ELSEVIER

Efficient discovery of immune response targets by cyclical refinement of QSAR models of peptide binding

Vladimir Brusic,* Kim Bucci,† Christian Schönbach,* Nikolai Petrovsky,‡ John Zeleznikow,§ and James W. Kazura†

**BIC-KRDL, Kent Ridge Digital Labs, Singapore*
†Division of Geographic Medicine, Case Western Reserve University and University Hospitals, Cleveland, OH, USA
‡Department of Medicine, Canberra Clinical School, University of Sydney, Woden ACT, Australia
§Department of Computer Science and Computer Engineering, Faculty of Science and Technology, La Trobe University, Bundoora 3083, Victoria, Australia

Peptides that induce and recall T-cell responses are called T-cell epitopes. T-cell epitopes may be useful in a subunit vaccine against malaria. Computer models that simulate peptide binding to MHC are useful for selecting candidate T-cell epitopes since they minimize the number of experiments required for their identification. We applied a combination of computational and immunological strategies to select candidate T-cell epitopes. A total of 86 experimental binding assays were performed in three rounds of identification of HLA-A11 binding peptides from the six pre-erythrocytic malaria antigens. Thirty-six peptides were experimentally confirmed as binders. We show that the cyclical refinement of the ANN models results in a significant improvement of the efficiency of identifying potential T-cell epitopes. © *2001 by Elsevier Science Inc.*

INTRODUCTION

T lymphocytes (T cells) mediate immune system protection against infection and cancer. To perform this function requires that they continuously survey body tissues for the presence of foreign antigens derived from viruses, bacteria, parasites, or tumors. T cells recognize foreign antigenic peptides displayed on the surface of target cells by specialized antigen receptors called major histocompatibility complex (MHC) molecules.[1,2] MHC molecules are divided into class I and class II; class I molecules bind peptides produced by degradation of intracellular proteins and class II molecules bind peptides of extracellular origin. MHC molecules thereby display contents of cells for surveillance by T cells. Presentation via the MHC-class-I pathway is dependent upon the affinity of peptide for MHC and the quantity of peptide available, as determined by intracellular antigen processing. Whereas MHC class-I-bound peptides elicit mainly cytotoxic (cell-destructive) responses, MHC class-II-bound peptides have a crucial role in immune regulation for the initiation, enhancement, and suppression of immune responses. Peptides that induce and recall T-cell responses are called T-cell epitopes.

More than 880 allelic variants of human leukocyte antigen (HLA, or human MHC) have been characterized.[3] HLA molecules encoded by allelic variants bind different peptide sets, characterized by specific binding motifs.[4] These sets can be overlapping or distinct. Each individual can potentially express six HLA class I and eight class II molecules. The majority of peptides binding HLA molecules are 8–11 (class I) or 11–20 (class II) amino acids long. The theoretical number of nine-amino acids-long (9-mer) peptides comprised of naturally occurring amino acids is 5.12×10^{11}. The identification of specific T-cell epitopes is thus a combinatorial problem. Efficient identification of T-cell epitopes necessitates combining experimental and computational methods. MHC-binding peptides that function as T-cell epitopes are promising candidates for new vaccines and immunotherapies.[5,6] For example, immunization with a single modified gp100 antigen peptide plus IL-2 produced objective cancer regression in 38% of patients with melanoma.[6]

In malaria, clinical protection was related to vigorous T-cell proliferation responses to MSP-1 and MSP-2 antigenic peptides,[7] suggesting that T-cell epitopes therein may be useful in a subunit vaccine against malaria. Even within one individual, T-cell responses against malaria antigens involve multiple

Color Plates for this article are on page 467.

Corresponding author: V. Brusic, BIC-KRDL, Kent Ridge Digital Labs, 21 Heng Mui Keng Terrace, Singapore 119613.

E-mail address: vladimir@krdl.org.sg (V. Brusic)

Journal of Molecular Graphics and Modelling 19, 405–411, 2001

655 Avenue of the Americas, New York, NY 10010

1093-3263/01/$–see front matter
PII S1093-3263(00)00099-1

combinations of different HLA-class I molecules and malaria peptides, as reported in a recent DNA vaccination study.[5] Therefore, epitope mapping of malaria antigens is important for malaria vaccine design. Epitope mapping is hampered, however, by the large number of experiments needed for systematic epitope determination. Our goal was the identification of malaria T-cell epitopes in the human population from a *Plasmodium falciparum* endemic coastal region of Papua New Guinea, which has a HLA-A11 frequency of ≥44%. Because of the large number of potential 9-mer peptides contained within the sequence of the even single malaria antigens of interest, we applied a combination of computational and immunological strategies to select candidate T-cell epitopes. This involved a methodology known as quantitative structure–activity relationships determination (QSAR), which correlates structural or property descriptors of compounds with their activities. This required building computer models for identification of HLA-A11 binding peptides from malarial antigens.

Computer models that simulate peptide binding to MHC are useful for selecting candidate T-cell epitopes since they minimize the number of experiments required for their identification. Data-driven models include quantitative matrices,[8–10] multivariate methods such as partial least squares (PLS),[11] artificial neural networks (ANNs),[12,13] and hidden Markov models (HMMs).[14] Structure-driven models based on molecular modelling[15] and rule-based models using binding motifs[16] have also been used for identification of potential MHC binding peptides. ANN models were reported as most accurate.[17,18] ANN models also have the advantage that they can be easily refined by the inclusion of new data and training with the expanded data sets. The requirements for the predictive models include high specificity SP = TN/(TN + FP) (TN signifies true negatives; FP, false positives; negatives correspond to nonbinders and positives correspond to binders), and high sensitivity SE = TP/(TP + FN) (TP, true positives; FN, false negatives). High specificity of the model, based on the low number of false positives, results in a modest number of experiments required to confirm predicted binders and thereby reduces the cost of identifying T-cell epitopes for vaccine development. High sensitivity of a predictive model ensures broad coverage as it corresponds to the percentage of experimental (or "true") binders correctly predicted.

The number of HLA class I binding peptides for a given HLA allele is estimated on average to represent 0.1–5% of the overlapping 9-mer peptides that span the entire length of a protein.[19] The low frequency of potential binding peptides within a protein antigen means that determining peptide binders using traditional experimental procedures can be laborious and slow. In a hypothetical case where a predictive model has a SE = 1 and SP = 0.95 and the prevalence of binders is 1%, experimental confirmation will require six experiments for the confirmation of each true binder. Although it has been previously proposed that computer models will be required to help scientists understand the full implication of their data, these models will necessarily be based on experimental data that can then be used to assist in the design of additional experiments[20], this approach has however not been practically demonstrated to date. We hereby show that the cyclical refinement of the ANN models results in a significant improvement of the efficiency of identifying potential T-cell epitopes. The cyclical process of model refinement/experiment suggestion is described in Figure 1.

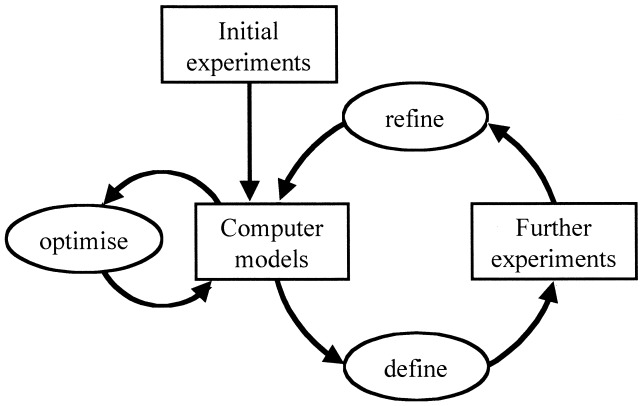

Figure 1. Cyclical refinement of computer models used to define further experiments. The larger loop represents alternating computational and experimental steps. The smaller loop represents an optional model refinement using computational optimization methods.

MATERIALS AND METHODS

Experimental Methods

Peptides Peptides were synthesized by fmoc chemistry, purified to ≥95% homogeneity by reverse phase HPLC (Chiron Mimotopes, Raleigh, NC), suspended in PBS (pH 7) to a concentration of 1 mg/ml and stored at −20°C until use. Putative binding peptides were predicted from *Plasmodium falciparum* protein sequences (GenBank accession number, sequence length in parentheses): liver stage-specific antigen-1 LSA-1 (A45592, 1909 AA), trombospondin related anonymous protein (also known as the sporozoite surface protein 2) TRAP/SSP2 (X13022, 559 AA), circumsporozoite protein CSP (AAA29544, 432 AA), sporozoite and liver stage antigen SALSA (AAB36334, 83 AA), sporozoite surface antigen STARP (CAA81224, 604 AA), and glutamate rich protein GLURP (AAA50613, 1262 AA). The peptide corresponding to amino acids 9–17 (STAPPAHGV) from MUC-1 mucin protein was used as positive control for measuring HLA-A11 surface expression since it has previously been shown to bind to HLA-A11 with high affinity.[21]

HLA-A11 stabilization assay Peptide binding affinity to HLA-A11 (A*1101 subtype) was estimated using an assay in which antibody that recognizes the conformation of stabilized class I dimers on the cell surface are used as a probe.[22,23] The .174/T2 (T2) cell line was obtained by fusion of the peptide transporter mutant .174 LCL with the T-cell line CEM.[24] For in vitro stabilization assays, T2/A11 cells (2×10^5) were incubated for 18 h in serum-free AIM-V medium (Life Technologies, Inc., Grand Island, NY) containing human β_2-microglobulin (5 μg/ml) (Calbiochem, La Jolla, CA), and peptide (100 μM) or no peptide. The cells were collected for FACS analysis to quantify HLA-A11 cell surface expression. Antibody to HLA-A11 (A11.1M, mouse IgG_3) (American Type Culture Collection, Rockville, MD) was added to cell preparations followed by FITC-conjugated goat anti-mouse IgG (Life Technologies, Rockville, MD). Binding results were expressed as a score based on the degree of binding relative to the positive control MUC-1 peptide. The score was calculated

using the equation: 10*MFIt/MFIc (MFIt and MFIc indicate, respectively, the mean fluorescence intensity of test and positive control peptide). The results were normalized to a scale of 3–8, in which 3 indicates no binding and 8 the highest binding value, reflecting the values assigned to peptides for the ANN training (8 for high-, 6 for moderate-, 4 for low-binding affinity, and 3 for nonbinding threshold). The binding assay score for each peptide was calculated from the mean of three tests; the binding of the positive control peptide was performed with all sets of experiments. The MFI for cells incubated without peptide ranged from 1.2 to 9.4 while the MFI for the cells incubated with the mucin positive control peptide ranged from 4.5 to 35. Variation was mainly related to the voltage settings of the flow cytometer used in different experiments.

COMPUTATIONAL METHODS

Design of Experiments

Peptides were selected using ANN-based predictive models for experimental measurement of HLA-A11 binding. This study targeted 9-mer peptides because this is the most common length of A11-binding peptides.[4,25] Peptides were selected in three discrete steps. First, all 9-mer peptides spanning the region 88-115 of the liver-stage specific LSA-1 protein were synthesized and tested for binding to HLA-A11. This region was selected because it was previously shown to contain at least one T-cell epitope.[26] Second, peptides selected by an ANN-based prediction using previously available peptide data were synthesized and tested for binding to HLA-A11. Finally, ANN models used in the second round were refined by training using the expanded data set, which included binding results obtained from the first two rounds of discovery and additional data. Peptides predicted from the initial ANN round were considered for experimental testing if their prediction score was >6 (LSA-1, SALSA, CSP, and TRAP/SSP2) or >7 (STARP and GLURP). In the final round of discovery, peptides with a prediction score of >5 were considered for experimental testing. A small number of peptides were selected using other selection criteria (e.g., those previously reported in the literature as binders).

Data

After eliminating duplicate peptides from repeat regions, the six antigens of interest contained 3,127 9-mer peptides–the target set. Peptide sequences used for the initial training of ANNs were extracted from the MHCPEP database[25] and from a set of known nonbinding peptides (V. Brusic, unpublished data). The initial set of ANN-training data comprised 98 binders (9 low-, 46 moderate-, and 43 high-affinity) and 145 nonbinding peptides to HLA-A11. ANN training data was characterized by a low redundancy level. The refinement set comprised 123 binders (11 low-, 64 moderate-, and 48 high-affinity) and 203 nonbinders. Additional peptides used for the refinement cycle were obtained by testing those predicted in the first two rounds of discovery and in published information.[27]

Predictive Models

ANN models were derived as previously described.[12,13] Briefly, PlaNet software[28] was used to train a fully connected three-layer feed forward ANN. The training set consisted of 9-mer binding and nonbinding peptides. The ANN architecture comprised 180 input units, corresponding to the binary representation of 9-mer peptides, two hidden layer units, and a single output unit (Color Plate 1). The fuzzy measures of high-, moderate-, low-, and no-affinity were defined as previously described[25], representing a rough grouping of the binding affinities of peptides. The learning algorithm was error back-propagation. Training was performed for 300 cycles. The respective values for momentum and learning rate were 0.5 and 0.2. Each prediction result was calculated as the average of four independent runs. The output prediction scores of the PlaNet software ANNs range includes integers 0–10. Binding scores used for ANN training were 1, 4, 6, 7, and 8 for no-, low-, moderate-, moderate-to-high-, and high-affinity binders, respectively.

Optimization Strategies

The training set consisted of a single copy of each peptide and the associated measure of binding affinity. Two strategies for model optimization were tested: the reinforcement and bonus strategies. The first step in each strategy was the identification of "lost peptides": true positives from the previous cycle that became false negatives in the subsequent cycle of prediction. Both optimization strategies involved (a) identification of lost peptides, (b) generation of optimized training sets, (c) generation of optimized predictive models, and (d) assessment of the model performance. The optimized training sets generated using the reinforcement strategy comprised multiple instances of each of the lost peptides and a single instance of all other peptides. The training sets generated by the bonus strategy consisted of a single instance of each peptide in which the measure of the binding affinity for each peptide was increased relative to nonoptimized sets.

In the reinforcement strategy, three instances of each of the lost peptides were used. In the bonus strategy, the binding value was increased by one for each lost peptide. Optimization results were assessed using the experimental measurements of peptide binding from the final round of experimentation.

Comparison with Other Predictive Methods

Binding results for the peptides selected and experimentally tested in this study were compared with rankings derived by using either binding motif (Table 1) or a quantitative matrix scoring available at BIMAS on the World Wide Web (URL: http://www-bimas.cit.nih.gov/molbio/hla_bind/).[29] Binding motif ranking is derived from the number of primary anchor, auxiliary anchor, and preferred residues.

RESULTS

A total of 86 experimental binding assays were performed in three rounds of identification of HLA-A11 binding peptides from the six pre-erythrocytic malaria antigens of interest. Thirty-six peptides were experimentally confirmed as binders. We identified seven HLA-A11 binding peptides from LSA-1 (peptides 49, 60, 68, 90, 94, 105, and 1854), six from TRAP/SSP2 (90, 224, 232, 307, 414, and 508), four from SALSA (18, 19, 48, and 72), four from CSP (18, 331, 380, and 411), eleven

Table 1. Binding motif of HLA-A*1101 (from reference 4). This motif indicates amino acids that facilitate binding when present at particular positions. The primary anchor residue is given in bold and auxiliary anchors are underlined

	Relative Position								
	1	2	3	4	5	6	7	8	9
Anchor									**K**
Auxiliary anchors		*V* *I* *F* *Y*	*M* *L* *F* *Y* *I* *A*				*L* *I* *Y* *V* *F*		
Other preferred residues	A	T	N D E O	P G D E K	P I F V M	I V M		R K N E O	R

from GLURP (41, 57, 174, 346, 394, 419, 695, 696, 775, 1199, and 1250), and four from STARP (16, 57, 522, 584).

In the first round of discovery, using overlapping peptides from LSA-1 88-115, three HLA-A11 binding peptides were identified from twenty binding assays. In the second round, which was performed independently of the first, 34 target peptides predicted as candidate binders were selected for experiments along with TRAP 508 and two predicted nonbinders from SALSA. Ten peptides (including LSA-1 94 tested in the first round) were confirmed as HLA-A11 binders. In addition, HLA-A11 binding peptide TRAP 508-516, which was not selected by the ANN prediction, was tested experimentally and confirmed as a binder, as previously reported.[30] In the final round of discovery, 29 target sequences were selected from predictions using the refined ANN. Twenty-two peptides were experimentally confirmed as HLA-A11 binders. The number of experiments per confirmed binder was 6.7, 3.6, and 1.3 (success rates of 15%, 29%, and 76%, Color Plate 2), respectively, for each of the selection strategies: overlapping peptides, initial ANN, and refined ANN. A summary of results of confirmed binders is presented in Table 1.

The ANNs of the refinement cycle were trained with the peptide set that included peptides whose binding affinity was determined in the first two rounds of discovery. We expected that these ANNs would recall the scores with which they were trained. For example, LSA-1 60, 90, 94, and 105 should have produced output scores of >5 when presented to the refined ANNs. Interestingly, only six of fourteen binders (43%) determined in the first two rounds of discovery had a recall score of >5. The increase in sensitivity of the refined network was therefore offset by the loss of a large proportion of true binders (as determined using the initial ANN predictions). Two simple strategies for improving the refinement cycles were applied (see Materials and Methods section). The bonus strategy improved the recall, as nine of fourteen (64%) lost peptides were correctly recalled, with no loss of correctly predicted binders. With the reinforcement strategy, all of the previously determined binders were correctly recalled (recall score >5), but five of twenty-two (23%) true binders determined in the refinement cycle of discovery were lost. These results indicate that optimization strategies can further improve the accuracy of the models.

Analysis of the HLA-A*1101 binding motif (see Table 1) in the set of confirmed binders showed that three HLA-A11 binding peptides (8%) did not have a primary anchor K at position 9. The LSA-1 94 peptide has the preferred residue R at position 9, while CSP 18 and CSP 411 have nonmotif residues S or Y at the primary anchor position. Four of the binding peptides (11%) have K at position 9, and no other motif residues. Twelve binders (33%) do not have any motif-defined auxiliary anchor residues at positions 2, 3, or 7. Thirteen of fifty peptides determined as nonbinders in stabilization assays had a primary anchor K9 and at least one preferred residue. A large proportion of confirmed binders (42%) conform poorly to the proposed motif. Binding motif is therefore useful for coarse classification but not for accurate predictions.

Peptide binding scores selected by the BIMAS scoring matrix were compared with the set determined in this study (see Table 2). Nineteen of 36 experimentally confirmed HLA-A11 binders (53%) were ranked among the top 63 predicted binders by the matrix score. Five more binders were ranked among the top 99 predicted binders. The scoring matrix ranked 12 of the confirmed binders (33%) lower than top 99 peptides.

In summary, these results show that:

1. The cyclical refinement of the ANN models for prediction of MHC binding peptides significantly improves predictive power of the refined models.
2. A large proportion (33%) of peptides identified by ANN predictions and experimentally confirmed as binders do not comply with the proposed binding motif.
3. A binding motif is useful for initial classification but not for accurate predictions.
4. The refined model predictions, relative to the initial model, increased the number of true-positive binders (predicted and experimentally confirmed as binders) and decreased the number of false-positive binders (peptides predicted as binders and experimentally determined as nonbinders).
5. Binding matrices have relatively high predictive power, but relatively low sensitivity. Binding matrices do not identify a large proportion of binding peptides (33% in this study).
6. Optimization cycles can further improve the predictive performance of predictive models.
7. The strength of ANN-based predictions is that they are more sensitive than matrix-based predictions.

The disadvantage is that some true positive binders from the earlier predictions can be lost in the refinement cycles. The refinement cycle thus should be done with care.

DISCUSSION

This study is part of a larger invesitgation aimed at determination of HLA-A11 restricted T-cell epitopes encoded by preerythrocytic *P. falciparum* antigens. Peripheral blood mononuclear cells from malaria-exposed individuals have been used to test the ability of peptides determined as binders in stabilization

Table 2. Summary results for peptides confirmed as HLA-A11 binders. The binding index was calculated as described in Materials and Methods: score 8 represents high relative binding. 3 represents no binding. Peptide selection using the initial ANN (second round) was based on the selection threshold >6 (>7 for STARP). However peptides[a] LSA-1 68, and GLURP 174, 394, and 419 were not tested in the second round. Previously reported HLA-A11 binding peptide[b] TRAP/SSP2 508 was synthesised and tested although it was not predicted as binder. Peptides with the prediction threshold of >5 using refined ANN were synthesised and tested. The "ANN recall" shows prediction scores (refined network) for peptides used to train the ANN - the prediction scores might be different to scores used for training; experimental binders that had recall scores below the selection threshold (>5) are shown in bold[c]. The "Motif" column indicates the numbers of primary anchors - auxiliary anchors - preferred residues within a peptide (see table 1). "Rank by BIMAS" indicates the peptide rank among all 3127 peptides, calculated using the BIMAS HLA-A11 matrix (see Materials and Methods). The ranks given in bold[d] are the best candidate binders predicted by BIMAS

Antigen	Position	Peptide	Binding Index	Prediction Initial ANN[a]	Prediction refined ANN	ANN Recall[c]	Motif	Rank by BIMAS[d]
LSA-1	49	RINEEKHEK	8	5	6		1-1-3	**16-22**
	60	VLSHNSYEK	7	7		7	1-1-1	**32-34**
	68	KTKNNENNK	8	7[a]	6		1-0-2	**5-6**
	90	KNVSQTNFK	7	5		**4**	1-0-0	93-99
	94	QTNFKSLLR	8	7		7	0-1-3	**47-63**
	105	GVSENIFLK	8	3		6	1-2-2	**1**
	1854	KSLYDEHIK	8	3	6		1-1-0	>99
TRAP	90	VNVFSNNAK	7	9		**5**	1-0-0	>99
	224	KNVIGPFMK	7	5	6		1-1-0	**44-45**
	232	KAVCVEVEK	7	7		9	1-1-2	**29-31**
	307	RGDNSSVQK	8	4	7		1-1-2	>99
	413	IPYSPLPPK	8	6	6		1-1-1	**47-63**
	508	LACAGLAYK	8	5[b]		**5**	1-0-0	81-92
SALSA	18	PQFIWSAEK	6	6	6		1-1-1	>99
	19	QFIWSAEKK	6	7		6	1-2-1	65-72
	48	SANGKDDVK	6	6		**4**	1-0-2	81-92
	72	KVQEKVLEK	7	3	6		1-2-4	**2**
CSP	18	ALFQEYQCY	5	6	6		0-1-1	>99
	331	NANANNAVK	6	8		**5**	1-0-1	81-92
	380	RIKPGSANK	8	3	6		1-1-2	**15-22**
	411	SVFNVVNSS	6	8		**5**	0-2-2	>99
GLURP	41	RGNVTSNIK	8	5	6		1-0-1	>99
	57	KIIRGSNDK	7	6	6		1-2-0	**11-12**
	174	STEPFPNQK	8	9[a]	7		1-0-5	**23-28**
	346	HVQDHALPK	8	6	6		1-2-2	**4**
	394	PNIESFEPK	5	8[a]	6		1-1-1	>99
	419	IIDDVPSPK	7	7[a]	8		1-1-3	**47-63**
	695	QSNNEPSEK	6	8		6	1-0-2	>99
	696	SNNEPSEKK	5	9		**5**	1-0-4	>99
	775	NVSEVVEEK	6	6	6		1-1-4	**8-10**
	1199	IDFEGLSRK	6	6	6		1-1-2	>99
	1250	VIQVVSTIK	8	4	6		1-1-2	**47-63**
STARP	16	WTTLLYSNK	7	<3	7		1-0-2	**23-28**
	57	GGGYSAALK	8	3	6		1-0-0	>99
	522	SMINAYLDK	7	8		**5**	1-2-0	**15-22**
	584	EVIYKYVGK	7	<3	7		1-3-0	**29-31**

assays to stimulate T-cell IFN-γ production. Preliminary results show that peptides determined as binders are more likely to stimulate production of this cytokine, confirming their role as T-cell epitopes. Independent observations[13] have shown concordance between ANN predictions of HLA class II binding peptides and experimentally-determined T-cell epitopes. Therefore, the combination of ANN predictions with experimental confirmation of MHC peptide binding is useful for

efficient determination of potential T-cell epitopes. A small number of peptides predicted as binders by the ANN predictions were experimentally shown to be HLA-A11 nonbinders. Although stabilization assays appear to be a reasonable predictor of T-cell epitopes, some peptides that are potent T-cell epitopes are poor binders in stabilization assays.[31] These peptides may be restricted through other HLA alleles or HLA-A11 subtypes. On the other hand, since we trained ANNs with peptide data determined from a variety of experimental methods,[25] a possible explanation is that these models capture a subset of peptides that, for reasons that are as yet unclear, do not bind well in a particular binding assay. Various binding assays may produce different results for the same peptides.[31] Standardization of binding assays has been recognized as a major problem, and the forthcoming International Histocompatibility Workshop[32] (Whistler 2001) will have a session toward establishing standards for peptide binding assays and related computer modelling.

Many of the known T-cell epitopes do not conform to proposed binding motifs. Studies involving mouse H-2K^d (Mata et al.[33]) as well as K^b, D^d, L^d, and K^k (Apostolopoulos et al.[34]) showed that the binding affinity to MHC molecules of potent T-cell epitopes might be relatively low. Prilliman et al.[35] studied the HLA-B15 (*1501) binding motif using large quantities of HLA molecules. They determined that consensus motifs do not account for a significant number of HLA-B15 binding peptides, thus demonstrating that many of the binding peptides presented to cytotoxic T lymphocytes (CTL) cannot be predicted using class I consensus motifs as search criteria. Among 16 HLA-A*1101 restricted T-cell epitopes shown in the MHC motif listing[4], four peptides (GSPATWTTR, type IV collagen 51-59; STAPPAHGV, MUC-1 9-17; TLYCVHQRI, HIV-1 p17 84-92; and RIETRSARH, rubella virus capsid protein) do not contain the primary anchor residue. ANN-based predictive models can capture distinct subsets of binding peptides, including those of relatively low affinity. They can capture motifs of higher complexity than the linear functions described by quantitative matrices. ANN models therefore combine flexibility with the ability to capture complex patterns. They can be refined to a high accuracy, as demonstrated in this study, and therefore represent ideal models for the cyclical refinement of the discovery process shown in Figure 1. Cyclical refinement, however, must be done with care. If the refinement does not include appropriate optimizations, it is possible that the decreased sensitivity (through the loss of previously identified positives) may offset increase in specificity of predictions (through minimization of the number of false positives).

CONCLUSIONS

We have demonstrated that combining computer models and experimentation in a cyclical fashion can be applied successfully to the identification of immunologically interesting peptides. Furthermore, we identified the requirement for the optimization step (Figure 1) for efficient model refinement.

Laboratory experimentation in immunology, particularly that related to antigen recognition and determination of specific targets of immune response, has become a combinatorial challenge. Understanding the mechanisms of immune recognition and specificity and selection of immune response targets is necessary before this information can be applied systematically to the design of vaccines and immunotherapeutics. We anticipate that dynamic models of immune interactions that can absorb the ever-increasing amount of data generated in the field and self-improve with the accumulation of data and knowledge will become standard methodology in immunology research. Standardization and exploitation of the synergies of modelling and experimental methods provide an efficient means for large-scale epitope screening. This study provides a first-level guideline for cyclical refinement of computer models and their integration with laboratory experiments.

ACKNOWLEDGMENTS

T2 cells transfected with HLA-A*1101 (T2/A11) were provided by Dr. O.J. Finn, University of Pittsburgh School of Medicine, Pittsburgh, PA, with permission from Dr. M. Massucci of the Karolinska Institute, Stockholm, Sweden. We thank Mike Sramkoski of the Case Western Reserve University Cancer Center for providing advice and technical assistance with flow cytometry. This work was supported in part by the National Science and Technology Board of Singapore and a grant from the United States Public Health Service (AI36478).

REFERENCES

1 Pamer, E., and Cresswell, P. Mechanisms of MHC class I–restricted antigen processing. *Annu. Rev. Immunol.* 1998, **16**, 323–358
2 Cresswell, P. Assembly, transport, and function of MHC class II molecules. *Annu. Rev. Immunol.* 1994, **12**, 259–293
3 Lefranc, M.P., Giudicelli, V., Ginestoux, C., Bodmer, J., Muller, W., Bontrop, R., Lemaitre, M., Malik, A., Barbie, V., and Chaume, D. IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.* 1999, **27**, 209–212
4 Rammensee, H.G., Friede, T., and Stevanovic, S. MHC ligands and peptide motifs: first listing. *Immunogenetics* 1995, **41**, 178–228
5 Wang, R., Doolan, D.L., Le, T.P., Hedstrom, R.C., Coonan, K.M., Charoenvit, Y., Jones, T.R., Hobart, P., Margalith, M., Ng, J., Weiss, W.R., Sedegah, M., de Taisne, C., Norman, J.A., and Hoffman, S.L. Induction of antigen-specific cytotoxic T lymphocytes in humans by a malaria DNA vaccine. *Science* 1998, **282**, 476–480
6 Rosenberg, S.A., Yang, J.C., Schwartzentruber, D.J., Hwu, P., Marincola, F.M., Topalian, S.L., Restifo, N.P., Sznol, M., Schwarz, S.L., Spiess, P.J., Wunderlich, J.R., Seipp, C.A., Einhorn, J.H., Rogers-Freezer, L., and White, D.E. Impact of cytokine administration on the generation of antitumor reactivity in patients with metastatic melanoma receiving a peptide vaccine. *J. Immunol.* 1999, **163**, 1690–1695
7 Migot-Nabias, F., Luty, A.J., Ringwald, P., Vaillant, M., Dubois, B., Renaut, A., Mayombo, R.J., Minh, T.N., Fievet, N., Mbessi, J.R., Millet, P., and Deloron, P. Immune responses against Plasmodium falciparum asexual blood-stage antigens and disease susceptibility in Gabonese and Cameroonian children. *Am. J. Trop. Med. Hyg.* 1999, **61**, 488–494
8 Parker, K.C., Bednarek, M.A. and Coligan, J.E. Scheme for ranking potential HLA-A2 binding peptides based on

independent binding of individual peptide side-chains. *J. Immunol.* 1994, **152**, 163–175
9 Hammer, J., Bono, E., Gallazzi, F., Belunis, C., Nagy, Z. and Sinigaglia, F. Precise prediction of major histocompatibility complex class II-peptide interaction based on peptide side chain scanning. *J. Exp. Med.* 1994, **180**, 2353–2358
10 Mallios, R.R. Class II MHC quantitative binding motifs derived from a large molecular database with a versatile iterative stepwise discriminant analysis meta-algorithm. *Bioinformatics* 1999, **15**, 432–439
11 Rovero, P., Riganelli, D.D., Fruci, D.,Vigano, S., Pegoraro, S., Revoltella, R., Greco, G., Butler, R., Clementi, S. and Tanigaki, N. The importance of secondary anchor residue motifs of HLA class I proteins: a chemometric approach. *Molec. Immunol.* 1994, **31**, 549–554
12 Brusic, V., Rudy, G., and Harrison, L.C. Prediction of MHC binding peptides using artificial neural networks. In: Stonier, R., and Yu, X.H. (eds*), Complex Systems: Mechanism of Adaptation.* IOS Press/Ohmsha, 1994, pp. 253–260. http://www.csu.edu.au/ci/vol2/vbb/vbb.html
13 Honeyman, M.C., Brusic, V., Stone, N.L., and Harrison, L.C. Neural network-based prediction of candidate T-cell epitopes. *Nat. Biotechnol.* 1998, **16**, 966–969
14 Mamitsuka, H. Predicting peptides that bind to MHC molecules using supervised learning of hidden Markov models. *Proteins* 1998, **33**, 460–474
15 Schueler-Furman, O., Elber, R., and Margalit, H. Knowledge-based structure prediction of MHC class I bound peptides: a study of 23 complexes. *Fold. Des.* 1998, **3**, 549–564
16 Nijman, H.W., Houbiers, J.G., Vierboom, M.P., van der Burg, S.H., Drijfhout, J.W., D'Amaro, J., Kenemans, P., Melief, C.J., Kast, W.M. Identification of peptide sequences that potentially trigger HLA-A2.1-restricted cytotoxic T lymphocytes. *Eur. J. Immunol.* 1993, **23**, 1215–1219
17 Brusic, V., Rudy, G., Honeyman, M.C., Hammer, J., and Harrison, L.C. Prediction of MHC class-II binding peptides using an evolutionary algorithm and artificial neural network. *Bioinformatics* 1998, **14**, 121–130
18 Borrás-Cuesta, F., Golvano J.J., García-Granero, M., Sarobe, P., Riezu-Boj J.I., Huarte, E. and Lasarte, J.J. Specific and general HLA-DR binding motifs: comparison of algorithms. *Hum. Immunol.* 2000, **61**, 266–278
19 Brusic, V., and Zeleznikow, J. Computational binding assays of antigenic peptides. *Lett. Pept. Sci.* 1999, **6**, 313–324
20 Miller, P.L. Medical informatics in clinical medicine and the biosciences. *Nat. Med.* 1995, **1**, 93–94
21 Domenech, N., Henderson, R.A., and Finn, O.J. Identification of an HLA-A11-restricted epitope from the tandem repeat domain of the epithelial tumor antigen mucin. *J. Immunol.* 1995, **155**, 4766–4774
22 Elvin, J., Potter, C., Elliott, T., Cerundolo, V., and Townsend, A. A method to quantify binding of unlabeled peptides to class I MHC molecules and detect their allele specificity. *J. Immunol. Methods* 1993, **158**, 161–171
23 Stuber, G, Leder, GH, Storkus, WT, Lotze, MT, Modrow, S, Szekely, L, Wolf, H, Klein, E, Karre, K, and Klein, G. Identification of wild-type and mutant p53 peptides binding to HLA-A2 assessed by a peptide loading-deficient cell line assay and a novel major histocompatibility complex class I peptide binding assay. *Eur. J. Immunol.* 1994, **24**, 765–768
24 Salter, R.D., and Cresswell, P. Impaired assembly and transport of HLA-A and -B antigens in a mutant TxB cell hybrid. *EMBO J.* 1986, **5**, 943–949
25 Brusic, V., Rudy, G., and Harrison, L.C. MHCPEP, a database of MHC-binding peptides: update 1997. *Nucleic Acids Res.* 1998, **26**, 368–371
26 Connelly, M., King, C.L., Bucci, K., Walters, S., Genton, B., Alpers, M.P., Hollingdale, M., and Kazura, J.W. T-cell immunity to peptide epitopes of liver-stage antigen 1 in an area of Papua New Guinea in which malaria is holoendemic.*Infect. Immun.* 1997, **65**, 5082–5087
27 Doolan, D.L., Hoffman, S.L., Southwood, S., Wentworth, P.A., Sidney, J., Chesnut, R.W., Keogh, E., Appella, E., Nutman, T.B., Lal, A.A., Gordon, D.M., Oloo, A., and Sette, A. Degenerate cytotoxic T cell epitopes from *P. falciparum* restricted by multiple HLA-A and HLA-B supertype alleles. *Immunity* 1997, **7**, 97–112
28 Miyata, Y. *A user's guide to PlaNet Version 5.6.* Computer Science Department, University of Colorado, 1991
29 Parker, K.C., Bednarek, M.A., Coligan, J.E. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J. Immunol.* 1994, **152**, 163–175
30 Sidney, J., Grey, H.M., Southwood, S., Celis, E., Wentworth, P.A., del Guercio, M.F., Kubo, R.T., Chesnut, R.W., and Sette, A. Definition of an HLA-A3-like supermotif demonstrates the overlapping peptide-binding repertoires of common HLA molecules. *Hum. Immunol.* 1996, **45**, 79–93
31 van Elsas, A., van der Burg, S.H., van der Minne, C.E., Borghi, M., Mourer, J.S., Melief, C.J., and Schrier, P.I. Peptide-pulsed dendritic cells induce tumoricidal cytotoxic T lymphocytes from healthy donors against stably HLA-A*0201-binding peptides from the Melan-A/MART-1 self antigen. *Eur. J. Immunol.* 1996, **26**, 1683–1689
32 13th IHWC newsletter, issue 3, October 1999. International Histocompatibility Working Group. (http://www.ihwg.org)
33 Mata, M., Travers, P.J., Liu, Q., Frankel, F.R., and Paterson, Y. The MHC class I-restricted immune response to HIV-gag in BALB/c mice selects a single epitope that does not have a predictable MHC-binding motif and binds to Kd through interactions between a glutamine at P3 and pocket D. *J. Immunol.* 1998, **161**, 2985–2893
34 Apostolopoulos, V., Haurum, J.S., and McKenzie, I.F. MUC1 peptide epitopes associated with five different H-2 class I molecules. *Eur. J. Immunol.* 1997, **27**, 2579–2587
35 Prilliman, K.R., Lindsey, M., Jackson, K.W., Cole, J., Bonner, R., and Hildebrand, W.H. Complexity among constituents of the HLA-B*1501 peptide motif. *Immunogenetics* 1998, **48**, 89–97