

# A Novel Predictive Technique for the MHC Class II Peptide–Binding Interaction

MATTHEW N DAVIES, CLARE E SANSOM, CLAUDE BEAZLEY,<sup>1</sup> AND DAVID S MOSS

Antigenic peptide is presented to a T-cell receptor through the formation of a stable complex with a Major Histocompatibility Complex (MHC) molecule. Various predictive algorithms have been developed to estimate a peptide's capacity to form a stable complex with a given MHC Class II allele, a technique integral to the strategy of vaccine design. These have previously incorporated such computational techniques as quantitative matrices and neural networks. We have developed a novel predictive technique that uses molecular modeling of predetermined crystal structures to estimate the stability of an MHC Class II peptide complex. This is the 1st structure-based technique, as previous methods have been based on binding data. ROC curves are used to quantify the accuracy of the molecular modeling technique. The novel predictive technique is found to be comparable with the best predictive software currently available.

## INTRODUCTION

Major Histocompatibility Complex (MHC) Class II molecules are specialized glycoproteins found on the surface of antigen-presenting cells (APCs). Their role is the presentation of antigenic peptides from foreign sources to the receptors of CD4<sup>+</sup> T cells. This function forms part of the body's adaptive immune response to an invasion by a pathogen. The CD4<sup>+</sup> T cells recognize a complex of the MHC Class II molecule and a foreign peptide fragment presented on the APC cell surface. The CD4<sup>+</sup> T cells fall into 2 categories. Inflammatory T cells respond by activating macrophages to destroy the antigenic material, whereas helper T cells stimulate B cells to generate antibodies specific to the antigen. In the Class II pathway, peptides cleaved from extracellular proteins either enter APCs via macrophage vesicles or are internalized by phagocytosis (1). The peptides are colocalized with MHC Class II molecules in intracellular membrane-bound vesicles called MIICs (MHC Class II compartments). Once inside the MIIC, the antigenic peptides will bind with MHC Class II molecules to form 1:1 complexes. These are then transported to the cell surface through direct fusion of the MIIC with the plasma membrane. Once on the cell surface, an MHC-peptide complex must interact with a T-cell receptor to stimulate a reaction. A significant number of interactions between T-cell receptors and MHC Class II peptide complexes will trigger a cascade of intracellular signals that depend on the identity of both the T cell and the APC.

A peptide's ability to stimulate a T-cell response is dependent partly upon its ability to form a stable complex with an MHC Class II molecule (2). Most individuals will express between 6 to 8 alleles of MHC molecules, all of which are capable of binding peptides. Polymorphism and polygeny both contribute to the

variety of molecular structures so that a large range of peptides can be bound and displayed on the cell surface. A single MHC allele may bind hundreds or thousands of different peptides, and binding such a broad spectrum requires a compromise between high affinity and broad specificity. It is known that the presence or absence of a given residue at a given position in the binding groove can be desirable to facilitate binding. However, it is entirely possible that the characteristics of the whole sequence are far more crucial to binding than the identity of the individual residues. Various predictive techniques have been developed to estimate a peptide's capacity to form a stable complex with a given MHC allele. Polymorphic variations cause different allelic variants of the MHC Class II molecule to have varying affinities for a given peptide. That a limited repertoire of MHC molecules can bind such a large variety of antigenic peptides implies that the recognition motif must be relatively unspecific; hence it is difficult to predict whether a given peptide will or will not bind a given allele.

The MHC Class II peptide-binding groove is composed of 2  $\alpha$ -helical regions that form a long cleft binding a single peptide unit overlaid with an antiparallel  $\beta$ -pleated sheet (3). The peptide is bound into the groove by a series of hydrogen bonds (4) and also by the protrusion of the side chains into small cavities along the peptide-binding site. These cavities are known as pockets, and they define a core region of 9 amino acids known to be essential for MHC binding. Pocket 1 (corresponding to position 1 in the core region) is a large hydrophobic cavity near the peptide N terminus that binds hydrophobic side chains, particularly aromatics such as tyrosine and phenylalanine (5). It appears to be the most crucial determinant of the binding interaction whereas pockets 4, 6, 7, and 9 are more permissive in their binding (6).

School of Crystallography, Birkbeck College, University of London, London, U.K. <sup>1</sup>Now at the Rosalind Franklin Center for Genomic Research, Hinxton Genome Campus, Cambridge, U.K.

Sequence-based computational approaches such as quantitative matrices (7–8) and neural networks (9) have had limited success in developing predictive algorithms. Sequence-based prediction systems weigh peptide residues by their independent positions within the binding groove. We present here a novel predictive method that uses molecular modeling of predetermined crystal structures to estimate the stability of an MHC Class II peptide complex by measuring the interaction energy of the 2 molecules. This allows for the consideration of the whole peptide structure rather than its component residues and is therefore a structural rather than a sequence-based technique.

## MATERIAL AND METHODS

Simulated annealing is a molecular dynamics technique that can be used to determine the global energy minimum of a predetermined crystal structure (10). We have developed this technique as a novel method of modeling the MHC Class II peptide-binding interaction. The system is heated up very rapidly and then very gradually kinetic energy is removed until there is none remaining in the system (11). At sufficiently high temperatures all torsions show high rotational frequencies, and thus the structure is able to overcome large energy barriers and move freely between minima in phase space. If the process is continued, the structure is gradually forced to move to lower energy conformations until it becomes trapped near the global minimum. A structure of the MHC Class II allele DR1-0101 bound to an endogenous peptide (PDB code:1AQD) (12) has been elucidated by X-ray diffraction at a resolution of 2.45Å. The structure was remodeled using the crystallographic modeling program 'O' (13) to generate the MHC alleles of DR1-0401, DR1-0301, and DR1-1101. These 4 alleles are all common in the Caucasian population. The polymorphic side chains of the MHC were mutated and modeled into rotamer conformations common for each residue. Energy minimization was then carried out on the structures using the AMBER Version 6 suite of programs (14). The AMBER force field 94 (15) was used to define the atomic interactions. Hydrogen atoms were added to the structure, and the system was fully solvated using water molecules in the TIP3 model (16). This function was performed by the *LEaP* program (14). All atoms in the simulation were explicitly represented. The energy of the solvated molecular complex was minimized using a steepest descent method that continued for 5000 one-femtosecond time steps or until the root mean square deviation between successive time steps had fallen below 0.01Å. The peptides were then remodeled and annealed by raising the temperature of the system from 0 to 500 K for a period of 4000 time steps and maintaining the system at that temperature for a further 3000 time steps. The system was then cooled to 0.2 K over a period of 23000 time steps before being rested at 1 K for a further 30000 time steps. Both the minimization and the annealing were performed using the *sander* program (14). The simulation conditions were optimized using an experimental binding data set (17) from annealing temperatures ranging from 300 to 1000 K. Increasing the acidity of the binding environment was found to be generally deleterious to the accuracy of the simulation. The approximate CPU of each individual simulation was 8 to 9 h on a 6-processor R12000 SGI Origin 2000.

**Table 1. Contingency table to predict binding/nonbinding of peptides**

	Known Binders	Known Nonbinders
Predicted Binders	True Positive	False Positive
Predicted Nonbinders	False Negative	True Negative

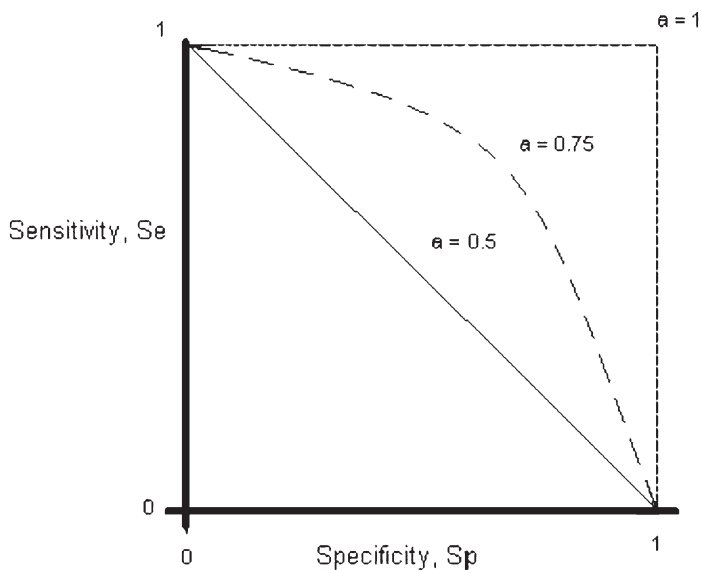
Following annealing, the energetic interaction between the MHC molecule and the bound peptide was analyzed using the *anal* program (14). This calculates the enthalpic interaction energy between the 2 molecules. A low-binding energy indicates a stable complex, hence there is a good chance that the peptide could be displayed in sufficient quantities on the surface of the APC to stimulate an immune response. We have used this energetic calculation as the basis of a system capable of quantifying the binding stability of the peptide and thus acting as a predictive technique. The accuracy of the simulation was optimized by varying conditions such as annealing temperature, cooling time, and pH to produce energetic outputs that correspond most strongly with the experimental data available.

To evaluate the viability of simulated annealing as a predictive mechanism for the MHC Class II binding interaction, it is necessary to provide some means of comparisons to the experimental data set. Two sets of IC<sub>50</sub> binding data were selected to provide an effective assessment of the predictive technique using Relative Operating Characteristic (ROC) curves (18). The 4 possible predictive outcomes are described in a contingency table (Table 1). The 1st data set was generated from the subfragments of a bee venom protein (19) and the 2nd was generated from a data set of peptides taken from the malarial parasite, *Plasmodium falciparum*, and the fungi, *Candida parapsilosis* and *Candida albicans* (20). Both data sets contained inhibitory concentration (IC<sub>50</sub>) binding data for the MHC Class II alleles DR1-0101, DR1-0401, DR1-1101, and DR1-0301, the only alleles for which there are matrices available in both the quantitative matrix-based programs SYFPEITHI (7) and TEPITOPE (8). The overall accuracy of the simulated annealing technique was then measured against both these to see which was the most precise and reliable predictive system.

## RESULTS

### ROC Curves

To assess effectively the quality of the various techniques, it is necessary to be able to compare both their sensitivity and specificity at predicting binding sequences. ROC curves are a diagnostic technique requiring data that may be divided into positive and negative results. To make this distinction, it is necessary to have a defined threshold for the data. A positive and a negative test set for MHC Class II binding has been generated from experimentally determined IC<sub>50</sub> binding data. Those peptides sequences that bind with a concentration < 1000nM were considered to be able to generate a stable complex with a given MHC Class II molecule and are therefore positive binders. Those with a concentration > 1000nM are therefore considered to be negative binders (21). A contingency table can illustrate the comparison between the known binders



**Figure 1.** ROC curves corresponding to  $\alpha$  values, 0.5, 0.75, and 1.0

(those determined through experimentation) against the predicted binders (those calculated using predictive software).

It is the ratios of the diagonal terms for the column total that allow us to determine the quality of the predictive technique. It is necessary to calculate the specificity ( $S_p$ ) and sensitivity ( $S_e$ ) of the data set. These are defined as

$$\text{Specificity } (S_p) = \frac{\text{Number of true positives}}{\text{Number of peptides in positive test set}}$$

and

$$\text{Sensitivity } (S_e) = \frac{\text{Number of true negatives}}{\text{Number of peptides in negative test set.}}$$

The generation of ROC curves allows for a consideration of the technique's capacity to maximize both the true positive proportion (specificity) and true negative proportion (sensitivity). If a high cutoff point is set, then the majority of peptides will be predicted to be nonbinders, thus reducing the number of true and false positives so that the sensitivity increases while specificity decreases. The opposite is true if the cutoff point is low, the number of true and false negatives is reduced so that sensitivity decreases and specificity increases. A ROC curve may be generated by plotting specificity against sensitivity for a range of values calculated by varying the cutoff point.

The accuracy of the prediction technique may be quantified by measuring the area under the ROC curve known as the  $\alpha$  value. The  $\alpha$  value varies from 0.5 (indicating an entirely random prediction) to 1.0 (indicating a perfect prediction system) (Figure 1).

Both data sets featured peptides between 17 and 19 amino acids in length. The termini were unbound, and because the binding data does not indicate which alignment of the peptide within the groove produces the most stable complex, it was necessary to

**Table 2.** Table of  $\alpha$  values for Texier and Southwood data sets

		Texier	Southwood
DR1-0101	SYFPEITHI	0.678	0.575
	TEPITOPE	0.692	0.605
	MD	0.755	0.555
DR1-0401	SYFPEITHI	0.600	0.500
	TEPITOPE	0.625	0.650
	MD	0.680	0.698
DR1-1101	SYFPEITHI	0.750	0.746
	TEPITOPE	0.704	0.667
	MD	0.688	0.635
DR1-0301	SYFPEITHI	0.678	—
	TEPITOPE	0.755	—
	MD	0.692	—

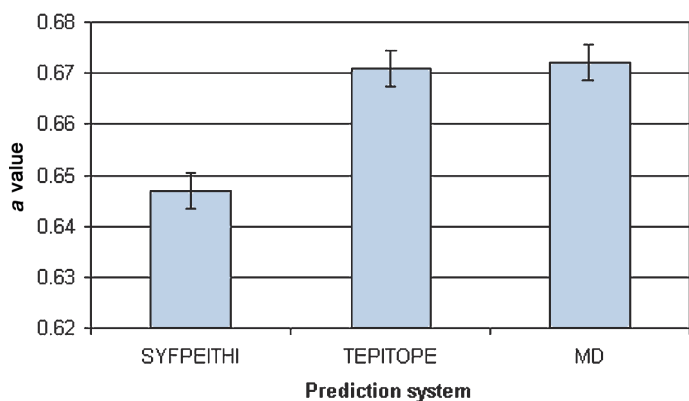
compute a binding score for every possible alignment of every peptide within the data sets. From the set of binding scores generated for each alignment of a peptide, the highest score was nominated as the predicted binding alignment, and the value for that alignment was recorded for the peptide. From these values, a set of scores was obtained for each predictive method with each data set and was used to generate a corresponding ROC curve (Figures 4 to 10).

Catherine Texier's group used a recognized allergen, the bee venom phospholipase A2, which is used in immunotherapy and has been observed to cause a T-cell response involving multiple epitopes (19). Seven alleles expressed prominently within Caucasian populations were selected, and binding assays performed on overlapping subfragments of the bee venom peptide.

It can be observed that the experimental results for DR1-0301 are markedly different from the other 3 alleles and that the predictive success of all 3 techniques is also much lower for this allele. In the case of the quantitative matrices, this may be due to the paucity of binding data available for that allele. The valine at position  $\beta$ 86 represents a significant change in the pocket environment between DR1-0301 and the other 3 alleles, and it is possible the molecular dynamics simulations have failed to properly optimize the parameters for this allele.

Southwood and others performed similar binding assays on peptide fragments taken from the malarial parasite, *Plasmodium falciparum*, and the 2 fungi, *Candida parapsilosis* and *Candida albicans* (20). No significant binding was observed for DR1-0301 and so only the DR1-0101, DRB1-0401, and DRB1-1101 alleles were analyzed.

The experimental data is unlike that generated by Texier in that the majority of the sequences do contain an alignment capable of forming a stable complex with DR1-0101 and DR1-0401. The ROC curves generated are less consistent than those generated by the Texier dataset. The DR1-0101 curve is of questionable value as all but one of the sequences are binders. This unbalances the dataset to the extent that generating a genuine curve was not possible. The absence of a curve for SYFPEITHI in the DR1-0401 dataset indicates that there was no correlation and the results it generated for the allele were effectively random. Conversely SYFPEITHI proves to be the most accurate for the DR1-1101 dataset but demonstrates mas-

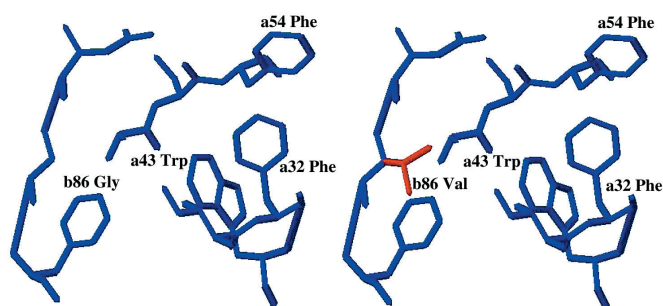


**Figure 2.** Comparative *a* values of MHC Class II predictive techniques

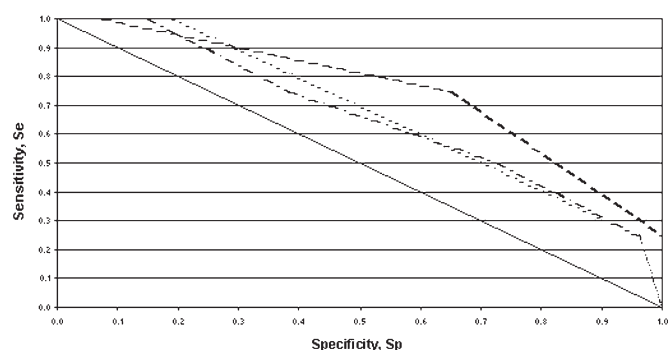
sive inconsistency as a prediction system. The simulated annealing was the most accurate in 3 of 7 cases and the 2nd most accurate in 1 case; it was the least accurate in 3 cases (Table 2). TEPITOPE outperformed SYFPEITHI in 5 of 7 cases. The mean *a* value of each predictive system is shown in Figure 2.

## DISCUSSION

The most vital determinant of peptide-binding success for HLA-DR molecules is the binding of the residue located in pocket 1. The character of pocket 1 is most heavily influenced by the identity of residue 86 on the  $\beta$  chain. This residue is Gly/Val dimorphic, valine in the case of DR1-0301 and glycine in the case of DRB1-0101, DRB1-0401, and DRB1-1101. The presence of a valine in  $\beta$ 86 significantly decreases the size of the binding pocket relative to the presence of a glycine and thus makes it unfavorable for aromatic residues such as tyrosine or phenylalanine (Figure 3). Instead aliphatic residues such as methionine or isoleucine are favored. It can also be seen how the 2 phenylalanines ( $\alpha$ 32 and  $\alpha$ 54) and single tryptophan ( $\alpha$ 43) form a hydrophobic barrier at the mouth of the cavity. The binding pockets 4, 6, 7, and 9 are much more permissive as to the type of side chain they will accommodate and do not make a significant contribution to the binding energy (22). As the binding of the peptide is dependent on the whole length of peptide, the interaction is not susceptible to disruption by a single inappropriate side chain.



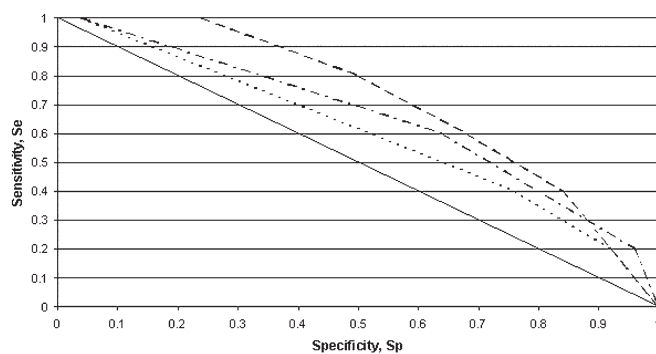
**Figure 3.** The glycine/valine dimorphism of the DR1-0101 pocket, remodeled from PDB structure laqcd.



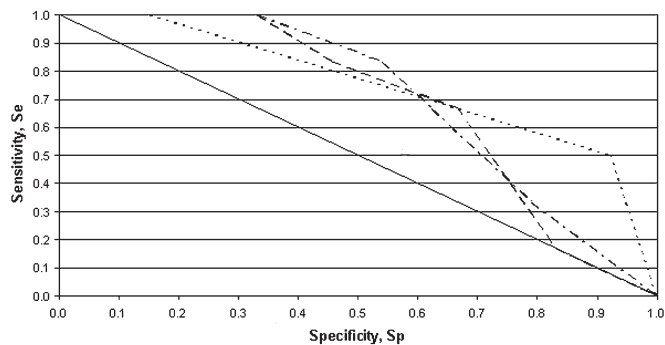
**Figure 4.** ROC values of Texier DR1-0101 predictions. — = random; ..... = SYFPEITHI; - · - · = TEPITOPE; - - - = MD.

One advantage of simulated annealing over quantitative matrices is that it does not require the implicit assumption that the peptide residues bind independently of each other. This is not known to be the case as adjacent side chains in a peptide sequence do interact, and the directions of the  $C^{\alpha}$ - $C^{\beta}$  bond will change depending on the neighboring residue (23). Therefore the energy contribution of a residue at a given position will not be the same in peptides of different sequences and a prediction technique that allows an analysis of the complete binding groove should have an intrinsic advantage over a system that considers the residues individually.

Although no technique proved to have the degree of precision necessary to be regarded as a reliable predictive mechanism, the simulated annealing shows great potential for improvement. The interaction energy calculated from simulated annealing was found to equal the best of the public domain sequence-based quantitative matrices for both accuracy and consistency, giving validity to the technique as a predictive algorithm. The technique will, however, need to be improved if it is to surpass the current software. The raw binding data suggests that the technique has a greater difficulty in eliminating false positives than false negatives. Thus whereas a peptide sequence that is not known not to bind experimentally may be predicted to form a stable complex,



**Figure 5.** ROC values of Texier DR1-0401 predictions. — = random; ..... = SYFPEITHI; - · - · = TEPITOPE; - - - = MD.

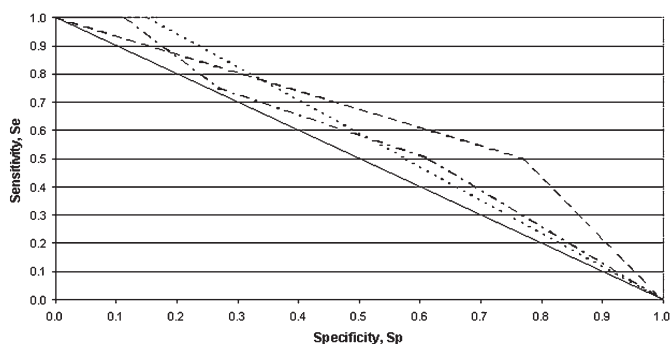


**Figure 6.** ROC values of Texier DR1-1101 predictions. — = random; ..... = SYFPEITHI; -.-.- = TEPITOPE; --- = MD.

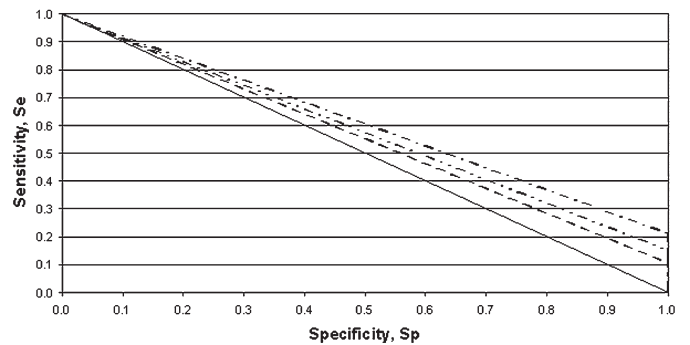
there are no occurrences of an experimental binder failing to form a stable complex in the simulations. This is encouraging as the most important use of the predictive software is to filter a given sequence or genome to identify probable T-cell epitopes for further experimental research. In this situation, accepting a sequence incorrectly is preferable to eliminating one incorrectly.

There are several problems in using simulated annealing as a predictive technique. First, the computational requirements for running a set of simulations are large, and a great deal of time is required to model the structures. This limits the number of structures that can be analyzed in a given period of time and compares very poorly with the near instantaneous results that quantitative matrices can produce. In addition, the results are not directly based upon any experimental data and can only be considered valid if the modeled environment in which the binding occurs is considered accurate enough. A possible source of error is the starting position of the peptide in the simulation. In vivo, the peptide must enter the binding groove of an empty MHC molecule and assume the polyproline type II helical formation. The simulation starts from the position of the peptide having already formed a helix within the groove. Therefore, all steric constraints that would prohibit the peptide both from entering the groove or forming into a helix are not accounted for by the simulation. If possible repulsive forces are not being represented, that would account for the simulation's failure to eliminate false positives.

A possible way to improve the quality of the simulation would be to incorporate the entropic estimates, allowing the cal-



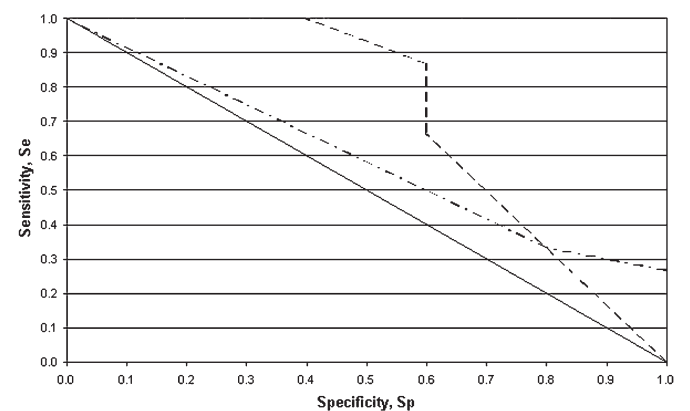
**Figure 7.** ROC values of Texier DR1-0301 predictions. — = random; ..... = SYFPEITHI; -.-.- = TEPITOPE; --- = MD.



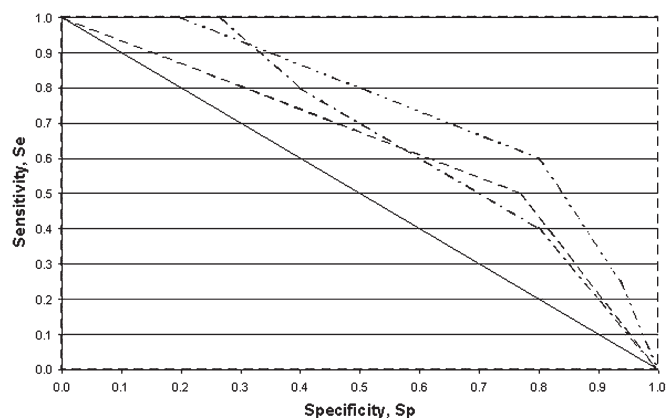
**Figure 8.** ROC values of Southwood DR1-0101 predictions. — = random; ..... = SYFPEITHI; -.-.- = TEPITOPE; --- = MD.

ulation of the free energy of interaction rather than just the enthalpy. It is entirely possible that the unexpectedly high energetic values for nonbinding sequences may be caused by a failure to incorporate the entropic contribution. The entropic energy may be divided into the change to the conformational entropy formation of the complex, the electrostatic contribution to the solvation free energy, and the hydrophobic contribution to the solvation free energy. Of these, the change in the conformational entropy may be considered the most important because it relates directly to the residues of the bound peptide. However, computation of the harmonic potential of the side chains will require the use of normal mode analysis, which would cause a large increase in the computational demands of the simulation. The simulations already require considerable time and computing power to run in comparison to the quantitative matrices. Incorporating the entropic contribution would further compromise the simulated annealing as a practical predictive algorithm. The application of GRID technology (24) to run simulations in parallel would greatly reduce the required run time for a large data set.

In this study the technique of simulated annealing does not prove itself to be a precise method of predicting MHC-binding affinity, although it does compare well with the best currently available public domain software. Whereas the



**Figure 9.** ROC values of Southwood DR1-0401 predictions. — = random; ..... = SYFPEITHI; -.-.- = TEPITOPE; --- = MD.



**Figure 10.** ROC values of Southwood DR1-1101 predictions. — = random; ..... = SYFPEITHI; - - - - = TEPITOPE; - . - . = MD.

quantitative matrix approach seems to have an inherent limitation in that it considers the binding pockets individually rather than the sequence as a whole, simulations have the potential to generate a plausible model of the whole MHC-peptide interaction. Currently, the technique is limited because of its demands on the computational system, but the increasing power and integration of computer systems may considerably reduce the necessary runtime. In the future, we shall develop techniques to include the entropic contribution in the calculation of the MHC-peptide interaction in the hope that they will significantly increase the accuracy of the prediction technique. We also hope to improve the technique by simulating each peptide in both the groove and free solution as a way to calculate its true affinity for the groove.

## ACKNOWLEDGMENTS

MN Davies would like to thank the Biotechnology and Biological Sciences Research Council and the Anthony Nolan Research Institute for their financial support. The authors would also like to extend their thanks to Dr Andy Purkiss for his help in setting up the simulations and Dr Paul Travers for his advice on all things immunological.

**Address correspondence and reprint requests to Clare E Sansom, School of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX. Phone: +44 (0)207-631-6800; fax: +44 (0)207-631-6803; e-mail: c.sansom@mail.cryst.bbk.ac.uk.**

Submitted August 7, 2003; accepted for publication September 29, 2003.

## REFERENCES

1. Germain RN, Hendrix LR. (1991) MHC class II structure, occupancy and surface expression determined by post-endoplasmic reticulum antigen binding. *Nature* 353:134-9.
2. McFarland BJ, Beeson C. (2002) Binding interactions between peptides and proteins of the class II major histocompatibility complex. *Med. Res. Rev.* 22: 168-203.
3. Madden DR. (1995) The three-dimensional structure of peptide-MHC complexes. *Annu. Rev. Immunol.* 13:587-622.
4. Chelvanayagam G, Eastel S. (1997) Peptides: two-faced, cheating go-betweens? Limits in the cellular immune system. *Immunogenetics* 46:516-9.
5. Jardetzky TS, Brown JH, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC. (1996) Crystallographic analysis of endogenous peptides associated with HLA-DR1 suggests a common, polyproline II-like conformation for bound peptides. *Proc. Natl. Acad. Sci. U.S.A.* 93:734-8.
6. Hammer J, Bono E, Gallazzi F, Belunis C, Nagy Z, Sinigaglia F. (1994) Precise prediction of major histocompatibility complex class II-peptide interaction based on peptide side chain scanning. *J. Exp. Med.* 180:2353-8.
7. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. (1999) SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 50:213-9.
8. Sturniolo T, Bono E, Ding J, Radrizzani L, Tuereci O, Sahin U, Braxenthaler M, Gallazzi F, Protti MP, Sinigaglia F, Hammer J. (1999) Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat. Biotechnol.* 17:555-61.
9. Brusic V, Rudy G, Honeyman G, Hammer J, Harrison L. (1998) Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network. *Bioinformatics* 14:121-30.
10. Brünger AT, Adam PD, Rice LM. (1997) New applications of simulated annealing in X-ray crystallography and solution NMR. *Structure* 5:325-34.
11. Hansson T, Oostenbrink C, van Gunsteren WF. (2002) Molecular dynamics simulations. *Curr. Opin. Struct. Biol.* 12:190-6.
12. Murthy VL, Stern LJ. (1997) The class II MHC protein HLA-DR1 in complex with an endogenous peptide: implications for the structural basis of the specificity of peptide binding. *Structure* 5:1385-96.
13. Kleywegt GJ, Jones TA. (1997) Model-building and refinement practice. *Methods Enzymol.* 277:208-30.
14. Case DA et al. (1999) AMBER 6. San Francisco, CA, Univ. of California.
15. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham III TE, DeBolt S, Ferguson DM, Seibel GL, Kollman PA. (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp. Physics Commun.* 91:1-41.
16. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926-35.
17. Marshall KW, Liu AF, Canales J, Perahia B, Jorgensen B, Gantz RD, Aguilar B, Devaux B, Rothbard JB. (1994) Role of the polymorphic residues in HLA-DR molecules in allele-specific binding of peptide ligands. *J. Immunol.* 152:4946-57.
18. Swets J. (1988) Measuring the accuracy of diagnostic systems. *Science* 240: 1285-93.
19. Texier C, Pouvelle S, Busson M, Herve M, Charron D, Menez A, Maillere B. (2000) HLA-DR restricted peptide candidates for bee venom immunotherapy. *J. Immunol.* 164:3177-84.
20. Southwood S, Sidney J, Kondo A, del Guercio MF, Appella E, Hoffman S, Kubo RT, Chesnut RW, Grey HM, Sette A. (1998) Several common HLA-DR types share largely overlapping peptide binding repertoires. *J. Immunol.* 160:3363-73.
21. Sette A, Buus S, Appella E, Smith JA, Chesnut R, Miles C, Colon SM, Grey HM. (1989) Prediction of major histocompatibility complex binding regions of protein antigens by sequence pattern analysis. *Proc. Natl. Acad. Sci. U.S.A.* 86(9):3296-300.
22. Gulukota K, Sidney J, Sette A, DeLisi C. (1997) Two complementary methods for predicting peptides binding major histocompatibility complex molecules. *J. Mol. Biol.* 267:1258-67.
23. Zarutskie JA, Sato AK, Rushe MM, Chan IC, Lomakin A, Benedek GB, Stern LJ. (1999) A conformational change in the human major histocompatibility complex protein HLA-DR1 induced by peptide binding. *Biochemistry* 38:5878-87.
24. Foster I, Kesselman C, Tuecke S. (2001) The anatomy of the grid: enabling scalable virtual organizations. *Intl. J. Supercomputer Appl.* 15(3):1-25.