OptiTope—a web server for the selection of an optimal set of peptides for epitope-based vaccines

Nora C. Toussaint* and Oliver Kohlbacher

Division for Simulation of Biological Systems, Wilhelm Schickard Institute for Computer Science, Center for Bioinformatics Tübingen, Eberhard-Karls-Universität Tübingen, Germany

ABSTRACT

Epitope-based vaccines (EVs) have recently been attracting significant interest. They trigger an immune response by confronting the immune system with immunogenic peptides derived from, e.g. viral- or cancer-related proteins. Binding of these peptides to proteins from the major histocompatibility complex (MHC) is crucial for immune system activation. However, since the MHC is highly polymorphic, different patients typically bind different repertoires of peptides. Furthermore, economical and regulatory issues impose strong limitations on the number of peptides that can be included in an EV. Hence, it is crucial to identify the optimal set of peptides for a vaccine, given constraints such as MHC allele probabilities in the target population, peptide mutation rates and maximum number of selected peptides. OptiTope aims at assisting immunologists in this critical task. With OptiTope, we provide an easy-to-use tool to determine a provably optimal set of epitopes with respect to overall immunogenicity in a specific individual (personalized medicine) or a target population (e.g. a certain ethnic group). OptiTope is available at http://www.epitoolkit.org/optitope.

INTRODUCTION

Vaccines rank among the greatest achievements of modern medicine. They utilize the adaptive part of the immune system to prevent infections as well as to fight chronic diseases and cancer.

A vital event in the triggering of adaptive immunity is the recognition of antigen-derived peptides bound to major histocompatibility complex (MHC) class I and II molecules by T-cell receptors. Since the MHC is highly polymorphic, each individual possesses a set of MHC class I and class II molecules of differing specificities, i.e. different patients typically bind different repertoires of peptides. Peptides capable of causing an immune response are called epitopes. These epitopes form the basis of so-called epitope-based vaccines (EVs). A mixture of well-chosen epitopes can evoke an immune response precisely directed at conserved and highly immunogenic regions of several antigens. Key criteria for selection are, e.g. overall immunogenicity, tolerance for antigenic mutations, population coverage and antigen coverage. Due to the manifold advantages of EVs, discussed in detail in a recent review (1), and their applicability in personalized medicine, they have recently been attracting significant interest.

A crucial step in the design of an EV is the selection of the epitopes: which set of epitopes yields the best immune response in a given population or individual? Constraining factors are economical and regulatory issues, which impose strong limitations on the number of peptides that can be included in an EV. This renders the epitope selection an interesting optimization problem. Nevertheless, this critical task is typically performed manually. Several computational approaches have been published (2–4). In a recent paper (4), we proposed a mathematical framework to find a provably optimal set of epitopes for an EV. Given a set of predicted or experimentally determined epitopes, the framework efficiently identifies the set most likely to elicit a broad and potent immune response in the target population.

Based on a specific application of this framework, OptiTope aims at assisting immunologists in the critical task of epitope selection. It is an easy-to-use web-based tool to efficiently determine an optimal set of epitopes in a specific individual or a target population.

RELATED WEB SERVERS

Recently, two web-based tools aiding in vaccine design have been published. Hotspot Hunter (5) identifies immunological hotspots on pathogenic proteins. The Mosaic Vaccine Tool Suite (6) provides a set of web-based tools for designing artificial recombinant proteins to be used in T-cell vaccines. While the Mosaic Vaccine Tool Suite focuses on a completely different type of vaccine,

---

*To whom correspondence should be addressed. Tel: +49-7071-29 70460; Fax: +49-7071-29 5152; Email: toussaint@informatik.uni-tuebingen.de

Hotspot Hunter precedes OptiTope in the EV design pipeline. The identified peptides are suited as input for OptiTope.

MATERIALS AND METHODS

Mathematical Framework

OptiTope is based on a recently proposed mathematical framework for the selection of an optimal set of peptides for epitope-based vaccines (4). A brief outline of this framework is given below.

The search for an optimal peptide set for an EV is interpreted as an optimization problem: out of a given set of candidate epitopes, choose a subset which, out of all subsets meeting certain requirements for a good vaccine (e.g. mutation tolerance, population coverage), displays maximum overall immunogenicity. Overall immunogenicity is defined to be the immunity induced in the target population and the following mathematical abstraction is proposed: given a set of epitopes and a set of MHC alleles, i.e. a target population, the overall immunogenicity of the epitope set is assumed to be comprised of the immunogenicities of its components with respect to the different MHC alleles. Furthermore, the probability of an MHC allele to occur within the target population directly affects the allele's contribution to the overall immunogenicity. (In this context, probability is commonly referred to as frequency.) A common allele weighs more than an uncommon allele. This yields a mathematical interpretation of overall immunogenicity as weighted sum over immunogenicities of epitopes. Formulation of this optimization problem as an integer linear program (7) allows finding the optimal peptide set quite efficiently.

The underlying assumption of independence and additivity of immunogenicities of individual epitopes is, of course, a simplification. However, lacking a more sophisticated model of the interplay of multiple epitopes in the induction of an immune response, this assumption represents the state-of-the-art. For a more detailed discussion of this problem and its implications, we refer to (4).

Target population or individual

Given the mathematical background of OptiTope, a target population or individual is sufficiently described by a set of MHC alleles and their respective probabilities. The NCBI dbMHC database (http://www.ncbi.nlm.nih.gov/gv/mhc) contains data on MHC allele probabilities in various human populations and geographic areas (8). These data were retrieved and made available to the OptiTope users in order to facilitate the input of a target population.

Epitope immunogenicity

In order to increase the usability, OptiTope needs to incorporate a method to predict immunogenicity. Unfortunately, the prediction of immunogenicity is a rather challenging problem and to our knowledge no sufficiently accurate solution exists. However, in (9), Sette *et al.* have demonstrated a correlation between immunogenicity and MHC class I binding affinity. It is therefore reasonable [and also very common (e.g. 3,10)] to use MHC class I binding affinity prediction methods for the prediction of immunogenicity. OptiTope employs three widely used MHC class I binding affinity prediction methods, namely BIMAS (11), SYFPEITHI (12) and SVMHC (13). Prediction methods for MHC class II will be included in the near future. Predictions of methods that are not included in OptiTope, e.g. NetMHCpan (14) for MHC class I or ProPred (15) for MHC class II, can be utilized via the third input type: a table of epitopes and their immunogenicities with respect to specific MHC alleles.

Since a positive prediction score does not necessarily imply immunogenicity, a threshold is required to separate non-immunogenic from immunogenic epitopes. OptiTope offers three kinds of thresholds: user-defined, percentage (16) and halfmax (16). The percentage thresholds are calculated based on a large set of naturally occurring peptides. Using, e.g. the 1% threshold, 1% of these peptides would be classified as immunogenic. The halfmax threshold corresponds to half of the maximal possible prediction score.

Implementation

OptiTope is incorporated into the website SBS EpiToolKit (16), which is based on the Zope application server (http://www.zope.org), and the content management system Plone (http://plone.org). For the user interface, we employ dynamic HTML with CSS and JavaScript. Python scripts are used for data validation and processing. OptiTope was thoroughly tested for compatibility with the popular web browsers Mozilla Firefox (version 3.0.5) and Microsoft Internet Explorer (version 7).

OptiTope uses the GNU Linear Programming Kit GLPK (version 4.32, http://www.gnu.org/software/glpk) and the GNU MathProg modeling language GMPL to formulate and solve the optimization problems.
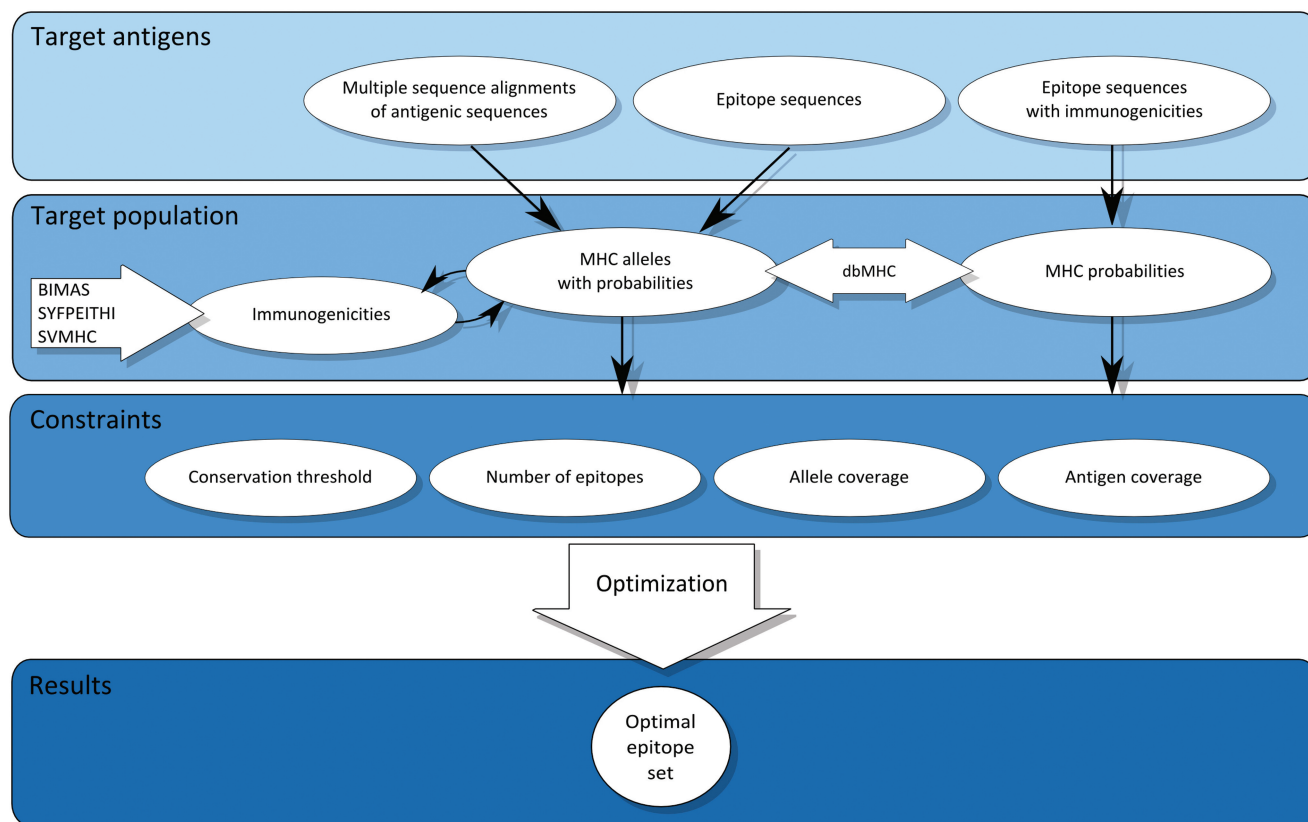
WEB INTERFACE

OptiTope requires the following input data: (i) sequences of known antigens, (ii) a target population, i.e. MHC alleles and corresponding probabilities and (iii) the user's requirements on the epitope set to be selected. The information given by the user is transformed into an optimization problem. If this problem is feasible, OptiTope will return an optimal set of epitopes along with additional information on their respective contribution to the overall immunogenicity. Otherwise, OptiTope will propose changes to the user's requirements that might yield a feasible optimization problem. The structure of the web interface is depicted in Figure 1.

An introductory tutorial is provided on the OptiTope home page to assist new users in learning how to use the web server.

Conceptual design

For ease of use, the web interface is divided into four steps: three input steps and one output step. Step-by-step, the user is asked to enter the required data.

Figure 1. Structure of the OptiTope web interface. OptiTope is divided into four parts: target antigens, target population, constraints and results. (i) Target antigens: three different formats of antigenic sequences can be entered: (a) a list of multiple sequence alignments of target-specific antigens, (b) a list of epitopes and (c) a table of epitopes with (experimentally determined) immunogenicities with respect to specific alleles. (ii) Target population: the information required to specify the target population (MHC alleles and corresponding probabilities) depends on the chosen input format. (iii) Constraints: OptiTope offers a set of constraints, which can be modified or excluded by the user. (iv) Results: if feasible, OptiTope presents an optimal set of epitopes.

Navigation through the individual steps is guided by a navigation bar at the top of each site. The navigation bar indicates the current step and contains corresponding instructions. Furthermore, it provides access to a more detailed help page. In order to keep the page layout clear, settings and options are hidden from the user by default. They can be accessed via the advanced options button underneath the navigation bar.

Target sequences

In the first step, the sequences of known target-specific antigens are entered. They can either be pasted directly or uploaded as a file. Three different formats are accepted: (i) a list of multiple sequence alignments (MSAs) in FASTA format, (ii) a list of epitopes of equal length, one epitope per line and (iii) a table of epitopes and their (experimentally determined) immunogenicities with respect to specific MHC alleles. Higher immunogenicity values ought to indicate stronger immunogenicities. Antigenic sequences entered as MSAs will be converted into consensus sequences. From these sequences, all peptides of a given length will be derived and will be considered as candidate epitopes. The user can adjust the peptide length to be applied via the advanced options.

Target population

In the second step, information on the target population has to be entered. This step is subdivided into two queries. The user is queried for (i) the MHC alleles to consider (if they have not already been entered in the previous step) and (ii) for their probabilities in the target population.

(i) MHC alleles can be selected by population or geographic area based on data (8) retrieved from the NCBI dbMHC database (http://www.ncbi.nlm.nih.gov/gv/mhc). The corresponding probabilities will be employed for the next query. Alternatively, the MHC alleles can be selected manually from an expandable allele tree (16) or by pasting a list of alleles.

(ii) In this step, a list of the selected MHC alleles along with probabilities (default values or values retrieved from the NCBI dbMHC database, respectively) is given. These probabilities can either be modified manually or they can be replaced by population or geographic area-specific probabilities from the NCBI dbMHC database via the advanced options. Individual MHC alleles can be excluded from further processing. Furthermore, low probability MHC

alleles can be excluded from the epitope selection process via a filter in the advanced options section.

If the user has not entered the immunogenicities of the candidate epitopes together with the target sequences, OptiTope will employ a prediction method to determine the respective immunogenicities. The prediction method to be employed can be selected via the advanced options.

Constraints

In the third step, the user is queried for the requirements on the epitope set to be selected. Depending on the data that have been entered in the previous steps (a summary of these data is given), a list of suitable constraints is displayed. The user can (de)select and modify these constraints. Potential constraints are:

- Maximum number of epitopes to select. This constraint defines the maximum number of epitopes OptiTope should select. It is the only obligatory constraint.
- Minimum epitope conservation. This constraint ensures that only epitopes that fulfill a user-defined conservation requirement will be considered.
- Minimum number of alleles to cover. An MHC allele is considered to be covered by an epitope set, if one of the epitopes is sufficiently immunogenic with respect

to the allele. If this constraint is selected, the optimal set of epitopes will be immunogenic with respect to the specified number of alleles or more.
- Minimum number of antigens to cover. An antigen is considered to be covered by an epitope set, if one of the epitopes is derived from this antigen. This constraint guarantees that the optimal epitope set will include epitopes from a specified number of antigens or more.

The advanced options offer the possibility to set an immunogenicity threshold, i.e. a minimum immunogenicity score required for a peptide to be considered immunogenic with respect to a specific allele. Only peptides which score above this threshold for at least one MHC allele will be considered during epitope selection.

Results

The results page gives a summary of the input data and the selected constraints as well as the results of the optimization.

If the optimization problem is feasible, a table containing the optimal set of epitopes will be displayed (Figure 2). For every epitope in the set the following information is given: its fraction of the overall immunogenicity, a list of the MHC alleles it covers and, if antigen information was given, the corresponding
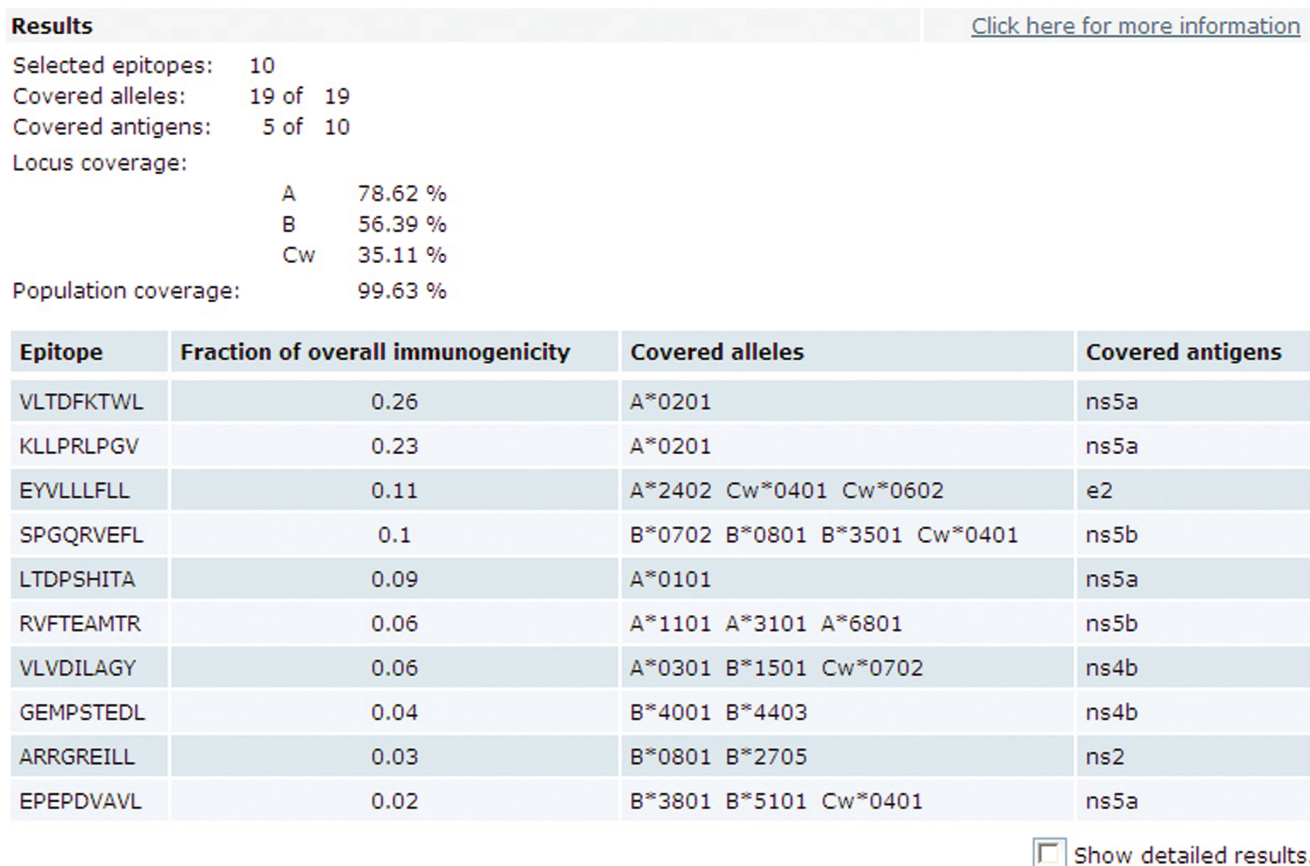
Results Click here for more information

Selected epitopes: 10
Covered alleles: 19 of 19
Covered antigens: 5 of 10
Locus coverage:

A	78.62 %
B	56.39 %
Cw	35.11 %

Population coverage: 99.63 %

Epitope	Fraction of overall immunogenicity	Covered alleles	Covered antigens
VLTDFKTWL	0.26	A*0201	ns5a
KLLPRLPGV	0.23	A*0201	ns5a
EYVLLLFLL	0.11	A*2402 Cw*0401 Cw*0602	e2
SPGQRVEFL	0.1	B*0702 B*0801 B*3501 Cw*0401	ns5b
LTDPSHITA	0.09	A*0101	ns5a
RVFTEAMTR	0.06	A*1101 A*3101 A*6801	ns5b
VLVDILAGY	0.06	A*0301 B*1501 Cw*0702	ns4b
GEMPSTEDL	0.04	B*4001 B*4403	ns4b
ARRGREILL	0.03	B*0801 B*2705	ns2
EPEPDVAVL	0.02	B*3801 B*5101 Cw*0401	ns5a

☐ Show detailed results.

Figure 2. A screenshot from the results page of OptiTope.

antigens. The user can switch to a more detailed results table, which contains additional information on epitope conservation and immunogenicities. Information on the size of the selected set, the number of covered alleles and on the number of covered antigens, if applicable, is displayed above the table. Furthermore, the coverage of each of the given MHC loci and the corresponding population coverage are given. (If locus A has a coverage of 75%, the probability of an individual from the target population carrying a covered allele at locus A is 75%. A population coverage of 80% corresponds to a probability of 80% for an individual from the target population to carry at least one of the covered alleles.) The results can be downloaded. A choice of two file formats is given: XLS (MS Excel) and CSV (comma separated values). For typical problem sizes, OptiTope finds an optimal set of peptides within seconds. Nevertheless, the user can choose to be notified of the completion of the request via e-mail.

If the optimization problem is infeasible, meaning that no set of epitopes from the given antigenic sequences fulfills all requirements, a basic analysis of the problem is performed. Based on this analysis, OptiTope suggests constraint modifications that might result in a feasible problem. If the basic analysis does not yield a possible explanation for the infeasibility, OptiTope will suggest to deselect individual constraints or to increase the number of epitopes to be selected.

CASE STUDY

In order to demonstrate the performance of OptiTope, we used it to select epitopes suitable for a hepatitis C virus (HCV) vaccine for the European population. An HCV dataset from (4) was utilized as antigenic sequence input. It consists of 10 MSAs corresponding to 10 different HCV proteins from four different strains, totaling in 4054 antigenic sequences. (This dataset is provided as an example dataset on the web server.) Default settings were used, i.e. immunogenicity prediction using BIMAS, a minimum conservation of 20%, coverage of at least 5 out of 10 antigens and of at least 10 out of 19 alleles. Within a few seconds, OptiTope returns a set of 10 epitopes covering 5 antigens and 10 alleles, yielding a locus coverage of ~54% for locus A, ~20% for locus B and ~35% for locus C. The corresponding population coverage is 94.28%. Three of these epitopes are known HCV epitopes and can be found in the Immune Epitope Database (IEDB, release 2008_4_1_3_28) (17). Another three epitopes are contained in known longer epitopes. Increasing allele coverage to 19, i.e. 100%, increases locus coverage for locus A to ~79% and to ~56% for locus B. Population coverage is increased to 99.63% (Figure 2). The selected set of epitopes includes five known HCV epitopes and another two epitopes contained in known longer epitopes. Again, this set of epitopes is found within a few seconds.

CONCLUSION

With OptiTope, we provide an easy-to-use tool that assists immunologists in designing EVs. Given a set of antigenic sequences of interest, a target population and special requirements of the user, OptiTope efficiently determines an optimal set of epitopes. To our knowledge, OptiTope is the first web-based approach for optimal vaccine design.

Currently, OptiTope only offers immunogenicity predictions for MHC class I, i.e. the only way to include MHC class II in the selection process is via the third input type: a table of epitopes and their immunogenicities with respect to specific MHC alleles. We plan to add MHC class II predictions and further MHC class I prediction methods in the future. A refinement of the analysis of infeasible problems in order to provide the user with more detailed information is also intended. Furthermore, we will enhance the results page by linking selected epitopes that can be found in the IEDB (17) to the corresponding IEDB site.

Due to the lack of a more sophisticated model of immunogenicity, OptiTope is forced to employ a commonly used additive model as well as prediction methods for MHC binding instead of immunogenicity. However, this does not pose fundamental limitations to the method.

REFERENCES

1. Purcell,A.W., McCluskey,J. and Rossjohn,J. (2007) More than one reason to rethink the use of peptides in vaccine design. *Nat. Rev. Drug Discov.*, **6**, 404–414.
2. Groot,A.S.D., Marcon,L., Bishop,E.A., Rivera,D., Kutzler,M., Weiner,D.B. and Martin,W. (2005) HIV vaccine development by computer assisted design: the GAIA vaccine. *Vaccine*, **23**, 2136–2148.
3. Vider-Shalit,T., Raffaeli,S. and Louzoun,Y. (2007) Virus-epitope vaccine design: informatic matching the HLA-I polymorphism to the virus genome. *Mol. Immunol.*, **44**, 1253–1261.
4. Toussaint,N.C., Dönnes,P. and Kohlbacher,O. (2008) A mathematical framework for the selection of an optimal set of peptides for epitope-based vaccines. *PLoS Comput. Biol.*, **4**, e1000246.
5. Zhang,G.L., Khan,A.M., Kellathur,N.S., Heiny,A.T., Lee,K.X., Kwoh,C.K., August,J.T. and Brusic,V. (2008) Hotspot Hunter: a computational system for large-scale screening and selection of candidate immunological hotspots in pathogen proteomes. *BMC Bioinformatics*, **9 (Suppl. 1)**, S19.
6. Thurmond,J., Yoon,H., Kuiken,C., Yusim,K., Perkins,S., Theiler,J., Bhattacharya,T., Korber,B. and Fischer,W. (2008) Web-based design and evaluation of T-cell vaccine candidates. *Bioinformatics*, **24**, 1639–1640.
7. Schrijver,A. (1986) *Theory Of Linear And Integer Linear Programming*., John Wiley & Sons, New York.
8. Meyer,D., Singe,R.M., Mack,S.J., Lancaster,A., Nelson,M.P., Erlich,H., Fernandez-Vina,M. and Thomson,G. (2007) Single locus

polymorphism of classical HLA genes. In Hansen,J.A. (ed.), *Immunobiology of the Human MHC: Proceedings of the 13th International Histocompatibility Workshop and Conference,* Vol. I, IHWG Press, Seattle, WA, pp. 653–704.

9. Sette,A., Vitiello,A., Reherman,B., Fowler,P., Nayersina,R., Kast,W.M., Melief,C.J., Oseroff,C., Yuan,L., Ruppert,J. *et al.* (1994) The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J. Immunol.*, **153**, 5586–5592.

10. Groot,A.S.D. and Moise,L. (2007) Prediction of immunogenicity for therapeutic proteins: state of the art. *Curr. Opin. Drug Discov. Devel.*, **10**, 332–340.

11. Parker,K.C., Bednarek,M.A. and Coligan,J.E. (1994) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J. Immunol.*, **152**, 163–175.

12. Rammensee,H., Bachmann,J., Emmerich,N.P., Bachor,O.A. and Stevanovic,S. (1999) SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics*, **50**, 213–219.

13. Dönnes,P. and Kohlbacher,O. (2006) SVMHC: a server for prediction of MHC-binding peptides. *Nucleic Acids Res.*, **34**, W194–W197.

14. Nielsen,M., Lundegaard,C., Blicher,T., Lamberth,K., Harndahl,M., Justesen,S., Røder,G., Peters,B., Sette,A., Lund,O. *et al.* (2007) NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and –B locus protein of known sequence. *PLoS ONE*, **2**, e796.

15. Singh,H. and Raghava,G.P. (2001) ProPred: prediction of HLA-DR binding sites. *Bioinformatics*, **17**, 1236–1237.

16. Feldhahn,M., Thiel,P., Schuler,M.M., Hillen,N., Stevanovic,S., Rammensee,H.G. and Kohlbacher,O. (2008) EpiToolKit – a web server for computational immunomics. *Nucleic Acids Res.*, **36**, W519–W522.

17. Peters,B., Sidney,J., Bourne,P., Bui,H.H., Buus,S., Doh,G., Fleri,W., Kronenberg,M., Kubo,R., Lund,O. *et al.* (2005) The immune epitope database and analysis resource: from vision to blueprint. *PLoS Biol.*, **3**, e91.