C. Sylvester-Hvid
M. Nielsen
K. Lamberth
G. Røder
S. Justesen
C. Lundegaard
P. Worning
H. Thomadsen
O. Lund
S. Brunak
S. Buus

SARS CTL vaccine candidates; HLA supertype-, genome-wide scanning and biochemical validation

Key words:
immunogenicity; vaccine; bioinformatics; Severe Acute Respiratory Syndrome

Acknowledgments:
The Danish MRC (grant 22-01-0272), the 5th Framework Programme of the European Commission (grant QLGT-1999–00173), the NIH (grant AI49213-02), and the Danish National Research Foundation supported this work. Support for this work has been obtained from SIGA Technologies.

Abstract: An effective Severe Acute Respiratory Syndrome (SARS) vaccine is likely to include components that can induce specific cytotoxic T-lymphocyte (CTL) responses. The specificities of such responses are governed by human leukocyte antigen (HLA)-restricted presentation of SARS-derived peptide epitopes. Exact knowledge of how the immune system handles protein antigens would allow for the identification of such linear sequences directly from genomic/proteomic sequence information (Lauemoller et al., *Rev Immunogenet* 2001: 2: 477–91). The latter was recently established when a causative coronavirus (SARS-CoV) was isolated and full-length sequenced (Marra et al., *Science* 2003: 300: 1399–404). Here, we have combined advanced bioinformatics and high-throughput immunology to perform an HLA supertype-, genome-wide scan for SARS-specific CTL epitopes. The scan includes all nine human HLA supertypes in total covering >99% of all individuals of all major human populations (Sette & Sidney, *Immunogenetics* 1999: 50: 201–12). For each HLA supertype, we have selected the 15 top candidates for test in biochemical binding assays. At this time (approximately 6 months after the genome was established), we have tested the majority of the HLA supertypes and identified almost 100 potential vaccine candidates. These should be further validated in SARS survivors and used for vaccine formulation. We suggest that immunobioinformatics may become a fast and valuable tool in rational vaccine design.

Authors' affiliations:

C. Sylvester-Hvid[1],
M. Nielsen[2],
K. Lamberth[1],
G. Røder[1],
S. Justesen[1],
C. Lundegaard[2],
P. Worning[2],
H. Thomadsen[1],
O. Lund[2],
S. Brunak[2],
S. Buus[1]

[1]Division of Experimental Immunology, Institute of Medical Microbiology and Immunology (IMMI), University of Copenhagen, Copenhagen N, Denmark

[2]Center for Biological Sequence Analysis (CBS), Technical University of Denmark, Lyngby, Denmark

Correspondence to:
Søren Buus
Institute for Medical Microbiology and Immunology
Building 18.3
Panum Institute
Blegdamsvej 3
2200 Copenhagen N
Denmark
Tel.: +45 35 32 78 85
Fax: +45 35 32 76 96
e-mail: sb@immi.ku.dk

Received 18 September 2003, revised 18 December 2003, accepted for publication 22 December 2003

Tissue Antigens 2004: **63:** 395–400
Printed in Denmark.

Severe Acute Respiratory Syndrome (SARS) has in about 7 months infected more than 8400 patients in over 30 countries and caused more than 800 deaths. The prospect of a deadly epidemic has had significant disruptive consequences in many health, social, economic, and political aspects of life. Coordinated by the WHO, classical measures including case detection, isolation, infection control, contact tracing, and follow-up surveillance have successfully contained the disease; however, the disease has proven resilient and it cannot be excluded that it will resurface given the right conditions. Should that occur, even the best health system would be severely strained if it had to resume and sustain the containment effort implemented in the spring of 2003. Ideally, the SARS-coronavirus (SARS-CoV) should be

eradicated. This would require detection assays that can track the disease and intervention measures that can break the chain of transmission. All of these procedures should be simple, yet effective. Unfortunately, no such diagnostic test is currently available, and controlling transmission by containment solely is complicated and extremely costly. Further complicating any eradication effort, a non-human reservoir appears to exist. Thus, a strong case for a SARS vaccine can be made. It would be of significant help in any eradication effort and, should that fail, it could protect infected individuals against the disease.

The SARS-CoV infects epithelial cells in the respiratory tract causing interstitial pneumonia (4). One would therefore expect that an effective vaccine should induce mucosal immunity such as that effected by secretory immunoglobulin A (IgA), which specifically prevents an infectious agent from penetrating the mucosal epithelium, and by cytotoxic T lymphocytes (CTLs), which specifically eradicate infected cells (5). IgA responses are generally considered the major protective mechanism; however, there are examples of CTLs, not antibodies, being responsible for early control of mucosal infection (5). Particularly noteworthy, this is the case for the infectious bronchitis virus of chicks, a prototype of the Coronaviridae family, where primary effector $CD8^+$ CTLs play a critical role in the elimination of virus during acute infection and subsequent control of the infection (6–10).

Human CTLs are specific for peptides presented in the context of human leukocyte antigen (HLA) molecules [generically known as "major histocompatibility complex (MHC) molecules"]. Prior to presentation, peptides are generated in the cytosol by limited proteolytic fragmentation of all available protein antigens, translocated to the endoplasmic reticulum, specifically sampled by the MHC molecules and exported to the cell surface, where they await CTL scrutiny. Importantly, the HLA is extremely polymorphic and the peptide binding specificity varies for the different polymorphic HLA molecules (1). It has, however, been suggested that the majority of all major human populations can be covered with three to nine "HLA supertypes", where the different members of each supertype bind similar peptides (3). If one knew exactly how peptides were generated and selected, then genomic/proteomic information could be used to predict the outcome of antigen presentation and forecast immunogenicity. Here, we have used advanced immunobioinformatical tools to mimic antigen presentation and, in a highly cost- and time-effective manner, predicted possible immunogenic epitopes.

The complete SARS genome/proteome was obtained from GenBank (NC004718) and virtually digested into all 9862 unique nonamer peptides (2). Thus, close to 10,000 binding predictions were made for each of the nine HLA supertypes. Artificial neural networks (ANNs) were used to predict the binding affinity quantitatively when the corresponding data were available [e.g. for A*0201 (11,12)]. The performance of the ANNs is high, as the correlation coefficient between predicted and measured binding is 0.85. The remaining HLA bindings were predicted using weight matrices derived from Gibbs sampling sequence-weighting methods with pseudocount correction for low counts as well as differential position-specific anchor weighting (Nielsen et al. manuscript in preparation). These weight matrices were calculated from available nonamer data from the SYFPEITHI and MHCPEP databases with the peptides clustered into the nine supertypes (A1, A2, A3, A24, B7, B27, B44, B58, and B62). The positive predictive value of the matrix-driven prediction has been found to be around 66%, whereas the negative predictive value has been found to be around 97% (Lamberth et al. unpublished observation). Proteasomal processing was predicted using NETCHOP 2.0 (13). NETCHOP 2.0 has been found to be superior to other proteasomal prediction algorithms (14). Peptides with a NETCHOP 2.0 score below 0.5 (i.e. poorly predicted proteasomal processing) were excluded from further analysis. Finally, we excluded all peptides that did not represent epitopes conserved in all SARS isolates. Figure 1 shows a representative example for a member of the HLA-A3 supertype, the HLA-A*1101 (this haplotype is particularly common in Southeast Asia).

Materials and methods

For each HLA supertype, the 15 top-ranking nonamer peptides were synthesized by standard 9-fluorenylmethyloxycarbonyl (FMOC) chemistry, purified by reversed-phase high-performance liquid chromatography (at least 80%, usually >95% purity) and validated by mass spectrometry. The interaction of these epitope candidates with the appropriate HLA was subsequently validated in a biochemical binding assay (15). Briefly, denatured and purified recombinant HLA heavy chains were diluted into a renaturation buffer containing HLA light chain, β_2-microglobulin, and graded concentrations of the peptide to be tested, and incubated at 18°C for 48 h allowing equilibrium to be reached. We have previously demonstrated that denatured HLA molecules can *de novo* fold efficiently, however, only in the presence of appropriate peptide (16). The concentration of peptide–HLA complexes generated was measured in a quantitative enzyme-linked immunosorbent assay and plotted against the concentration of peptide offered (15) (Fig. 2). Because the effective concentration of HLA (3–5 nM) used in these assays is below the equilibrium dissociation constant (K_D) of most high-affinity peptide–HLA interactions, the peptide concentration leading to half-saturation of the HLA is a reasonable approximation of the affinity of the interaction. An initial screening procedure was employed whereby a single high concentration (20,000 nM) of peptide was incubated with one or more HLA molecules. If no complex

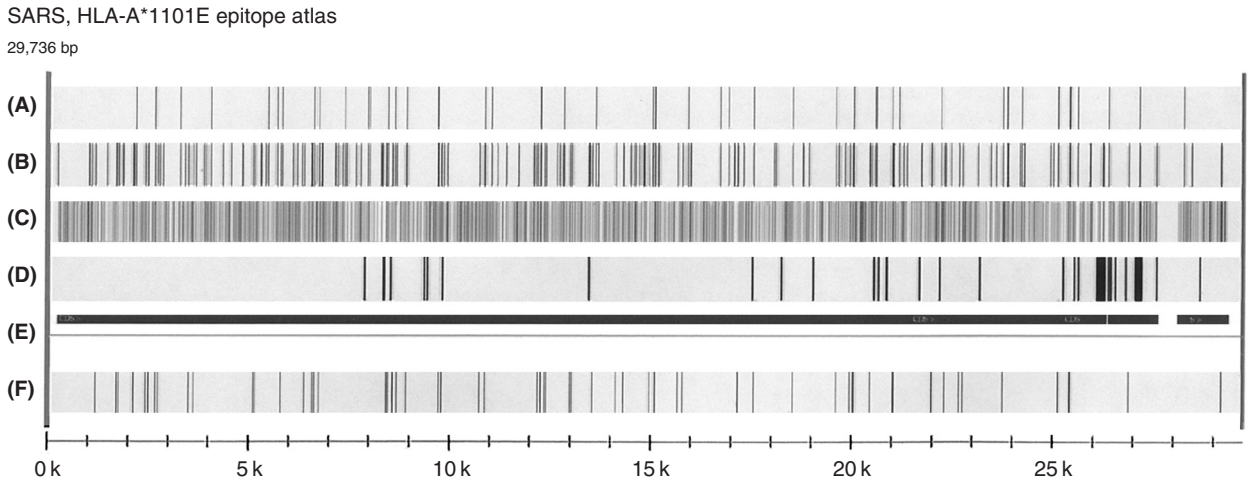

Fig. 1. **A representative example of the genome-wide scanning for putative epitopes restricted to A1.** (A) Predicted strong binding peptides (equilibrium dissociation constant, $K_D < 50$ nM); (B) predicted intermediate binding peptides ($K_D < 500$ nM); (C) predicted proteasomal cleavage (NETCHOP $2.0 > 0.5$); (D) sequence variation estimated from 12 SARS isolates; (E) assigned translated regions; and (F) combined selection of peptides (binding <500 nM, proteasomal cleavage >0.5, and sequence variation = 0).

formation was found, the peptide was assigned as a non-binder to the HLA molecule(s) in question, conversely, if complex formation was found in the initial screening, a full titration of the peptide was performed to determine the affinity of binding.

Results and discussion

The resulting binding isotherms were analyzed by one-site hyperbola regression (Prism® GraphPad) determining the concentration of HLA employed (3–5 nM, data not shown), the K_D of the interaction

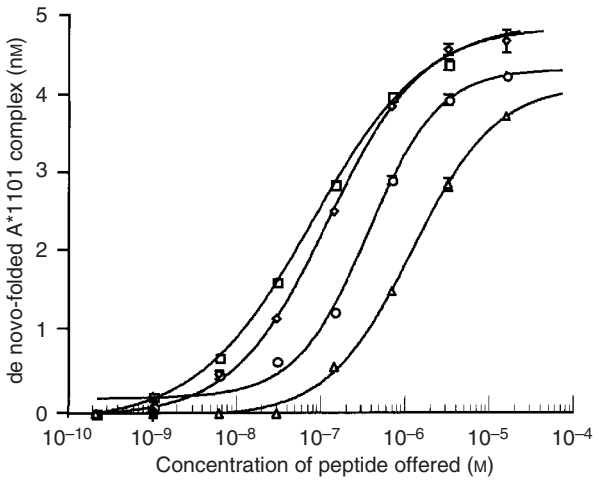

Fig. 2. **The concentration of *de novo*-folded peptide–human leukocyte antigen (HLA)-A*1101 complexes was measured in a quantitative enzyme-linked immunosorbent assay and plotted against the concentration of peptide offered (15).** Peptides: LIGANYLGK (△); MTNRQFHQK (◇); ITCVVIPSK (○); GVAMPNLYK (□).

(Table 1) and the goodness of the curve fit (R^2 was always >0.95 and in the majority of cases it was >0.98, data not shown) (15). In general, intermediate and high-affinity binders have K_Ds better (i.e. lower) than 500 nM and 50 nM, respectively, and the higher the affinity, the more likely the peptide is going to be a T-cell epitope (17). Table 1 summarizes the data for HLA-A*0301 and HLA-A*1101. The peptides are ranked according to the predicted affinity and with few exceptions, the top-ranking predictions could be confirmed as bona fide binders. Thus, 11 of the 15 peptides tested for A*0301 binding and 14 of the 15 peptides tested for A*1101 binding bound with a K_D lower than 500 nM. This would indicate that there might be even more SARS-derived binders than the 15 per HLA supertype, which are predicted here.

Figure 3 shows a graphical representation of the predicted and validated HLA binding of the SARS-derived peptides. Eight of the nine HLA supertypes have been completed (as of December 2003): A1 (represented by HLA-A*0101), A2 (represented by HLA-A*0201), A3 (represented by HLA-A*0301, and in some cases by HLA-A*1101), A24 (represented by HLA-A*2402), B7 (represented by HLA-B*0702), B44 (represented by HLA-B*4001), B58 (represented by HLA-B*5801), and B62 (represented by HLA-B*1501). A total of 952 peptide-HLA combinations were examined. The performance of the prediction tools is high as 894 (94%) of the predictions could be confirmed. For the 120 positive binding predictions, 84 (70%) could be confirmed, whereas 36 (30%) could not; for the 832 negative predictions; 810 (97%) could be confirmed, whereas 22 (3%) were unexpectedly found to represent binders (however, only of intermediate affinity, data not shown). Although these observations are biased by the selection of the top-ranking candidates, the present data

Peptide binders to human leukocyte antigen (HLA)-A*0301 and HLA-A*1101

Peptide sequence (single-letter code)	Equilibrium dissociation constant K_D (nM)	
	rA*0301	rA*1101
A3		
EVMPVSMAK	473	19
KTFPPTEPK	186	70
ATFSVPMEK	265	28
KVIQPRVEK	595	168
RLYYDSMSY	52	237
AVLQSGFRK	259	80
AVDPAKAYK	1674	124
YIFFASFYY	1176	347
KCYGVSATK	2069	8376
QLFKPLTKK	215	237
KLFAAETLK	376	234
RVFNNYMPY	358	42
ALRANSAVK	197	1760
VVYRGTTTY	42	117
VTFQGKFKK	321	91
A1101		
STDDCFANK	2360	80
ATVVIGTSK	232	30
ATNNVFRLK	401	233
SSNVANYQK	442	23
AVAVHDFFK	572	281
KMQRMLLEK	305	341
LIGANYLGK	999	1470
GTLSYDNLK	365	48
ASLPTTIAK	56	19
GVAMPNLYK	60	73
MTNRQFHQK	88	145
ITCVVIPSK	1881	290
AITTSNCAK	646	54
AIKCVDIVK	3044	327
SSSLTSLLK	276	51

Peptide binders to HLA-A*0301 (top frame) and HLA-A*1101 (bottom frame), sorted according to predicted binding strength, were synthesized and the affinities of binding to A*0301 and A*1101 were determined. The peptide sequence is given in single-letter code and the measured binding affinity is given as the K_D.

Table 1

demonstrate the selection and diversification power of the HLA system. Each HLA molecule selects a very specific peptide repertoire; for the top-ranking peptides, only 22 (or 2–3%) of the 952 combinations involves cross-responses, where a peptide predicted to be a top-ranking binder to one HLA molecule turns out to be a binder to a member of another HLA supertype, i.e. it supports the contention that HLA supertypes effect significant diversification of anti-SARS CTL responses. Conversely, the overlap between different members of the same HLA supertype appears to be extensive. Thus, 13 of the 15 peptides predicted to be good binders to A*0301 were found to bind to another member of the A3 supertype, HLA-A*1101. Similarly, nine of the 15 peptides predicted to be good binders to A*1101 were found to bind to HLA-A*0301 (Table 1). Thus, it may not be necessary to know the exact HLA haplotype of any single individual; one may still expect to achieve considerable coverage of the major human populations just by selecting good binders for each of the different supertypes. This could have important implications for vaccine development.

Once all nine supertypes have been tested, we would project to have found well over 100 different vaccine candidates. These would all have been predicted to be successfully processed by the proteasome and biochemically validated for HLA binding. Therefore, there should be a high probability that these peptides are indeed presented to CTLs. Once that occurs, there is an approximately 50% chance of being able to raise a CTL response (18). Thus, our data do in all likelihood include some 50 CTL epitopes. To identify these from the >100 binding peptides, one could search for the corresponding CTL reactivities in peripheral blood of SARS survivors using robust and reasonably simple technologies such as interferon-γ secretion from stimulated whole blood T cells (19) (e.g. Quantiferon CMI®, Cellestis, Valencia, CA). This would constitute an important independent validation of both the proteasome and HLA predictors. Doing this in an efficient and timely manner would be of considerable importance, and one could suggest that such a strategy should be coordinated by the WHO and included in any future SARS-like effort. Alternatively, one could use the biochemically validated peptide epitopes without further information. Most humans express six different HLA molecules, and it would be highly unlikely that one would have failed to identify at least one CTL epitope in given individual. A vaccine could be formulated as a polytope design where the different peptides are linked together such that the polytope contains one or more peptides for each HLA supertype (20). This would maximize the likelihood that at least one anti-SARS CTL response is raised in any given vaccinated individual. With the number of possible epitopes available here, one could even formulate several different polytopes. Their availability would further increase the chance of generating a multi-epitope response and this would reduce the risk of the SARS virus escaping immune attack. These polytopes could be administrated in several different forms such as DNA vaccines, virus-like particles, immune-stimulating complexes (ISCOMs), etc. (20–27). These administration forms are ideally suited to exploit the potential of a fast epitope-identification approach like the one presented here; however,

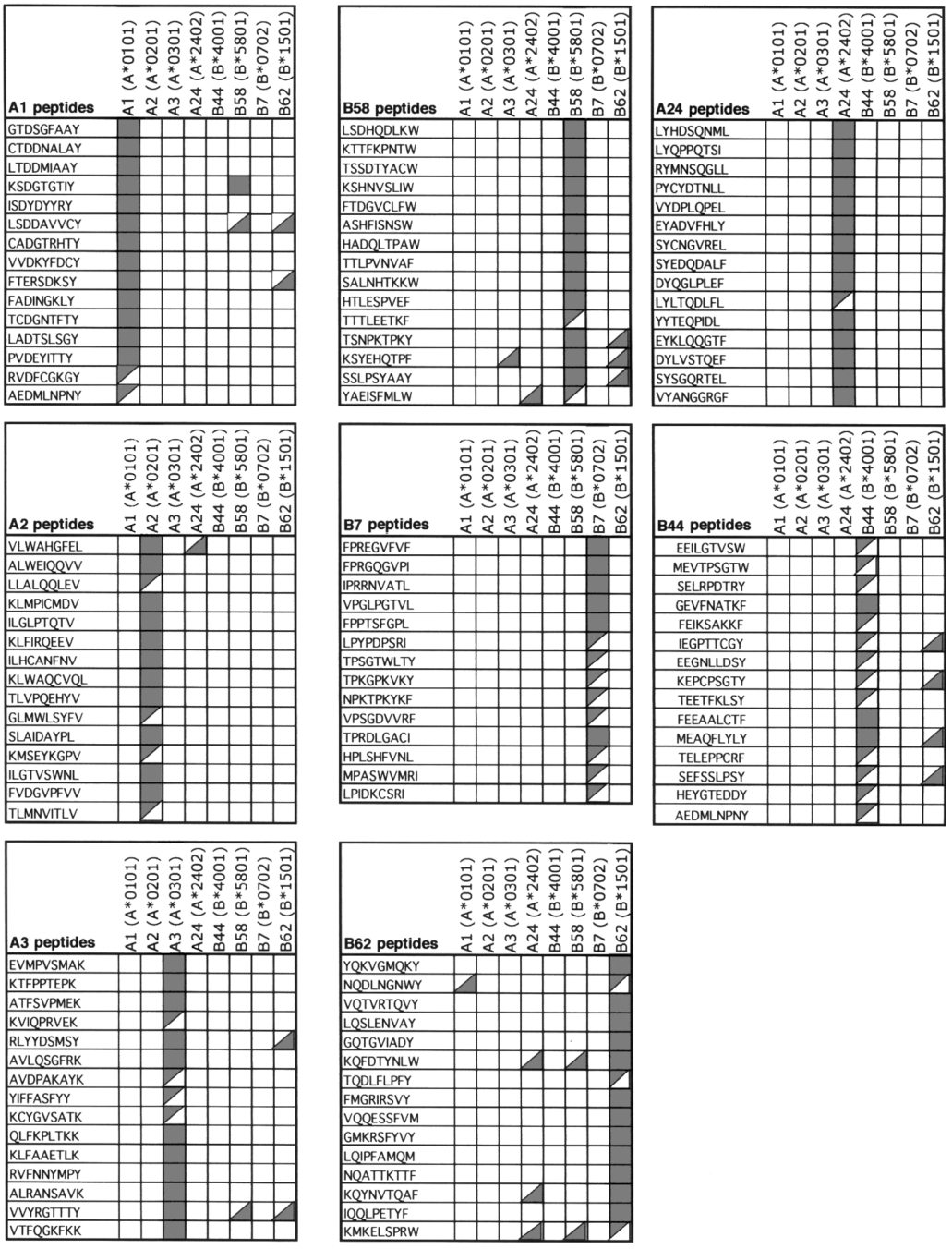

Fig. 3. **A graphical representation of the predicted and validated human leukocyte antigen (HLA) binding of 119 selected SARS-derived peptides binding to eight different HLA molecules.** It illustrates combinations that were (A) predicted and confirmed binders (true positives, ■); (B) predicted, but not confirmed, binders (false positives, ◤); (C) predicted, but not confirmed, non-binders (false negatives, ◢); and (D) predicted and confirmed non-binders (true negatives, □).

they are still at the experimental stage. In this context, it may be worth noting that the US Food and Drug Administration (FDA) has eased the regulatory requirements for the approval of anti-bioterrorism vaccines.

In the near future, the genome of any pathogen can be fully sequenced in a matter of days. The bioinformatics tools currently being developed and perfected will be able to use such genomic information to predict immune epitopes computationally, and the corresponding immunological tools will currently be able to validate these predictions in a matter of weeks to months. We predict that epitope identification in the near future will be as fast as a DNA sequencing in handling whole organisms. With the dissemination of

these tools, one could envision that clinicians and scientists anywhere would be able to analyze pathogens of their interest (or agent of bioterrorism, or tumor cell) for the purpose of fast identification of immunogenic epitopes (1). The timeline of the present SARS epidemic has demonstrated how fast modern science can identify a pathogen and decipher its genome. Using this information in a fast and rational design of vaccines, immunobioinformatics promises to take this development one step further.

References

1. Lauemoller SL, Kesmir C, Corbet S, Fomsgaard A, Holm A, Claesson MH et al. Identifying cytotoxic T cell epitopes from genomic and proteomic information: "The human MHC project". *Rev Immunogenet* 2001: **2**: 477–91.
2. Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YS et al. The genome sequence of the SARS-associated coronavirus. *Science* 2003: **300**: 1399–404.
3. Sette A, Sidney J. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics* 1999: **50**: 201–12.
4. Holmes KV. SARS coronavirus: a new challenge for prevention and therapy. *J Clin Invest* 2003: **111**: 1605–9.
5. Eriksson K, Holmgren J. Recent advances in mucosal vaccines and adjuvants. *Curr Opin Immunol* 2002: **14**: 666–72.
6. Seo SH, Collisson EW. Specific cytotoxic T lymphocytes are involved in in vivo clearance of infectious bronchitis virus. *J Virol* 1997: **71**: 5173–7.
7. Seo SH, Collisson EW. Cytotoxic T lymphocyte responses to infectious bronchitis virus infection. *Adv Exp Med Biol* 1998: **440**: 455–60.
8. Collisson EW, Pei J, Dzielawa J, Seo SH. Cytotoxic T lymphocytes are critical in the control of infectious bronchitis virus in poultry. *Dev Comp Immunol* 2000: **24**: 187–200.
9. Seo SH, Pei J, Briles WE, Dzielawa J, Collisson EW. Adoptive transfer of infectious bronchitis virus primed alphabeta T cells bearing CD8 antigen protects chicks from acute infection. *Virology* 2000: **269**: 183–9.
10. Pei J, Briles WE, Collisson EW. Memory T cells protect chicks from acute infectious bronchitis virus infection. *Virology* 2003: **306**: 376–84.
11. Buus S, Lauemoller SL, Kesmir C, Worning P, Frimurer M, Corbet S et al. Sensitive quantitative predictions of peptide-MHC binding by a "Query by Committee" artificial neural network approach. *Tissue Antigens* 2003: **62**: 5.
12. Nielsen M, Lundegaard C, Worning P, Lauemoller SL, Lamberth K, Buus S et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci* 2003: **12**: 1007–17.
13. Kesmir C, Nussbaum AK, Schild H, Detours V, Brunak S. Prediction of proteasome cleavage motifs by neural networks. *Protein Eng* 2002: **15**: 287–96.
14. Saxova P, Buus S, Brunak S, Kesmir C. Predicting proteasomal cleavage sites: a comparison of available methods. *Int Immunol* 2003: **15**: 781–7.
15. Sylvester-Hvid C, Kristensen N, Blicker T, Ferre H, Lauemøller SL, Wolff XA et al. Establishment of a quantitative ELISA-based assay capable of determining peptide–MHC class I interaction. Tissue Antigens 2002: 59.
16. Pedersen LØ, Nissen MH, Hansen NJV, Nielsen LLB, Laeumøller SL, Blicker T et al. Efficient assembly of recombinant major histocompatibility complex class I molecules with preformed disulfide bonds. *Eur J Immunol* 2001: **31**: 2986.
17. Sette A, Vitiello A, Reherman B, Fowler P, Nayersina R, Kast WM et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J Immunol* 1994: **153**: 5586–92.
18. Yewdell JW, Bennink JR. Immunodominance in MHC class I restricted T lymphocyte responses. *Annu Rev Immunol* 1999: **17**: 51–88.
19. Petrovsky N, Harrison LC. Cytokine-based human whole blood assay for the detection of antigen-reactive T cells. *J Immunol Methods* 1995: **186**: 37–46.
20. Thomson SA, Elliott SL, Sherritt MA, Sproat KW, Coupar BE, Scalzo AA et al. Recombinant polyepitope vaccines for the delivery of multiple CD8 cytotoxic T cell epitopes. *J Immunol* 1996: **157**: 822–6.
21. Thomson SA, Sherritt MA, Medveczky J, Elliott SL, Moss DJ, Fernando GJ et al. Delivery of multiple CD8 cytotoxic T cell epitopes by DNA vaccination. *J Immunol* 1998: **160**: 1717–23.
22. Woodberry T, Gardner J, Mateo L, Eisen D, Medveczky J, Ramshaw IA et al. Immunogenicity of a human immunodeficiency virus (HIV) polytope vaccine containing multiple HLA A2 HIV CD8(+) cytotoxic T-cell epitopes. *J Virol* 1999: **73**: 5320–5.
23. Mateo L, Gardner J, Chen Q, Schmidt C, Down M, Elliott SL et al. An HLA-A2 polyepitope vaccine for melanoma immunotherapy. *J Immunol* 1999: **163**: 4058–63.
24. Liu WJ, Liu XS, Zhao KN, Leggatt GR, Frazer IH. Papillomavirus virus-like particles for the delivery of multiple cytotoxic T cell epitopes. *Virology* 2000: **273**: 374–82.
25. Le TT, Drane D, Malliaros J, Cox JC, Rothel L, Pearse M et al. Cytotoxic T cell polyepitope vaccines delivered by ISCOMs. *Vaccine* 2001: **19**: 4669–75.
26. Sette A, Livingston B, McKinney D, Appella E, Fikes J, Sidney J et al. The development of multi-epitope vaccines: epitope identification, vaccine design and clinical evaluation. *Biologicals* 2001: **29**: 271–6.
27. Livingston BD, Newman M, Crimi C, McKinney D, Chesnut R, Sette A. Optimization of epitope processing enhances immunogenicity of multiepitope DNA vaccines. *Vaccine* 2001: **19**: 4652–60.