2	HA	H-2-IAK	MOUSE	29	29	29	29	29	29	29	81	67	95	100	76	71	67	95	29	29
167	LEFITEGFTWTGVTQNGGSNAC	H3N2	HA	H-2-IAK	MOUSE	27	32	27	27	32	27	27	82	68	95	100	77	73	68	95	27	27
168	LELRSRYWA		NP	HLA-B44		100	100	100	100	100	100	100	100	100	100	100	100	89	89	100	100	100
169	LELRSRYWAI		NP	HLA-B44		100	100	100	100	100	100	100	100	100	100	100	100	90	90	100	100	100
170	LELRSRYWAIRTRSGGNTNQQRAS	H3N2	NP	H-2-D CLASS II;HLA-DR4	HUMAN;MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	96	96	100	100	100
171	LENFRAYVDGFEPNG	H3N2	PA		MOUSE	100	93	100	100	100	100	100	100	100	100	100	100	100	100	100	93	100
172	LENLQAYQKR	H7N1	M1	HLA-DQ	HUMAN	90	100	100	100	100	100	100	90	90	100	100	90	90	90	100	90	100
173	LGLDIETATRAGKQIVERI	H1N1	NS1	H-2-KD	MOUSE	100	84	100	79	84	84	100	79	74	74	84	74	74	74	74	84	100
174	LIDALLGDP	H3N2	HA	H-2-IAD	MOUSE	44	44	56	44	44	44	56	100	100	100	100	100	100	100	100	44	44
175	LIEKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKI	H3N2	HA		HUMAN	26	40	40	40	40	40	40	97	100	26	97	100	97	100	97	37	37
176	LKGKFQTAAQRAMMDQVRES	H3N2	NP		HUMAN	100	95	95	100	100	100	90	100	100	100	100	100	95	95	100	100	100
177	LKLATGMRNVPEKQT	H3N2	HA		MOUSE	53	47	47	47	47	47	47	100	100	100	100	100	100	100	100	53	53
178	LKTTPRPLRLPNGPP	H3N2	PA		MOUSE	93	87	100	87	87	87	87	80	87	80	87	80	80	87	80	93	93
179	LKYNGIITETI	H1N1	NA	H-2-IAD;H-2-KD	MOUSE	91	100	100	100	100	100	100	36	36	36	36	36	45	36	36	91	91
180	LLCVVLLGFIMWACQKGNIRCNICI	H3N2	HA		HUMAN	24	40	40	40	40	40	40	100	100	24	96	100	96	100	100	36	24
181	LLQNSQVYSLIRPNE	H3N2	NP		MOUSE	100	93	100	93	93	93	100	100	100	100	100	100	100	100	100	87	93
182	LPACVYGPAVASGYD	H3N2	NP	H-2-B CLASS II	MOUSE	100	100	100	93	93	93	93	93	93	100	100	93	93	87	100	93	93
183	LPFDKPTIM	H3N2	NP	HLA-B*3501	HUMAN	67	89	78	89	89	89	89	78	67	89	100	67	67	67	89	67	67
184	LPFDKSTIM	H3N2	NP	HLA-B*3501	HUMAN	67	89	78	89	89	89	78	89	78	100	89	78	78	78	100	67	67
185	LPFDRTTVM	H1N1	NP	HLA-B7		67	78	100	78	78	78	78	78	67	67	67	67	67	67	67	67	67
186	LPFEKSTVM	H3N2	NP	HLA-B*3501	HUMAN	67	67	67	67	67	67	56	89	100	78	67	100	100	100	78	67	67
187	LPRRSGAAGAAVKG	H3N2	NP		HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	93	100
188	LRGSVAHKSCLPACV	H3N2	NP		MOUSE	100	100	100	100	93	100	100	100	100	100	100	100	100	100	100	100	100
189	LRSRYWAI	H3N2	NP	HLA-B*2702	HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	88	88	100	100	100
190	LRVLSFIRGTKVSPRGKLSTRG	H3N2	NP		HUMAN	86	86	91	86	86	91	91	95	95	95	100	95	95	95	95	82	82
191	LSLRNPILV	H1N1	PB1	H-2-DB;H-2-KB	MOUSE	56	44	44	44	44	44	44	44	44	44	44	44	44	44	44	56	56
192	LSQMSKEVNARIEPF	H3N2	PA		MOUSE	100	100	100	100	100	100	100	93	93	100	93	93	93	93	100	100	100
193	LSSRISIYWTIVKPGDVLVI	H3N2	HA	H-2-IEK	MOUSE	35	40	40	40	40	40	35	85	90	95	100	90	85	85	95	40	40
194	LSWKQVLAELQDIEN	H3N2	PA		MOUSE	93	87	100	93	93	100	100	100	93	100	100	93	87	93	100	93	93
195	LTKGILGFVFTLTVPSERG	H2N2	M1	HLA-A2.1	HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
196	LVKTITNDQIEVTNATELVQSSSTGRICDSPHRIL	H3N2	HA		HUMAN	23	23	23	23	23	23	23	100	100	91	89	100	91	97	91	26	26
197	LYEKVKSQL	H1N1	HA	H-2-KD	MOUSE	44	100	100	100	100	100	100	56	56	44	56	56	56	56	56	67	67
198	LYIWGIHHPSTNQEQTSLYVQAS	H3N2	HA	H-2-IAD;H-2-IED	MOUSE	52	39	43	39	39	39	52	78	74	91	96	74	70	70	91	48	52
199	LYIWGVHHPSTNQEQTSLYVQAS	H3N2	HA	H-2-IAD	MOUSE	57	43	39	43	43	43	57	83	78	96	100	78	74	74	96	52	48
200	LYQNVGTYV	H2N2	HA	H-2-KD	MOUSE	67	56	56	56	56	56	56	56	56	56	56	56	56	56	56	67	67
201	LYQNVGTYVS	H2N2	HA	H-2-KD		70	50	60	50	50	50	60	50	50	50	50	50	50	50	50	70	70
202	MELVRMIKRGINDRN	H3N2	NP		MOUSE	93	87	100	87	87	87	93	93	93	93	93	93	93	87	93	93	93
203	METMESSTLELRSRY	H3N2	NP		MOUSE	93	73	100	87	73	87	100	93	87	93	87	87	73	73	93	80	80
204	MGLIYNRM	H1N1;H2N2	M1	H-2-KB	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
205	MIKRGINDRNFWRGE	H3N2	NP	H-2-B CLASS II	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	93	100	100	100
206	MLIIW	H2N2	HA	H-2-M3	MOUSE	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60
207	MMIWHSNLNDATYQR	H3N2	NP	H-2-B CLASS II	MOUSE	100	87	100	87	87	93	100	93	93	93	93	93	100	93	93	100	93
208	MRTFFGWKEPNVVKP	H3N2	PA		MOUSE	100	93	100	93	93	93	100	87	80	87	87	87	73	73	87	73	87
209	MRTGKSSIMRSDAPI	H3N2	HA	H-2-B CLASS II	MOUSE	47	40	40	40	40	40	40	93	93	100	100	93	87	87	100	40	40
210	MVLSAFDERRNKYLE	H3N2	NP		MOUSE	100	100	100	100	100	100	100	93	93	93	100	93	93	93	93	93	93
211	NACIESIRNGTYDHD	H3N2	HA		MOUSE	33	60	67	60	60	60	67	93	93	33	93	93	93	93	93	60	67
212	NAYVSVVTSNYNRRF	H1N1	HA	H-2-IAD	MOUSE	73	80	100	87	80	93	87	33	33	33	33	33	33	33	33	47	47
213	NCTLIDALLGDPH	H3N2	HA	H-2-IAD	MOUSE	38	38	38	38	38	38	46	100	100	92	92	100	92	100	92	38	38
214	NDKPFQNVNRITYGAC	H3N2	HA	HLA-DR	HUMAN	44	50	44	50	50	44	44	94	100	94	94	94	100	100	94	38	38
215	NETWDLFVERSKAFSNC	H3N2	HA	H-2-IAD	MOUSE	29	29	29	29	29	29	29	94	82	100	100	88	82	82	100	29	29
216	NFSVIFDR	H1N1	NS1		MOUSE	100	100	100	100	100	100	100	100	100	88	100	100	100	88	100	88	100
217	NGRKTRIAYERMCNI	H3N2	NP		MOUSE	93	100	100	100	100	100	93	93	93	93	93	93	93	93	93	93	93
218	NGYIEGKL	H1N1	PA	H-2-KB		100	100	100	100	100	100	100	88	88	88	100	88	88	88	88	88	88

Table 7. Conservancy analysis of T cell linear epitope sequences

No.	Sequence	Influenza Source Subtype	Source Protein	MHC Restriction Allele(s)	Host Species	H1N1							H3N2								H5N1	
						A/Brevig Mission/1/18	A/New Caledonia/20/99	A/PR/8/34	A/Taiwan/1/86	A/Texas/36/91	A/USSR/90/77	A/WS/33	A/Bangkok/1/79	A/Beijing/353/89	A/England/42/72	A/Hong Kong/1/68	A/Leningrad/360/86	A/New York/5/2004	A/Panama/2007/99	A/UDORN/307/72	A/Hong Kong/156/97	A/Viet Nam/1194/2004
219	NNPHRILDGIDC	H3N2	HA	H-2-IAK	MOUSE	42	42	42	42	42	42	42	67	67	92	100	67	58	58	100	42	42
220	NPAHKSQLVWMACHS	H3N2	NP		MOUSE	100	93	100	93	93	93	100	100	100	93	93	100	100	100	93	100	100
221	NSEGTGQAADLKSTQAAIDQINGKLNRLI EKTNEKFH	H3N2	HA		HUMAN	22	54	54	54	54	54	54	97	100	22	97	100	95	97	97	43	49
222	NSNGNLIAPRGYFKMRTGKS	H3N2	HA	H-2-IEK	MOUSE	40	45	50	45	45	45	40	95	90	100	100	90	85	85	100	40	40
223	NVKNLYEKVK	H1N1	HA	HLA-A*11		40	100	100	100	100	100	100	40	40	40	40	40	40	40	40	80	80
224	NVVKPHEKGINPNYL	H3N2	PA		MOUSE	100	87	100	87	87	93	100	80	80	87	87	80	80	80	87	87	93
225	NYFTSEVSHCRATEY	H3N2	PA		MOUSE	93	93	100	93	93	93	93	93	93	93	93	93	93	93	93	93	93
226	PIGTCSSECITPNGSIPNDKPFQNVNRITY GAC	H3N2	HA		HUMAN	42	45	39	42	42	39	42	97	100	94	91	97	94	94	94	39	39
227	PKKTGGPIYKRVD	H3N2	NP		HUMAN	85	100	85	100	100	100	92	92	92	100	100	92	100	92	100	85	85
228	PKYVKQNTLKLA		HA	HLA-DR1	HUMAN	50	42	42	42	42	42	42	100	92	100	100	92	92	92	100	75	75
229	PKYVKQNTLKLAT	H3N2	HA	H-2-IED;HLA-DR1;HLA-DR4;HLA-DR7;HLA-DRB1*0101;HLA-DRB1*010101;HLA-DRB1*0102;HLA-DRB1*0301;HLA-DRB1*0302;HLA-DRB1*0401;HLA-DRB1*0402;HLA-DRB1*040301;HLA-DRB1*0405;HLA-DRB1*07;HLA-DRB1*0701;HLA-DRB1*0901;HLA-DRB1*1101;HLA-DRB1*1102;HLA-DRB1*1103;HLA-DRB1*11	HUMAN;MOUSE	54	46	46	46	46	46	46	100	92	100	100	92	92	92	100	77	77
230	PKYVKQNTLKLATG	H3N2	HA		MOUSE	57	50	50	50	50	50	50	100	93	100	100	93	93	93	100	79	79
231	PKYVRSAKLRMVT		HA	H-2-IED		85	100	100	92	92	92	92	46	46	46	46	46	46	46	46	54	54
232	PLKAEIAQRLEDV	H7N7	M1	HLA-DRB1*0101;HLA-DRB1*010201	HUMAN	100	92	100	100	100	100	100	100	100	100	100	100	100	100	100	100	92
233	PNENPAHKSQLVWMACNS	H3N2	NP		HUMAN	94	100	94	100	100	100	94	94	94	100	100	94	94	94	100	89	94
234	PNFSSLENFRAYVDG	H3N2	PA	H-2-DB	MOUSE	100	87	100	93	93	93	100	93	93	93	93	93	93	93	93	100	100
235	PNGPPCSQRSKFLLM	H3N2	PA		MOUSE	93	87	100	87	87	93	93	87	93	87	87	87	93	93	87	93	93
236	PNGYIEGK	H1N1	PA		MOUSE	100	100	100	100	100	100	100	88	88	88	100	88	88	88	88	88	88
237	PRMCSLMQGSTLPRR	H3N2	NP	H-2-B CLASS I	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
238	PSFDMSNEGSYFFGDNAEEYDN	H3N2	NP		HUMAN	100	100	100	100	100	100	100	100	100	100	95	100	32	95	100	100	32
239	PSTNQEQTSLYVQAS	H3N2	HA	H-2-B CLASS II	MOUSE	47	33	33	33	33	33	47	73	67	93	100	67	60	60	93	40	40
240	QDLEKYVEDTKIDLWS	H3N2	HA		HUMAN	31	38	38	38	38	38	38	100	100	31	100	100	100	100	100	38	38
241	QDLEKYVEDTKIDLWSYNAELLVALENQ HTIDLTDS	H3N2	HA		HUMAN	25	56	56	56	56	56	53	100	100	25	100	97	97	100	100	53	53
242	QKLPGNDNSTATLCLGHHAVPNGTLVKT ITNDQIE	H3N2	HA		HUMAN	23	23	23	23	23	23	23	97	100	97	94	100	97	97	94	23	23
243	QMVQAMITIGTHPSS	H3N2	M1		MOUSE	93	87	93	87	87	87	93	87	87	87	87	87	87	87	87	93	87
244	QNVNRITYGACPRYVKQNTLKLATGMRN VPEKQT	H3N2	HA		HUMAN	47	47	44	47	47	44	44	94	100	94	94	97	100	100	94	50	50
245	QVYSLIRPNENPAHK	H3N2	NP	H-2-B CLASS II	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	87	93
246	RASVGKMIDGIGRFY	H3N2	NP		HUMAN	87	87	93	93	93	100	100	100	100	100	93	100	93	100	100	80	80
247	RATEYIMKGVYINTA	H3N2	PA	H-2-B CLASS II	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
248	RELILYDKEEIRRIW	H3N2	NP		MOUSE	100	93	100	93	93	93	100	93	93	93	93	93	93	93	93	100	100
249	RENAEDMGNGCFKIYHKCDNACIGSIRN GTYDH	H3N2	HA		HUMAN	21	58	64	58	58	58	64	100	100	27	94	100	100	100	100	64	67
250	RFYIQMCTEL	H1N1	NP	H-2-D CLASS I	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
251	RGEETIEERFEITGT	H3N2	PA		MOUSE	100	100	100	100	100	100	100	93	87	100	100	80	87	87	100	100	93
252	RGLQRRRFVQNALNGNG	H1N1	M1	HLA-A2	HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
253	RGVQIASNENMETME	H3N2	NP	H-2-DB	MOUSE	93	73	100	87	73	87	100	93	87	93	87	87	80	80	93	80	87
254	RIAYERMCNILKGKF	H3N2	NP	H-2-B CLASS I	MOUSE	100	100	100	100	100	100	100	93	93	93	93	93	93	93	93	100	100
255	RIEPFLKTTPRPLRL	H3N2	PA		MOUSE	100	93	100	93	93	93	93	80	80	87	87	80	73	80	87	100	100
256	RKLKREITF	H3N2	M1	SAOE-G*02;SAOE-G*12	COTTON-TOP TAMARIN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	78	89
257	RLEDVFAGK		M1	HLA-A1;HLA-A3	HUMAN	100	89	100	100	100	100	100	100	100	100	100	100	100	100	100	100	89
258	RLIQNSLTI	H1N1;H2N2;H3N2	NP	H-2-DB;HLA-A*0201	MOUSE	89	100	100	100	100	100	100	100	100	100	100	100	100	100	100	89	89
259	RLIQNSLTIERMVLS	H2N2	NP		MOUSE	93	100	100	100	100	100	100	100	100	100	100	100	93	93	100	93	93
260	RLIQNSLTIERMVLSAFDERRNK	H3N2	NP	H-2-D CLASS II	MOUSE	96	100	100	100	100	100	100	96	96	96	100	96	91	91	96	91	91
261	RMCNILKGKFQTAAQ	H3N2	NP		MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
262	RNGPMTNTVHYPKIY	H3N2	PB2		MOUSE	80	73	100	87	73	87	87	87	87	87	87	87	80	80	80	80	73
263	RNLPFDRTTIMAAFN	H3N2	NP		MOUSE	80	87	100	87	87	87	87	73	67	80	80	67	67	67	80	80	80
264	RPLRLPNGPPCSQRS	H3N2	PA		MOUSE	93	80	100	80	80	87	93	73	80	73	80	73	80	80	73	93	93
265	RRATAILRK	H3N2	PB2	HLA-B*27052/KB	MOUSE	100	89	100	89	89	89	100	89	89	89	89	89	89	89	89	100	100
266	RRSFEIKKL	H3N2	PB1	HLA-B*27052/KB	MOUSE	100	100	100	100	100	100	100	89	89	89	89	89	89	89	89	89	89
267	RRSGAAGAAVK	H1N1;H3N2	NP	HLA-B27	HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	91	100
268	RSALILRGSVAHKSC	H3N2	NP		MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
269	RSAYERMCNILKGK	H2N2	NP		MOUSE	93	93	93	93	93	93	93	100	100	93	100	100	100	100	93	93	93

Table 7. Conservancy analysis of T cell linear epitope sequences

No.	Sequence	Influenza Source Subtype	Source Protein	MHC Restriction Allele(s)	Host Species	H1N1							H3N2								H5N1	
						A/Brevig Mission/1/18	A/New Caledonia/20/99	A/PR/8/34	A/Taiwan/1/86	A/Texas/36/91	A/USSR/90/77	A/WS/33	A/Bangkok/1/79	A/Beijing/353/89	A/England/42/72	A/Hong Kong/1/68	A/Leningrad/360/86	A/New York/5/2004	A/Panama/2007/99	A/UDORN/307/72	A/Hong Kong/156/97	A/Viet Nam/1194/2004
270	RSDAPIDTCISECITPNGSI	H3N2	HA	H-2-IAK	MOUSE	50	50	40	45	45	45	45	90	90	95	100	90	85	85	95	35	35
271	RSLASWIQNEFNKAC	H3N2	PA		MOUSE	93	87	100	93	93	93	100	93	93	93	93	93	93	93	93	100	93
272	RSQQTVTPNIGSRPWVRGQSSRISIYWTIVKPGDIL	H3N2	HA		HUMAN	42	44	44	47	47	47	42	89	94	89	86	92	89	92	92	44	42
273	RSYLIRAL	H1N1;H2N2	PB1	H-2-KB	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	88	88	100	88	88
274	RTFSFQLI	H1N1;H3N2	NS2	H-2-KB;MAMU-A*02	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	88
275	RTGMDPRMCSLMQGS	H3N2	NP		MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
276	RTLYQNVGTYV	H2N2	HA	H-2-KD	MOUSE	55	45	45	45	45	45	45	45	45	45	45	45	45	45	45	55	55
277	RTLYQNVGTYVSVGTSTLNK	H2N2	HA	H-2-D CLASS I;H2-KD	MOUSE	55	40	50	40	35	40	45	30	30	30	30	30	30	30	30	70	70
278	RVDGKWMRELVL	H3N2	NP		HUMAN	83	92	83	100	92	100	83	100	83	100	100	83	92	100	100	75	75
279	RYWAIRTR	H3N2	NP	HLA-B*2705	HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	88	88	100	100	100
280	SAAFEDLR	H3N2	NP	HLA-DQW5	HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
281	SAAFEDLRVLSFIKG	H1N1	NP	HLA-B37	HUMAN	87	87	100	87	87	87	87	87	87	87	93	87	87	87	87	87	87
282	SAAFEDLRVLSFIRG	H3N2	NP	HLA-B37	HUMAN	93	93	93	93	93	93	93	93	93	93	100	93	93	93	93	93	93
283	SCLENFRAYV	H3N2	PA	H-2-D CLASS I;H2-DB	MOUSE	90	90	90	100	100	100	90	100	100	100	100	100	100	100	100	90	90
284	SDMRAEIIRMMEG	H3N2	NP		MOUSE	85	85	85	85	85	85	77	100	100	100	100	100	100	100	100	85	85
285	SDRVMVSPLAVTWWN	H3N2	PB2		MOUSE	100	93	100	93	93	100	100	100	100	100	100	100	100	100	100	100	100
286	SDYEGRLI	H1N1;H2N3	NP	H-2-K CLASS I;H2-KK;MAMU-A*11	MOUSE	100	88	100	88	88	100	100	100	100	100	100	100	100	100	100	88	100
287	SDYEGRLIQNSLTI	H1N1;H3N2	NP	H-2-DB;H-2-KD;H-2-KK;HLA-B37	MOUSE	93	93	100	93	93	100	100	100	100	100	100	100	100	100	100	86	93
288	SETQGTEKLTITYSS	H3N2	PB2		MOUSE	100	100	100	100	100	100	100	93	93	93	100	93	93	93	93	93	100
289	SFERFEIFPKE	H1N1	HA	H-2-IED	MOUSE	82	100	100	100	100	100	91	36	36	45	45	36	36	36	45	45	45
290	SFYRNVVWLIKK	H5N9	HA	H-2-D CLASS I	MOUSE	75	67	67	67	75	67	75	33	33	42	33	33	33	33	33	92	92
291	SGPDNGAVAV	H1N1	NA	H-2-DB		100	100	100	90	100	90	90	40	40	40	50	40	40	40	40	100	100
292	SGPLKAEIAQRLE	H1N1	M1	HLA-DR1;HLA-DR7 DW11;HLA-DRB1*0304;HLA-DRB1*0405;HLA-DRB1*0701;HLA-DRB1*0802;HLA-DRB1*1101;HLA-DRB1*1401;HLA-DRB1*1402	HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	92
293	SGPLKAEIAQRLEDV	H1N1	M1	HLA-B37;HLA-DR1	HUMAN	100	93	100	100	100	100	100	100	100	100	100	100	100	100	100	100	93
294	SGSFVQHPELTGL	H1N1	NA		MOUSE	100	100	100	100	100	100	100	38	38	38	38	38	38	38	38	92	100
295	SIRNGTYDHDVYRDE	H3N2	HA		MOUSE	33	53	60	53	53	53	60	100	100	33	93	100	100	100	100	53	60
296	SKAFSNCYPYDVPDYASL	H3N2	HA		MOUSE	39	33	39	33	33	33	39	100	94	94	100	100	94	94	100	44	39
297	SLVGIDPFKLLQNSQVYSLIRP	H3N2	NP		HUMAN	95	91	95	95	95	95	95	100	100	100	100	100	100	100	100	91	91
298	SMIEAESSVKEKDMT	H3N2	PA		MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	93	100	100	93	100
299	SNEGSYFF	H3N2	NP	SAOE-G*08	COTTON-TOP TAMARIN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	88
300	SNLNDATYQRTRALV	H3N2	NP		MOUSE	100	93	100	93	93	93	100	93	93	93	93	93	100	93	93	100	100
301	SQLVVMACHSAAFED	H3N2	NP		MOUSE	100	93	100	93	93	93	100	100	100	93	93	100	100	100	93	100	100
302	SRYWAI		NP	HLA-B8		100	100	100	100	100	100	100	100	100	100	100	100	83	83	100	100	100
303	SRYWAIRTR	H1N1;H3N2	NP	HLA-B*08;HLA-B*2703;HLA-B*2705;HLA-B*27052/KB;HLA-B27	HUMAN;MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	89	89	100	100	100
304	SSFSFGGFTFKRTSG	H3N2	PB2		MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	93	100
305	SSISFCGV	H1N1	NA	H-2-KB		100	100	100	100	100	100	88	63	63	63	63	63	63	63	63	100	100
306	SSLENFRAYV	H1N1;H3N2;H3N8	PA	H-2-DB	MOUSE	100	80	100	90	90	90	100	90	90	90	90	90	90	90	90	100	100
307	SSYRRPVGI	H1N1;H2N2;H3N2	PB1	H-2-KB	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
308	STNQEQTSLYVQASGRVTVS	H3N2	HA	H-2-B CLASS II;H-2-D CLASS II;H-2-IAD	MOUSE	40	30	30	25	30	25	40	80	75	95	100	75	65	70	95	30	30
309	SVQRNLPFDKPTIMAAFTGNTEG	H2N2;H3N2	NP		MOUSE	87	96	87	91	91	91	96	91	87	96	100	87	87	87	96	83	87
310	SVSSFERFEIFPK	H1N1	HA	H-2-IED	MOUSE	92	100	100	100	100	100	92	38	38	38	38	38	38	38	38	38	38
311	SYNAELLVAL	H3N2	HA	H-2-D CLASS I	MOUSE	40	80	80	80	80	80	80	100	100	40	100	100	90	100	100	70	70
312	SYNAELLVALENQHTI	H3N2	HA		HUMAN	31	69	69	69	69	69	69	100	100	31	100	94	94	100	100	63	63
313	TALANTIEV	H1N1;H2N2;H3N2	PB1	H-2-D CLASS I;H2-DB	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	89
314	TELKLSDYEGRLIQNS	H3N2	NP		MOUSE	100	94	100	94	94	100	100	100	100	100	100	100	100	100	100	94	100
315	TELKLSDYEGRLIQNSLTIER	H3N2	NP	H-2-D CLASS II	MOUSE	95	95	100	95	95	100	100	100	100	100	100	100	95	95	100	90	95
316	TETIKSWRKKILRTQ	H1N1	NA		MOUSE	80	87	100	93	87	93	93	60	60	53	53	60	67	67	53	80	80
317	TGKICNNPHRILDGIDCTLI	H3N2	HA	H-2-IAD	MOUSE	30	30	30	30	30	30	30	75	75	95	100	75	70	70	100	30	30
318	TGKICNNPHRILDGIDCTLID	H3N2	HA	H-2-IAK	MOUSE	29	29	29	29	29	29	29	76	76	95	100	76	71	71	100	29	29
319	TIMAAFNGNTEGRTS	H3N2	NP		MOUSE	93	93	100	87	87	87	93	87	87	93	93	87	87	87	93	93	93
320	TKGILGFVFTLTV	H1N1	M1	HLA-A2	HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
321	TLIDALLG	H3N2	HA	H-2-IAD	MOUSE	50	50	50	50	50	50	50	100	100	100	100	100	100	100	100	50	50
322	TNEKFHQIEKEFSEVE	H3N2	HA		MOUSE	38	38	38	38	38	38	38	100	100	38	100	100	100	100	100	38	38
323	TNTVHYPKIYKTYFE	H3N2	PB2		MOUSE	87	80	100	93	80	93	93	87	87	80	87	87	80	80	80	87	80
324	TRALVRTGMDPRMCS	H3N2	NP	H-2-B CLASS I	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
325	TREILTKTTVDHMAI	H3N2	PB2		MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
326	TRRSQQTIIPNIGSRPWVRGLS	H3N2	HA	H-2-IAD	MOUSE	32	32	32	32	32	32	32	95	91	91	95	86	77	82	95	41	36
327	TSLYVRASGRVTVSTK	H3N2	HA	HLA-DR;HLA-DRB1*0101;HLA-DRB1*0401	HUMAN	38	38	38	38	38	38	38	94	94	94	88	94	75	75	94	38	38

Table 7. Conservancy analysis of T cell linear epitope sequences

No.	Sequence	Influenza Source Subtype	Source Protein	MHC Restriction Allele(s)	Host Species	H1N1							H3N2								H5N1	
						A/Brevig Mission/1/18	A/New Caledonia/20/99	A/PR/8/34	A/Taiwan/1/86	A/Texas/36/91	A/USSR/90/77	A/WS/33	A/Bangkok/1/79	A/Beijing/353/89	A/England/42/72	A/Hong Kong/1/68	A/Leningrad/360/86	A/New York/5/2004	A/Panama/2007/99	A/UDORN/307/72	A/Hong Kong/156/97	A/Viet Nam/1194/2004
328	TYQRTRALV	H1N1;H2N3;H3N2	NP	H-2-D CLASS I;H2-DB;H-2-KD;PATR-A*0901	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
329	TYQRTRALVRTG	H1N1;H3N2	NP	H-2-D CLASS I;H2-KD	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
330	TYQRTRALVRTGMD	H3N2	NP	H-2-KD	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
331	TYQRTRALVRTGMDP	H1N1;H2N2;H3N2	NP	H-2-KD	HUMAN;MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
332	TYVSVGTST	H2N2	HA	H-2-KD	MOUSE	67	56	67	56	56	56	56	44	44	44	44	44	44	44	44	89	89
333	TYVSVGTSTL	H2N2	HA	H-2-KD	MOUSE	60	50	60	50	50	50	50	40	40	40	40	40	40	40	40	90	90
334	VAHKSCLPACVYGP	H2N2	NP	H-2-S CLASS II	MOUSE	100	100	100	100	93	100	93	100	100	100	100	100	100	100	100	93	93
335	VETPIRNEW	H1N1	M2	HLA-B27;HLA-B44	HUMAN	89	67	100	67	67	100	100	67	67	100	100	67	100	67	67	67	89
336	VFPNEVGARILTSE	H1N1	PB2	H-2-D CLASS I	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
337	VGLISLILQI	H1N1	NA		MOUSE	90	70	100	80	80	80	90	50	50	50	50	50	50	50	50	80	60
338	VKILPKDRWTQH	H2N2	HA	HLA-DRW11	HUMAN	33	33	42	33	33	33	42	42	42	42	42	42	42	42	42	42	42
339	VLAELQDIENEEKIP	H3N2	PA		MOUSE	100	87	100	93	93	100	100	100	93	100	100	93	87	93	100	93	100
340	VLMEWLKTRPILSPLTKGIL	H1N1	M1	HLA-DPB1*0401	HUMAN	95	95	100	95	95	95	100	95	95	95	95	95	95	95	95	95	95
341	VLRGFLIL	H1N1	PB2	H-2-KB		100	100	100	100	100	100	100	100	100	100	100	100	88	88	100	100	100
342	VLWGIHHPPNSK	H1N1	HA	H-2-IED	MOUSE	67	75	92	67	67	67	67	50	42	42	42	50	42	42	42	58	67
343	VPNGTLVKTITNDQIEVTNAT	H3N2	HA	HLA-DR	HUMAN	29	29	29	29	29	29	29	100	100	100	95	100	95	95	100	29	29
344	VRESRNPGNAEIEDLIFLARS	H3N2	NP	H-2-D CLASS II	MOUSE	100	95	90	95	95	100	95	95	95	100	100	95	100	100	100	100	100
345	VSDGGPNLY	H1N1;H2N2	PB1	HLA-A*0101;HLA-A1;MAMU-A*02	HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
346	VSPLAVTWWNRNGPM	H3N2	PB2		MOUSE	93	87	100	87	87	93	93	93	93	100	100	93	93	93	93	93	93
347	VTGLRNIPS	H1N1	HA		MOUSE	89	100	100	100	100	100	100	56	56	56	56	56	56	56	56	67	67
348	VTGLRNIPSI	H1N1	HA	H-2-IED;MAMU-A*02		90	100	100	100	100	100	100	50	50	50	50	50	50	50	50	60	60
349	VTGLRNIPSIQS	H1N1	HA	H-2-IED	MOUSE	92	100	100	100	100	100	92	50	50	50	50	50	50	50	50	50	50
350	VTGLRNIPSIQSR	H1N1	HA	H-2-IED	MOUSE	92	100	100	100	100	100	92	54	54	54	54	54	54	54	54	54	54
351	VTQNGGSNACKRGPG	H3N2	HA		MOUSE	33	33	33	33	33	33	33	73	60	93	100	67	67	60	93	33	33
352	VTVSTKRSQQTVTPNI	H3N2	HA		HUMAN	38	31	38	38	38	38	38	88	94	81	81	94	88	94	88	31	31
353	VTWWNRNGPMTNTVH	H3N2	PB2		MOUSE	80	73	100	87	73	87	87	93	93	93	93	93	87	87	87	87	80
354	VYQILAIYATVAGSLSLAIMMAG	H2N2	HA	H-2-D CLASS I;H2-KD	MOUSE	30	61	57	61	61	61	61	26	26	26	26	26	26	26	26	74	74
355	WGIHHPSTNQEQTSL	H3N2	HA		MOUSE	60	47	53	47	47	47	67	73	67	87	93	67	60	60	87	60	67
356	WIQNEFNKACELTDS	H3N2	PA		MOUSE	93	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	93
357	WVELIRGRPKEK	H1N1	NA		MOUSE	83	67	100	83	83	92	75	67	67	67	67	67	67	67	67	100	92
358	YACKRGSVNSFFSRLNWLHKSEYKYPALNVTMPNN	H3N2	HA		HUMAN	37	31	34	34	34	31	34	86	94	74	71	91	77	80	71	31	31
359	YASLRSLVASSGTLEF	H3N2	HA		HUMAN	38	38	38	38	38	38	38	100	100	94	100	100	100	100	100	44	44
360	YDVPDYASLRSLVASS	H3N2	HA	HLA-DRB1*0101;HLA-DRB1*0701	HUMAN	38	38	38	38	38	38	38	100	100	94	100	100	100	100	100	38	38
361	YIWGIHHPSTNQEQTSLY	H3N2	HA	H-2-IED	MOUSE	56	44	50	44	44	44	61	78	72	89	94	72	67	67	89	56	61
362	YNIRNLHIPEVCLKW	H3N2	PB1		MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
363	YPALNVTMPNNGKFDKLYIWGVHHPSTDRDQTS	H3N2	HA		HUMAN	48	42	39	45	45	45	45	85	91	79	82	85	88	91	85	39	36
364	YRRVNGKWM	H1N1	NP	H-2-DB	MOUSE	78	67	100	78	67	78	78	89	78	78	78	78	67	89	78	67	67
365	YRYGNGVWI	H1N1	NA	H-2-DB		78	89	100	100	100	100	89	44	44	44	44	44	44	44	44	78	78
366	YSLVGIDPFRLLQNS	H3N2	NP		MOUSE	100	80	100	87	87	87	100	93	93	93	93	93	93	93	93	100	100
367	YTLDEESRARIKTRL	H3N2	PA		MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
368	YVKQNTLKLATGMRNV	H3N2	HA		HUMAN	50	44	44	44	44	44	44	100	100	100	100	100	100	100	100	69	69
369	YVQASGRVTVSTRRS	H3N2	HA	H-2-IAB	MOUSE	40	40	40	40	40	40	40	87	87	87	100	87	80	87	93	40	40
370	YVSVGTSTLNK	H2N2	HA	H-2-KD	MOUSE	64	45	64	55	45	55	55	45	45	45	45	45	36	36	45	82	82
371	STNQEQTSLYVQA	H3N2	HA	H-2-IAD	MOUSE	46	38	38	38	38	38	46	69	62	92	100	62	54	54	92	38	38
372	STNQEQTSLYVQAS	H3N2	HA	H-2-IAD	MOUSE	43	36	36	36	36	36	43	71	64	93	100	64	57	57	93	36	36
373	TGKICNNPHRILDGIDCTLIDALLGDPHCDVFQNETWDL	H3N2	HA	H-2-K CLASS II	MOUSE	21	23	23	21	21	21	21	82	79	92	100	82	74	77	97	31	28
374	AMDSNTLEL	H5N1	NP	HLA-A2.1/KB	MOUSE	78	67	67	67	56	67	67	67	67	67	78	67	67	67	67	89	100
375	ATVAGSL	H2N2	HA	H-2-KD	MOUSE	71	71	71	71	71	71	71	43	43	43	43	43	43	43	43	71	71
376	EDLTFLARS	H1N1	NP		HUMAN	89	100	100	100	100	89	89	78	78	89	89	78	89	89	89	89	89
377	FMYSDFHFI	H1N1	PA	HLA-A*0201;HLA-A*0202;HLA-A*0206	HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
378	GIAPLQLGK	H1N1	HA	HLA-A*0301;HLA-A11		100	89	100	89	89	100	100	44	44	44	44	44	44	44	44	44	44
379	GIHHPSNSK	H1N1	HA	HLA-A11		44	56	100	67	67	67	67	67	44	56	56	56	44	56	56	44	56
380	GLISLILQI	H1N1	NA	HLA-A*0201		89	78	100	89	89	89	89	56	56	56	56	56	56	56	56	78	67
381	GLKGGPSTE	H1N1	M2	HLA-A2/KB	MOUSE	89	89	100	89	89	89	89	89	89	89	89	89	89	89	89	89	67
382	ILGFVFTLTV	H1N1	M1	HLA-A*0201;HLA-A*020101;HLA-A*0203	HUMAN;MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
383	IRGWAIYS	H1N1	NA	H-2-IED	MOUSE	88	75	100	88	88	100	88	50	50	50	50	50	50	50	50	88	75
384	IYQILAIYSTVASSL	H1N1	HA	H-2-KD	MOUSE	40	93	100	93	93	93	93	33	33	33	33	33	33	33	33	87	93
385	IYQILAIYSTVASSLVL	H1N1	HA	H-2-KD	MOUSE	35	94	100	94	94	94	94	35	35	35	35	35	35	35	35	82	88
386	IYQILAIYSTVASSLVLLVS	H1N1	HA	H-2-KD	MOUSE	30	95	100	95	95	95	95	30	30	30	30	30	30	30	30	70	75
387	IYQILAIYSTVASSLVLLVSLGA	H1N1	HA	H-2-KD	MOUSE	30	96	100	96	96	96	96	30	30	30	30	30	30	30	30	61	65
388	KSMREEYRK	H1N1	M2	HLA-A11	HUMAN	89	78	100	78	78	89	89	89	89	89	89	89	89	89	89	78	78
389	MLLRSAIGQV	H1N1	PA	HLA-A*0201		100	100	100	100	100	100	90	90	90	90	80	90	90	90	90	90	80
390	NETWDLFVERS	H3N2	HA	H-2-IAD	MOUSE	45	45	45	36	45	36	45	91	82	100	100	82	82	82	100	45	36
391	NMLSTVLGV	H1N1	PB1	HLA-A*0201;HLA-A*0206;HLA-A*6802	HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
392	RMVLASTTAK	H1N1	M1	HLA-A*0301;HLA-A*11;HLA-A11	HUMAN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Table 7. Conservancy analysis of T cell linear epitope sequences

No.	Sequence	Influenza Source Subtype	Source Protein	MHC Restriction Allele(s)	Host Species	H1N1							H3N2								H5N1	
						A/Brevig Mission/1/18	A/New Caledonia/20/99	A/PR/8/34	A/Taiwan/1/86	A/Texas/36/91	A/USSR/90/77	A/WS/33	A/Bangkok/1/79	A/Beijing/353/89	A/England/42/72	A/Hong Kong/1/68	A/Leningrad/360/86	A/New York/5/2004	A/Panama/2007/99	A/UDORN/307/72	A/Hong Kong/156/97	A/Viet Nam/1194/2004
393	RTLDFHDSNVK	H1N1	HA	HLA-A*3301;HLA-A*6801;HLA-A11	HUMAN	36	100	100	100	100	100	91	45	45	45	45	45	45	45	45	100	100
394	RVLSFIKGTK	H1N1	NP	HLA-A*3101;HLA-A*6801;HLA-A11	MOUSE	70	70	100	70	70	80	80	80	80	80	90	80	80	80	80	70	70
395	SIIPSGPLK	H1N1	M1	HLA-A*3101;HLA-A11	HUMAN	89	89	100	89	89	89	89	89	89	89	89	89	89	89	89	100	100
396	SLCPIRGWAI	H1N1	NA	HLA-A*0201;HLA-A*0206	HUMAN	90	80	100	90	100	100	100	40	40	40	40	40	40	40	40	90	80
397	SLENFRAYV	H1N1	PA	HLA-A*0201;HLA-A*0203;HLA-A*6802	HUMAN	100	78	100	89	89	89	100	89	89	89	89	89	89	89	89	100	100
398	VTAACSHAGK	H1N1	HA	HLA-A*0301;HLA-A11		80	70	100	90	70	80	80	40	40	40	40	40	40	40	40	40	50
399	YFKIHTGKSS	H3N2	HA		HUMAN	50	50	50	50	50	50	50	90	90	80	80	90	80	80	80	40	40
400	YGKQNTLKLA	H3N2	HA		HUMAN	40	50	40	50	50	40	40	90	90	90	90	90	90	90	90	60	60
401	YIKQDTLKLA	H3N2	HA		HUMAN	40	40	40	40	40	40	40	80	80	80	80	80	80	80	80	50	50
402	YIKQNTLKLA		HA		HUMAN	40	40	40	40	40	40	40	90	90	90	90	90	90	90	90	60	60
403	YIKQNTLKLS	H3N2	HA		HUMAN	40	40	40	40	40	40	40	80	80	80	80	80	80	80	80	50	50
404	YVKENTLKLA	H3N2	HA		HUMAN	40	40	40	40	40	40	40	90	90	90	90	90	90	90	90	70	70
405	YVKQDTLKLA	H3N2	HA		HUMAN	40	40	40	40	40	40	40	90	90	90	90	90	90	90	90	60	60
406	YVKQHTLKLA	H3N2	HA		HUMAN	40	40	40	40	40	40	40	90	90	90	90	90	90	90	90	60	60
407	YVKQNSLKLA	H3N2	HA		HUMAN	40	40	40	40	40	40	40	90	90	90	90	90	90	90	90	70	70
408	YVKQNTLKLA	H3N2	HA		HUMAN	40	40	40	40	40	40	40	100	100	100	100	100	100	100	100	70	70
409	YVKQNTLKVA	H3N2	HA		HUMAN	40	40	40	40	40	40	40	90	90	90	90	90	90	90	90	60	60
410	YVKQNTLRLA	H3N2	HA		HUMAN	50	40	40	40	40	40	40	90	90	90	90	90	90	90	90	70	70
411	YVKQSTLKLA	H3N2	HA		HUMAN	50	50	50	50	50	50	50	90	90	90	90	90	90	90	90	60	60
412	YVKQTTLKLA	H3N2	HA		HUMAN	50	40	40	40	40	40	40	90	90	90	90	90	90	90	90	60	60

Identity level color code:

Yellow: 100%

Magenta: ≥90%

Green: ≥80%

Table 8. Distribution of linear epitopes conserved at different identity levels

Identity (%)	Antibody		T cell	
	Human (%)*	Avian (%)**	Human (%)	Avian (%)
<80	78.7	82.7	8.5	15.8
≥80	10.7	9.3	46.6	45.6
≥90	8.0	5.3	29.9	27.2
100	2.7	2.7	15.0	11.4
Number of epitopes	**75**		**412**	

* Fraction of epitopes conserved in the human H1N1 and H3N2 strains

** Fraction of epitopes conserved in the avian H5N1 strains

Table 9. Conservancy analysis of antibody conformational epitope sequences

No.	Sequence	Source Species Subtype	Source Protein	Antibody Type(s)	Host Species	H1N1							H3N2								H5N1	
						A/Brevig Mission/1/18	A/New Caledonia/20/99	A/PR/8/34	A/Taiwan/1/86	A/Texas/36/91	A/USSR/90/77	A/WS/33	A/Bangkok/1/79	A/Beijing/353/89	A/England/42/72	A/Hong Kong/1/68	A/Leningrad/360/86	A/New York/5/2004	A/Panama/2007/99	A/UDORN/307/72	A/Hong Kong/156/97	A/Viet Nam/1194/2004
1	A198, S199, R201	H3N2	HA	MONOCLONAL; POLYCLONAL	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
2	D147, H150, H197, D198, E199, K221, D251	H3N2	NA	MONOCLONAL		43	43	43	43	43	43	43	57	86	57	57	71	71	100	57	57	43
3	D147, H150, R152, T153, P154, H197, D198, E199, W218, S219, K220, K221, I222, G248, R249, A250, D251	H3N2	NA	MONOCLONAL	MOUSE	29	35	41	35	35	35	35	76	88	71	71	82	82	100	76	29	29
4	D188,Q189, R201	H3N2	HA	MONOCLONAL	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
5	E152, G155, Q188, N189, Q192, E194	H1N1	HA	MONOCLONAL	MOUSE	50	50	100	50	50	50	50	50	50	50	50	50	50	50	50	50	50
6	E189, D190, E219	H1N1	HA	MONOCLONAL		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
7	G129, Y155, S157, G158, S159	H3N2	HA	MONOCLONAL	MOUSE	60	60	60	60	60	60	60	80	60	60	80	60	60	60	100	60	60
8	G129, Y155, S157, G158, S159, A160, Q189	H3N2	HA	MONOCLONAL	MOUSE	57	57	57	57	57	57	57	57	57	57	71	57	57	57	86	57	57
9	G129, Y155, S159	H3N2	HA	MONOCLONAL	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	67
10	G135, K140, G142, S145, G158	H3N2	HA	MONOCLONAL	MOUSE	60	60	60	60	60	60	60	60	60	80	100	60	80	80	100	60	60
11	G47,I48,T49,N50,K51,V52,N53,S54,I55,I56,D57,K58,T318,G319,P320,R321,N322	H5N2	HA	MONOCLONAL	MOUSE	29	35	35	35	35	35	35	35	35	35	35	35	29	29	35	35	35
12	G49,K50, L59,D60, I62,D63, P74,H75, V78,F79, R90, K92, F94, P143, D271, P273,I274,D275	H3N2	HA	MONOCLONAL	MOUSE	28	28	28	28	28	28	33	67	61	83	100	67	56	61	89	28	28
13	I369, A370, S371, L399, N400, T401, D402, W403, P433, K434, E435, D436, K437, Q317, Y318, I319, C320, S321, P343, G344, N345, N346, N347, N348	H11N9	NA	MONOCLONAL		42	29	38	29	29	29	38	25	25	25	25	25	25	25	25	38	38
14	I51, C52, N53, N54, T276, C277, I278, S279	H3N2	HA	MONOCLONAL	MOUSE	38	38	38	38	38	38	50	63	63	100	100	63	50	50	100	38	38
15	K129, G159, E189	H1N1	HA	MONOCLONAL	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
16	K129, V132, K157	H1N1	HA	MONOCLONAL	MOUSE	67	67	67	67	67	67	67	100	100	67	100	100	100	100	100	67	67
17	K142, G143, K172, E219, G225, A227	H1N1	HA	POLYCLONAL	CHICKEN	67	67	67	67	67	100	67	67	67	67	67	67	67	67	67	67	67
18	K148, S198, S364, D395, N396	H3N8	NA	MONOCLONAL	MOUSE	60	60	60	60	60	60	60	80	60	80	60	60	60	60	80	60	60
19	K50, L59,D60, I62,D63, P74,H75, V78,F79, R90, K92, F94, G142, D271	H3N2	HA	MONOCLONAL	MOUSE	29	29	29	29	29	29	29	71	64	86	100	71	50	57	93	29	29
20	L70, L71, V73, R74, S75, E115	H1N1	HA	MONOCLONAL	MOUSE	67	67	100	67	67	67	83	50	50	50	50	50	50	50	50	67	67
21	N329, S367, A369	H11N9	NA	MONOCLONAL		67	67	67	100	67	67	67	67	67	100	100	67	67	67	100	100	67
22	N329, T332, K336, Y341, N344, S367, I368, K432	H13N9	NA	MONOCLONAL	MOUSE	63	50	50	50	50	50	38	50	50	50	50	50	50	50	50	50	50
23	P124, N125, E154, S156, P158, K159, K161, N162, S163	H1N1	HA	MONOCLONAL	MOUSE	56	44	78	44	44	44	78	44	44	44	44	44	44	44	44	56	56
24	P326, R327, N329, P328, N344, N347, I366, S367, I368, A369, S370, S372, L399, N400, T401,D402, W403, P431, K432	H11N9	NA	MONOCLONAL	MOUSE	42	26	37	26	26	26	32	37	26	37	26	26	32	37	37	37	37
25	P326, R327, P328, D329, N344, N347, I366, S367, I368, A369, S370, S372, L399, N400, T401,D402, W403, P431, K432	H11N9	NA	MONOCLONAL	MOUSE	37	26	37	26	26	26	32	32	26	32	32	26	26	32	32	32	32
26	P326, R327, P328, N329, N344, N347, I366, S367, R368, A369, S370, S372, L399, N400, T401,D402, W403, P431, K432	H11N9	NA	MONOCLONAL	MOUSE	47	32	42	32	32	32	37	37	32	37	32	32	32	37	37	42	42
27	P328,N329,D330,P331,T332, Y341, G343,N344, I366, I368,A369,S370, S372, N400,T401, W403	H13N9	NA	MONOCLONAL	MOUSE	31	38	31	38	31	31	38	31	31	31	31	31	31	31	31	31	31
28	P330, N331, N346, N347, S368, I369, A370, S371, S373, N400, T401, W403, K434	H11N9	NA	MONOCLONAL		46	38	38	38	38	38	31	31	31	38	38	31	31	31	38	46	46
29	Q100, T156, G196	H7N7	HA	MONOCLONAL	MOUSE	67	67	67	67	67	67	67	100	100	100	100	100	100	100	100	67	67
30	Q189, V196, A198	H3N2	HA	MONOCLONAL	MOUSE	67	100	67	100	100	100	100	67	67	100	100	67	67	67	100	100	67
31	S127, D135, T179, N183, T188, L216	H9N2	HA	MONOCLONAL	MOUSE	67	67	67	67	67	67	67	67	67	67	67	67	67	67	67	67	67
32	S127, T129, K147, P152, T179, T188	H9N2	HA	MONOCLONAL	MOUSE	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50
33	S136, G139, S141, R220, D221	H1N1	HA	MONOCLONAL	MOUSE	100	100	100	80	100	80	100	60	60	60	60	60	60	60	60	60	60
34	S186, R220, R229, I230	H3N2	HA	MONOCLONAL	MOUSE	75	75	75	100	100	100	75	75	75	100	100	75	75	75	100	75	75
35	S364, D395, N396	H3N8	NA	MONOCLONAL	MOUSE	100	100	100	67	100	67	67	100	67	100	67	67	67	67	100	100	100
36	S367, S372, N400	H11N9	NA	MONOCLONAL		67	100	67	100	100	67	67	100	100	100	100	100	100	100	100	100	100
37	S368, S373, N400, T401, K434		NA	MONOCLONAL	MOUSE	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	80
38	T131, G134, S136, S145, W153, T155, K156, S17, G158, E190, S193, L194, L226	H3N2	HA	MONOCLONAL	MOUSE	69	54	38	38	38	38	77	54	46	69	92	54	54	31	85	54	54
39	T318, G319, L320, R321, N322, G47, I48, T49, N50, K51, V52, N53, S54, V55, I56, E57, K58		HA	MONOCLONAL	MOUSE	29	41	29	29	29	29	35	41	41	41	41	41	29	29	41	35	35
40	V165, G169, I178, S203, G236, S270	H1N1	HA	MONOCLONAL	MOUSE	67	67	100	67	67	67	83	67	67	67	67	67	67	67	67	67	67
41	Y155, S157, S159	H3N2	HA	MONOCLONAL	MOUSE	67	67	67	67	67	67	67	100	67	67	67	67	67	67	100	67	67
42	Y155, S157, G158, N193	H3N2	HA	MONOCLONAL	MOUSE	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75
43	Y155, S159, T188, D189, T193, A198, S199, R201	H3N2	HA	MONOCLONAL	MOUSE	50	50	38	50	50	50	50	63	50	50	50	50	50	50	63	50	50
44	K432, W403, T401, N400, S372, S370, A369, I368, S367, I366, P, 342, Y341, N329, T332, P331, D330, N344, G343, D330, P328, N200	H11N9	NA	MONOCLONAL	MOUSE	40	25	35	25	25	25	25	35	25	35	25	25	30	35	35	35	35

Identity level color code:
Yellow: 100%
Magenta: ≥90%
Green: ≥80%

Table 10. Conservancy analysis of protective antibody/T cell epitopes

Epitope Type	Sequence	Source Species	Source Protein	Protected Host Species	H1N1							H3N2								H5N1	
					A/Brevig Mission/1/18	A/New Caledonia/20/99	A/PR/8/34	A/Taiwan/1/86	A/Texas/36/91	A/USSR/90/77	A/WS/33	A/Bangkok/1/79	A/Beijing/353/89	A/England/42/72	A/Hong Kong/1/68	A/Leningrad/360/86	A/New York/5/2004	A/Panama/2007/99	A/UDORN/307/72	A/Hong Kong/156/97	A/Viet Nam/1194/2004
Antibody	MSLLTEVETPIRNEWGCRCNDSSD	A/WSN/33(H1N1)	M2	MOUSE	96	63	96	63	63	100	100	63	63	100	100	63	96	63	63	83	88
Antibody	SLLTEVETPIRNEWGCRCNDSSD	A/AICHI/2/68(H3N2); A/PUERTO RICO/8/34(H1N1); A/USSR/90/77(H1N1)	M2	MOUSE; RHESUS MONKEY	96	65	96	65	65	100	100	65	65	100	100	65	96	65	65	83	87
Antibody	NVPEKQTRGIFGAIAGFIE	INFLUENZA A VIRUS	HA	MOUSE	74	74	74	74	74	74	74	100	100	95	95	100	100	100	95	58	58
Antibody	WTGVTQN	A/AICHI/2/68(H3N2); A/TEXAS/1/77(H3N2)	HA	MOUSE	43	43	43	43	43	43	43	86	71	100	100	86	100	86	100	43	43
Antibody	SKAFSNCYPYDVPDYASL	A/MEMPHIS/6/86(H3N2); A/TEXAS/1/77(H3N2)	HA	RABBIT	39	33	39	33	33	33	39	100	94	94	100	100	94	94	100	44	39
Antibody	G47,I48,T49,N50,K51,V52,N53,S54,I55,I56,D57,K58,T318,G319,P320,R321,N322	A/MALLARD DUCK/PA/10218/84(H5N2)	HA	MOUSE	29	35	35	35	35	35	35	35	35	35	35	35	29	29	35	35	35
Antibody	G49,K50, L59,D60, I62,D63, P74,H75, V78,F79, R90, K92, F94, P143, D271, P273,I274,D275	A/AICHI/2/68(H3N2); A/X-31(H3N2)	HA	MOUSE	28	28	28	28	28	28	33	67	61	83	100	67	56	61	89	28	28
Antibody	L70, L71, V73, R74, S75, E115	A/PUERTO RICO/8/34(H1N1); A/PUERTO RICO/8/34/MOUNT SINAI(H1N1)	HA	MOUSE	67	67	100	67	67	67	83	50	50	50	50	50	50	50	50	67	67
Antibody	T318, G319, L320, R321, N322, G47, I48, T49, N50, K51, V52, N53, S54, V55, I56, E57, K58	INFLUENZA A VIRUS	HA	MOUSE	29	41	29	29	29	29	35	41	41	41	41	41	29	29	41	35	35
T cell	MGLIYNRM	A/ANN ARBOR/6/60(H2N2); A/PUERTO RICO/8/34(H1N1)	M1	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
T cell	TYQRTRALV	INFLUENZA A VIRUS; A/ANAS ACUTA/PRIMORJE/695/76(H2N3); A/MEMPHIS/1/71H-A/BELLAMY/42N); A/NT/60/68/(H3N2); A/PORT CHALMERS/1/73(H3N2); A/PUERTO RICO/8/34(H1N1)	NP	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
T cell	TYQRTRALVRTG	INFLUENZA A VIRUS; A/PUERTO RICO/8/34(H1N1); A/TEXAS/1/77(H3N2); A/X-31(H3N2)	NP	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
T cell	SSYRRPVGI	INFLUENZA A VIRUS; A/ANN ARBOR/6/60(H2N2); A/NT/60/68/(H3N2); A/PUERTO RICO/8/34(H1N1)	PB1	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
T cell	ASNENMETM	INFLUENZA A VIRUS; A/ANAS ACUTA/PRIMORJE/695/76(H2N3); A/PUERTO RICO/8/34(H1N1); A/X-31(H3N2); INFLUENZA A VIRUS H3N2	NP	MOUSE	100	67	100	89	67	89	100	89	78	89	78	78	78	78	89	78	89
T cell	RTFSFQLI	A/AICHI/2/68(H3N2); A/MEMPHIS/102/72(H3N2); A/PUERTO RICO/8/34(H1N1)	NS2	MOUSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	88
T cell	PNGYIEGK	A/PUERTO RICO/8/34(H1N1)	PA	MOUSE	100	100	100	100	100	100	100	88	88	88	100	88	88	88	88	88	88
T cell	VTGLRNIPS	A/PUERTO RICO/8/34(H1N1)	HA	MOUSE	89	100	100	100	100	100	100	56	56	56	56	56	56	56	56	67	67
T cell	ALNNRFQIKGVELKS	A/MEMPHIS/1/71(H3N2)	HA	MOUSE	33	47	40	47	47	47	47	100	100	33	100	100	100	100	100	53	47

Identity level color code:

100%
≥90%
≥80%

((epitope[TW] OR epitopes[TW] OR mimotope[TW] OR ((MHC[tw] OR "major histocompatibility complex"[tw] OR HLA[tw]) AND (peptide[tw] OR peptides[tw])) OR "TCR recognition"[tw] OR ("Class"[tw] AND "I motif"[tw]) OR supermotif[tw] OR immunogenic linear OR ("peptide-based"[tw] AND CTL[tw]) OR phage displa*[tw] OR "antibody binding"[tw] OR "protective immune response"[tw] OR antibody recog*[tw] OR "cytotoxicity assay"[tw] OR "new monoclonal"[tw] OR "novel antibody"[tw] OR ((monoclonal antibod*[tw]) AND "binding site"[tw]) OR ((KA[tw] OR KD[tw]) AND (monoclonal[tw] OR mAb[tw])) OR "neutralizing antibody"[tw] OR "peptide vaccine"[tw] OR (peptide conjugate vaccine*[tw]) OR ((CD8[tw] OR CD4[tw]) AND “T cells”[tw] AND (peptide[tw] OR peptides[tw])) OR (“antigenic repertoire”[tw]) OR ((peptide[tw] OR peptides[tw]) AND “antibody reactivity”[tw]) OR ("Class II"[tw] AND (binding [tw] OR bound[tw] OR peptide[tw] OR peptides[tw])) OR "immunogenic peptide"[tw])) **AND** (("Influenza Virus"[Text Word]) OR ("Influenza A virus"[Text Word]) OR ("Influenza B virus"[Text Word]) OR ("Influenza C Virus"[Text Word]) OR ("Influenza D Virus"[Text Word]) OR ("Influenzavirus A"[Text Word]) OR (“Flu”[Text Word]) OR ("Influenzavirus B"[Text Word]) OR ("Influenzavirus C"[Text Word]) OR (influenza[Text Word])) **AND** (hasabstract[text] AND English[Lang] AND ("1900"[PDat]:"2006/03/27"[PDat])) **NOT** (Review[PT] OR Editorial[PT] OR meta-Analysis[PT] OR Comment[PT])

Legend:
Epitope keywords – red colored text
Influenza keywords – blue colored text
Filters – green colored text
NOT keywords – brown colored text