Immunology and Cell Biology (2002) **80**, 286–298

Special Feature

A combined bioinformatic approach oriented to the analysis and design of peptides with high affinity to MHC class I molecules

CARLOS A DEL CARPIO, TABEA HENNIG, SIMONETTA FICKEL and ATSUSHI YOSHIMORI

Laboratory for Bioinformatics, Department of Ecological Engineering, Toyohashi University of Technology, Tempaku, Toyohashi, Japan

Summary We report on a new method to compute the antigenic degree of peptides from available experimental data on peptide binding affinity to class I MHC molecules. The methodology is a combination of two strategies at different levels of information. The first, at the primary structure level, consists in expressing the peptides binding activity as a profile of amino acid contributions, amino acid similarity being accounted for by their characteristic physicochemical properties and their position within the sequence. The higher level of the strategy is based on a meticulous analysis of the contact interface of the peptides with the cleft constituting the receptor region of a particular class I MHC molecule. Interaction interfaces are inferred by docking the peptide onto the receptor groove of the MHC molecule; evaluation of the affinity of the peptide to the receptor is then performed by analysis of the electrostatic and hydrophobic energies on points of the interaction interface.The result is a robust system for analysis of peptide affinity to class I MHC molecules since while the first analysis dictates the composition of active sequences at the amino acid level, the second translates this information to the atomic level, where the molecular interaction can be analyzed in terms of the intrinsic interatomic forces and energies. Evaluation results for the methodology are encouraging since high affinity peptides are reflected by high scores at both levels of information, and are proportionally lower for peptides of medium and lower affinity for which interaction surfaces show relatively lower electrostatic complementarity and hydrophobic correlation than for the former.

Key words: activity profile, antigenic peptides, binding affinity, drug design, genetic algorithms, major histocompatibility complex, molecular interaction, self organizing maps.

Introduction

Immune responses are regulated and initiated by MHC molecules, which bind short peptides (resulting from intracellular processing of proteins) and display them on the cell surface in order that TCR can recognize them.[1–3] For a peptide to be recognized by a T cell, its binding to any type of MHC molecule is imperative; not all peptides have the same binding activity. Accordingly, understanding the molecular mechanism underlying immunological responses and phenomena requires the determination of the binding affinity (BA) of a peptide to a MHC molecule.

While MHC class II molecules usually bind peptides of fluctuating lengths (between 10 and 30 amino acid residues)[1,4] and for which anchor amino acids are not well defined, MHC class I molecules seem to be better defined, both in terms of length and number of anchor amino acids of the peptides they bind.[1,5,6] Nevertheless, many studies have been carried out in order to determine quantitatively BA for peptides of different lengths to MHC class I molecules, and the contribution of every amino acid constituting the sequence to the overall BA.[5–8] These studies are based on the idea of an MHC class I molecule binding matrix that quantitatively expresses the importance of an amino acid within the sequence of a peptide on its overall binding strength. Fundamentally, these quantitative matrices are a scheme to express binding motifs in a rather refined way. Each entry in the matrix is a coefficient expressing the importance of the amino acid and its position within the sequence. These coefficients can then be used to compute and predict MHC binding strength indices as well as peptide similarity indices.

Several computational methodologies have been exploited experimentally to correlate obtained peptide BA to MHC class I molecules and the amino acid composition of the peptides, in order to derive the optimal table or matrix expressing the binding coefficients. These range from relatively simple simultaneous equations and inequalities solution methods[5] to more refined pattern recognition and learning techniques[1,6] based on neural networks and evolutionary algorithms.

In the present study, we propose a multilevel information-processing methodology to assess peptide affinity for MHC class I molecules. The first level of information processing consists of relating the primary structure of peptides to their experimentally measured activity. Here we focus on 9-mer peptides and the construction of a profile of peptide binding activity to MHC class I molecules. The resulting profile can be used to predict BA of any 9-mer sequence.

Correspondence: Dr Carlos A Del Carpio, Laboratory for Bioinformatics, Department of Ecological Engineering, Toyohashi University of Technology, Tempaku, Toyohashi 441–8580, Japan.
Email: carlos@translell.eco.tut.ac.jp

Received 19 February 2002; accepted 19 February 2002.

The technique underlying the methodology is based on a genetic algorithm (GA),[9] which, combined with the profile analysis for detection of related proteins introduced by Gribskov,[10] leads to a set of coefficients that express quantitatively the contribution of each amino acid in the peptide sequence to its binding activity. The GA evolves the coefficients obtained from the similarity analysis, so that computed binding activities correlate optimally to the respective experimental binding activities, yielding a table of optimal correlation coefficients, or a matrix of binding activity indices for each amino acid.

These scores express the contribution of particular amino acids at particular positions to the overall binding activity of the peptide. The tables enable the identification of anchor amino acids within the oligopeptide, that is, those amino acids that contribute to a higher degree to the activity of the peptide when they are located at determined positions within the amino acid sequence. Calculation of the binding activity for an arbitrary sequence of amino acids is straightforward using the matrix of binding activity scores.

The second level of information processing consists of the analysis of the complex formed by a peptide and the receptor region of the MHC molecule. Prediction of the structure of the ligand peptide is carried out, followed by a docking simulation of the peptide into the receptor cleft of the MHC molecule. The interaction interface is then extracted from the most plausible decoy. The analysis is carried over a set of several types of MHC class I-binding peptides, and tendencies in hydrophobic correlation as well as electrostatic complementarities are ascertained that may drive the binding of the molecules and formation of the complex. This analysis is performed using the system for assessment of bio–macromolecular interaction Macromolecular Interaction Assessment System (MIAX) reported by Del Carpio *et al*.[11–14] The tendencies derived are then applied to assess the probability of ligand–receptor binding at the atomic level.

The sequence of analysis proposed here can be directly applicable in modelling peptide BA to MHC molecules, eliminating unnecessary experimental assays. Data on peptide binding strength and activity can thus facilitate the design of clinically useful and immunologically silent peptidic drugs, as well as immunotherapeutics and vaccines for autoimmune diseases and cancer.

Methods

Peptide interaction can be described as a function of mainly ionic forces, hydrophobic interactions, and hydrogen-bond type interactions, the latter being calculated as part of the well-characterized electrostatic interactions. Binding of peptides to larger molecules, such as MHC molecules, can be evaluated by means of these energies. The stability of the complex is dictated by the stability attained by the structure of both the ligand and the receptor at interaction.

Considering the crystal structure of MHC molecules and reports on their peptide binding grooves, which show that these regions undergo only slight adjustments at interaction with the ligand, maintaining essentially a similar shape in all peptide complexes,[5,15] it can be hypothesized that peptide binding to MHC molecules is primarily driven by the structural characteristics of the ligand peptide. Based on this hypothesis, it was proposed the binding activity of peptides could be related to their amino acid composition.

However, in order to establish the nature of the predominant interaction energies associated with the complex formation, a detailed analysis of the receptor–ligand interface was performed over a set of randomly selected peptides for which experimental activity values have been reported. Insights obtained by this analysis were then applied to assess the feasibility of complex formation of peptides predicted to have high activity by the primary sequence analysis.

Applying this combined strategy leads to a robust system for the analysis of peptide affinity to MHC class I molecules. We applied this method to a series of 9-mer peptides for which the binding characteristics were available.

Activity profiles for MHC class I-binding peptides

Gribskov's profile analysis is a technique to compare amino acid sequences.[10] A group of previously aligned sequences with common 3-D and/or functional characteristics is used as the probe for which the profile is computed. The profile itself *M (p,a)* is an $m \times n$ matrix of scores; m being the number of columns corresponding to the number of amino acids considered in the calculation plus an extra one for deletion/insertion penalties, while n is the length of the amino acids constituting the probe. The values of the profile for a determined amino acid a is computed by the expression:

$$M(p, a) = \sum_{b=1}^{20} W(p, b) \times Y(a, b) \tag{1}$$

where *Y(a,b)* is the similarity of amino acids a and b, extracted from Dayhoff's distance matrix, and *W(p,b)* is a weight representing the amino acid b at position p, which in Gribskov's method is computed by an averaging scheme (the frequency of appearance of each amino acid at a certain position within the amino acid sequences constituting the probe).

Here, we propose a methodology to predict MHC class I molecule BA, which shares similarities with the Gribskov's profile analysis because we also seek a matrix or profile similar to *M (p,a)*, but with the fundamental difference that similarity coefficients are obtained through an artificial evolution process that evolves in order to express optimally the contribution of every amino acid and its position, within the primary sequence of amino acid residues, to the overall BA of the peptide.

The profile obtained here, the MHC binding matrix (BM), is a 20×9 matrix. This represents all naturally occurring amino acids (20) and the probe, which in this case is made up of peptides of fixed length (9-mers). An extension to larger peptides is straightforward. Each element *BM(a,p)* represents the contribution of amino acid a at position p to the overall BA of the peptide. The profile *BM(a,p)* is calculated by the following expressions:

$$BM(a, b) = f(p, b) \cdot S(a, b), \ \text{if } a = b \tag{2}$$

$$BM\,(a,p) = \sum_{a \neq b} \{1 - v(p)\}^2 f(b, p) \cdot S(a,b), \ \text{if } a \neq b \tag{3}$$

where *BA (a,p)* is the BA contribution score for amino acid a at position p; *f (b,p)* is the frequency of appearance of amino acid b in position p; and *v (p)* is a coefficient standing for the deviation of the distribution of the frequency of appearance of the amino acids at position p. This coefficient is calculated for the sequences constituting the set of sample peptides (object set) aligned at position p with the expression:

$$v(p) = \frac{n \sum_{j=1}^{20} f(j, p) - \left(\sum_{j=1}^{20} f(j, p) \right)^2}{n(n-1)} \tag{4}$$

where n is the number of 9-mer peptides in the object set.

Finally, *S (a,b)* is the similarity index for amino acids *a* and *b*, which is computed by artificial evolution using a GA described in what follows.

The objective of the methodology is to find an optimal matrix (BM) whose elements are the values calculated by equations 2 and 3, which can be used to compute directly the BA for any peptide of length *n* and sequence a_1, a_2, $a_3 \ldots a_n$ by an additive scheme of the form:

$$BA = BM(a_1, 1) + \ldots + BM(a_i, i) + \ldots + BM(a_n, n) \quad (5)$$

where, *BM* (a_i,*i*) is the BA score for any amino acid a_i (out of the 20 naturally occurring amino acids) at position *i*. To compute the BM elements expressing the contribution of every amino acid at its respective position in the sequence to the overall affinity of the peptide using equations 2 and 3, a matrix *S (a,b)* must be obtained that reflects the similarity between any pair of amino acids *a* and *b*.

To achieve this goal, we propose a genetic algorithm that evolves strings of numerical scores (Fig. 1), with each value or score representing an amino acid that is used to compute its similarity to any other of the 19 amino acids, *S (a,b)*, using a scheme introduced by Grantham.[16]

This scheme, adapted here to the computation of pair-wise amino acid residue similarities, consists of the computation of a mutation matrix, D_{ij}.[17] Figure 1 shows an arbitrary string of g_i scores evolved by the GA and used in that calculation of the similarity matrix *S (a,b)* for a pair of amino acids *i* and *j* using the expression:

$$D_{ij} = [\alpha(g_i - g_j)] \quad (6)$$

where g_i and g_j are the indices representing the amino acids in the gene string (chromosome or individual; Fig. 1) in the GA proposed here. Here α is then computed as:

$$\alpha = \left(\frac{1}{\overline{D}}\right)^2 \quad (7)$$

and stands for a scaling factor calculated from the mean value *D* of the off-diagonal elements of the symmetrical matrix *D* (*i*,*j*). Finally, considering a scale in which 0 denotes total amino acid dissimilarity and 1 complete similarity, the similarity matrix is computed using equation 8.

$$S_{ij} = 1 - D_{ij} \quad (8)$$

Initially, a population of chromosomes with random values g_i for each score is generated. Subsequently, with the operations of the GA (selection, mutation and crossover), these values are improved to a degree in which pair-wise amino acid similarities (S_{ij}), calculated using those values as dictated by equations 6, 7 and 8 can be used to express optimally peptide BA to MHC class I molecules.

One of the most important steps in solving multivariational methods using GA is the evaluation of the eligibility and fitness of a determined individual (or chromosome) through the evolution process, by calculation of the penalty or object function in the GA. Since our purpose is to express the BA of any peptide to the MHC molecule as dictated by equation 5, the *BM* (a_i,*i*) scores must express optimally the contribution of amino acid a_i at position *i* to the BA of the peptide. Because the BM scores are directly related to amino acid pair wise similarities, as expressed by equations 2 and 3, optimal amino acid pair wise similarity values are necessary. The optimal BM scores sought are those that, when used in equation 5, reproduced with the highest accuracy the BA values for the peptides in the object set of samples. Accordingly, the fitness function used in the present work consists of the calculation of the BA values for all the peptides in the object set, following the procedure described before, and correlating them to the experimental values reported in the literature. In this way, the GA evolves in the direction of improving the correlation coefficient of calculated versus experimental values. In this case, the binding scores computed using the elements of the BM, following the scheme of equation 5, are in inverse relation to the experimental coefficients. Therefore, the units of the computed binding scores are arbitrary (peptides showing low computed scores being less active than those having high scores), while the experimental data used are in nM IC_{50} units (low values stand for highly active peptides and high values for less active peptides).

The GA improves the values for each index that represents each amino acid in the chromosome. This improvement leads, in turn, to the optimal similarity matrix for the amino acid pairs from the point of view of MHC class I molecule BA, and consequently to the optimal BM from which binding affinity values can be computed for any amino acid sequence. The complete process to obtain the optimal BM is illustrated in the form of a flow diagram in figure 2.

Analysis of the interaction and computation of electrostatic and hydrophobic binding scores

To further assess the feasibility of interaction and complex formation of MHC class I molecules with any arbitrary peptide at the atomic interaction level, we propose a methodology based on analysis of the forces driving bonding among the molecules on the ligand–receptor interaction interface, which is formed at complex formation. The analysis is then applied to the peptides predicted by the methodology, leading to a more complete picture of the interaction of the peptide with the receptor groove. This analysis is performed evaluating the complementarity and/or correlation of energies and forces leading to bonding on the interaction interface of the complex MHC receptor molecule and the peptide molecule.

To extract information corresponding to the interface of the ligand–receptor interaction, the tertiary structure of the peptide is first retrieved from a database or predicted using 3-D coordinate prediction programs (in the present study we performed the predictions using essentially the molecular mechanics-based programs GAX[18] and AMBER,[19] and the molecular dynamics system TINKER.[20–22] By a docking process performed using the system for bio–macromolecular interaction assessment (MIAX) developed by Del Carpio *et al.*,[11–14] candidate decoys for the complex peptide–MHC molecule are generated, the lowest energy one being selected for the evaluation described here.

The interface is determined by computing the set of points on the surface of atoms of the receptor groove occluded by the atoms of the peptide in the complex. The procedure consists of the calculation of the set of solvent accessible surface points (SASP) on each atom of the receptor and the ligand, and the selection of those atoms of the receptor occluded at interaction. To perform this calculation, each atom is represented by a sphere of radius equal to the sum of its van

A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
7.95	0.11	7.65	7.89	0.42	1.58	0.6	0.22	5.57	6.03	5.54	7.63	1.65	5.55	8.37	8.37	3.05	5.84	5.53	5.51

Figure 1 Arbitrary string of g_i scores evolved by the genetic algorithm (GA) and used in calculation of the similarity matrix *S* (*a*,*b*).

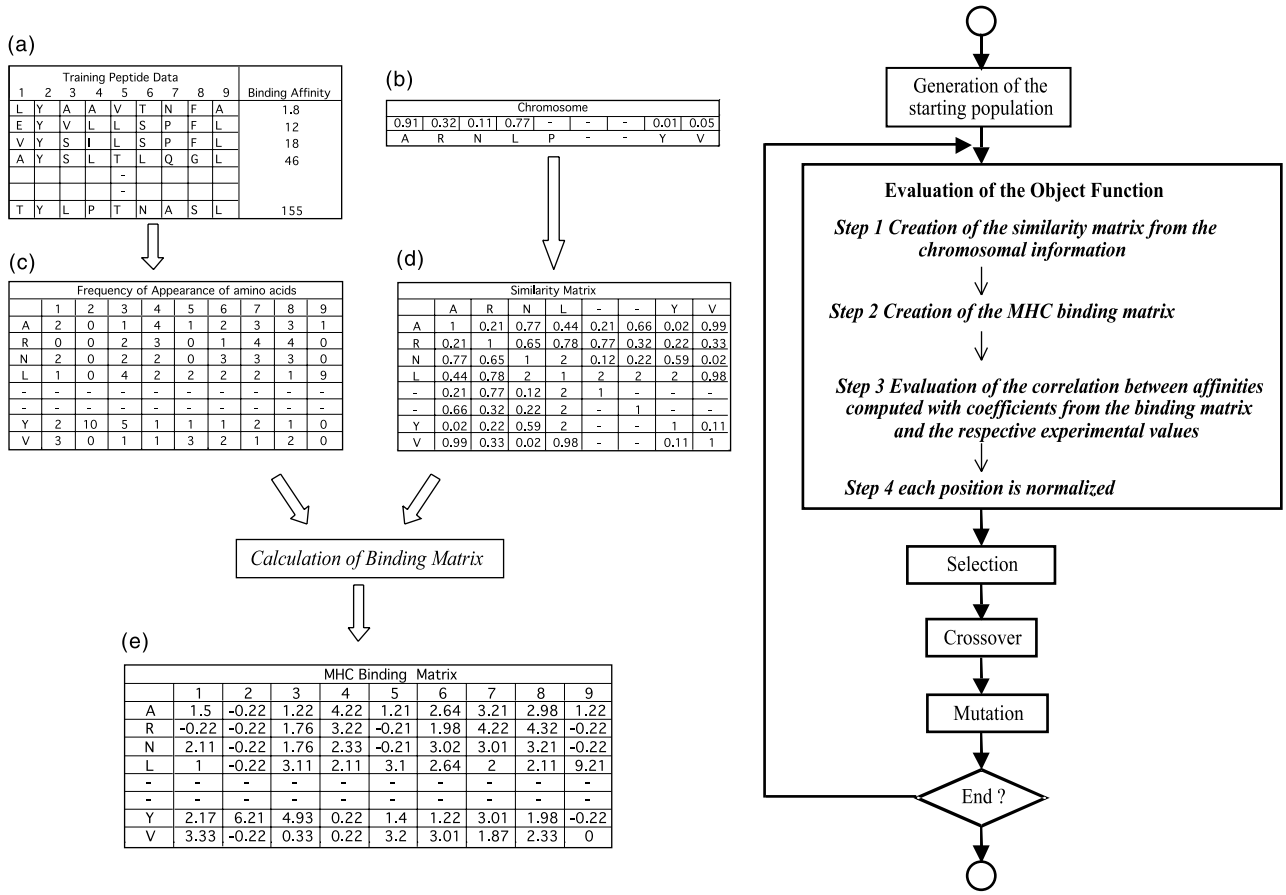

Figure 2 Left: Scheme to compute the BM: (a) Peptide sequence and affinity data. (b) Chromosome of g_i values. (c) Matrix of frequency of amino acid appearance at position *i*. (d) Similarity matrix. (e) Binding matrix (BM).

Right: Flow chart for the genetic algorithm.

der Waals radius plus the solvent diameter (1.4 Å for water). The surface is represented by 92 surface points computed as described by Del Carpio *et al*.[23]

The analysis of the inferred interface begins with the calculation of the physicochemical properties expressing bonding and interaction. Since the main interaction factors among proteins and peptides are electrostatic interactions (involving ionic and hydrogen bonding) and hydrophobic interactions, potentials related to these properties are calculated on every point of the interface. Electrostatic potentials are calculated by equations 8a and 8b, while hydrophobic potentials are computed using equation 9.

$$\varepsilon_0 = 78.0 + 38.0\{(0.3 \times r)^2 + 0.6 \times r + 2.0\}e^{-0.3r} \quad (8a)$$

$$ep = \sum \frac{332.0}{\varepsilon_0} \frac{chrg}{r} [\text{kcal/mol}] \quad (8b)$$

$$MHP = \sum E_{tri} e^{(r_i - d_i)} \quad (9)$$

where *ep* is the electrostatic potential and *MHP* is the molecular hydrophobicity potential.

Considering a cut distance of 12 Å, potentials created by atoms of the receptor and those created by the ligand atoms are computed on each point of the interaction interface.

The hypothesis underlying the present analysis is that the interaction forces leading to the bonding of the molecules are of opposite sign if electrostatic, and are positive values for both molecules when hydrophobic interactions are considered.

To analyse these complementarity relationships in a simple way, the interaction interface (an array of 3-D points) is projected into a 2-D array of neurones by means of a self-organizing map or Kohonen neural network.[24] This procedure is described in detail by Del Carpio *et al*.[14] Since the calculated map is an array of points that preserve the 3-D distribution of the surface, the values for electrostatic and hydrophobic potential are plotted on the maps, resulting in corresponding electrostatic and hydrophobic maps for receptor and ligand. These 2-D maps allow a simple scoring scheme to compute complementarity of interaction forces on the interaction interface. This consists of comparing each corresponding neurone on the maps for the ligand and receptor, and the summing of all neurones with complementary relationships (in the case of the electrostatic maps) and of the same sign (in the case of hydrophobic maps). This results in two scores that describe the balance of forces on the interface of the molecules forming the complex. These coefficients can then be related to the BA of the peptide.

The process of calculation of the score of interaction is exemplified for peptide CLTSTVQLV and receptor HLA-A2.1 in Fig. 3. The 3-D structure of the peptide is calculated by the polypeptide 3-D structure prediction system GAX[18] with further refinement using the systems AMBER[19] and TINKER.[20–22] The docking process is

performed using the system MIAX.[11–14] The complex configuration selected among the decoys output by MIAX is that possessing the lowest energy. Figure 4 illustrates the maps of electrostatic potentials for the ligand and receptor molecules, respectively.

The calculation of the binding scores for electrostatic potential is also illustrated in Fig. 4. The left map shows the distribution of the electrostatic potential from low negative values (red) to high positive values (blue), corresponding to the field created by atoms of the receptor. The right map shows the distribution of the potentials created by the atoms of the ligand when an ideal probe of charge (+1e) is placed on every SASP. The electrostatic binding score is defined as the total sum of corresponding neurones having opposite signs. Similarly, after the calculated hydrophobic potential values are assigned to the neurones of the self organizing map (SOM) for both molecules, the scores for hydrophobic bonding are computed as the total sum of all corresponding neurones in both maps for which the values are positive.

To establish if any correlation exists between experimental values and the computed interaction scores suggested in the present study, a set of peptides for which the affinities were known *a priori* were treated with our methodology. The set of peptides for which this correlation was studied is presented in Table 1, and binding scores versus BA are plotted in Fig. 5 for the electrostatic interaction and in Fig. 6 for hydrophobic interaction. In both figures, normalized values are plotted against the experimental binding affinities.

A trend can be observed in the relationship between the binding scores and the affinities of the peptides. Low (negative) values showing electrostatic complementary on the interface are associated with high values of affinity, while lower complementarity is observed for low affinity values. A similar trend can be observed for the hydrophobic potentials on the interaction interface.

On the basis of this correlation, we propose a systematic way for assessing the complex formation of a MHC class I molecule with any peptide predicted by the methodology based on amino acid binding profiles. Determination of the structure of the complex peptide–MHC molecule enables the inference of the interface of interaction on which the potentials created by the atoms of the receptor and the ligand can be evaluated for complementarity. High complementarity may mean a stable complex (i.e. high probability of complex formation and affinity). Low complementarity may reflect a low probability of complex formation or interaction among both molecules. The former result would enable the identification of highly active peptides, while the latter would confirm the presence of a silent antigenic peptide with low affinity for the MHC molecule.

Evaluation of the peptides from these two points of view leads to a robust scheme of evaluation of molecules (such as the functional peptides studied here). The simplicity of treating the affinity in function of the sequence of amino acid residues enables the fast screening of peptides with affinities for certain MHC molecules. At this level of evaluation, qualitative characteristics of the sequence, such as primary and secondary anchors, can be inspected rapidly. However, the second part of the analysis allows for a more detailed evaluation of the interaction pattern that characterizes not only the anchor amino acids of the peptide, but the 3-D structure of the peptide, which is not evident at the primary structure information level. These analyses lead to a characterization of the interaction and formation of the complex structure. The interaction and reactive specificity cannot be completely characterized as a pairwise amino acid property and characterization in terms of atom-pair interaction is also difficult because no specific covalent bond is created during the interaction. Rather, this is a soft bonding, driven by characteristics of the specific environment created by the local distribution of atoms resulting from a particular order in primary sequence and the tertiary structure adopted by the peptide.

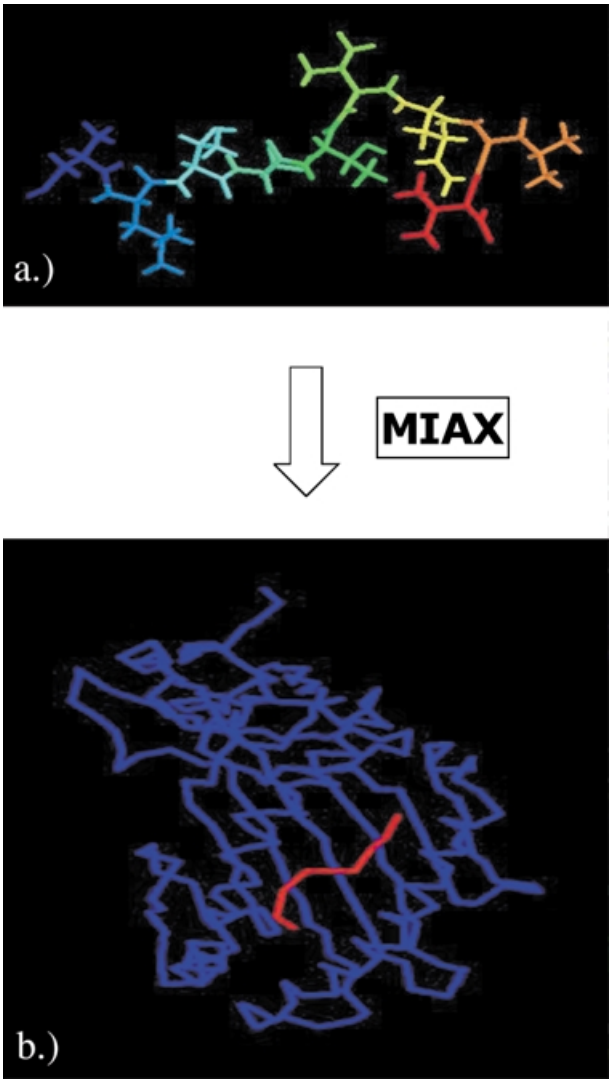

Figure 3 Docking process (a) Peptide tertiary structure prediction. (b) Docking or peptide to the MHC I molecule using MIAX.

Results

First, we apply the proposed methodology to sets of peptides in order to characterize their function in terms of the sequence of the amino acids; and second, we attempt to establish the characteristics of the interaction leading to the complex peptide–MHC class I molecule. Finally, we evaluate the methodology using the MHC class I molecules HLA-A24 and HLA-A2.1, for which reported data on binding affinities are available.

Binding affinity profile for HLA-A24 binding peptides

Table 2 summarizes the object set or data set employed to derive the amino-acid-based binding profile for peptides binding to HLA-A24 as well as for model validation and prediction assessment. The primary sequences of all the peptides in the training set are shown, together with their respective experimental BA values to HLA-A24 and references to the source (Table 2).

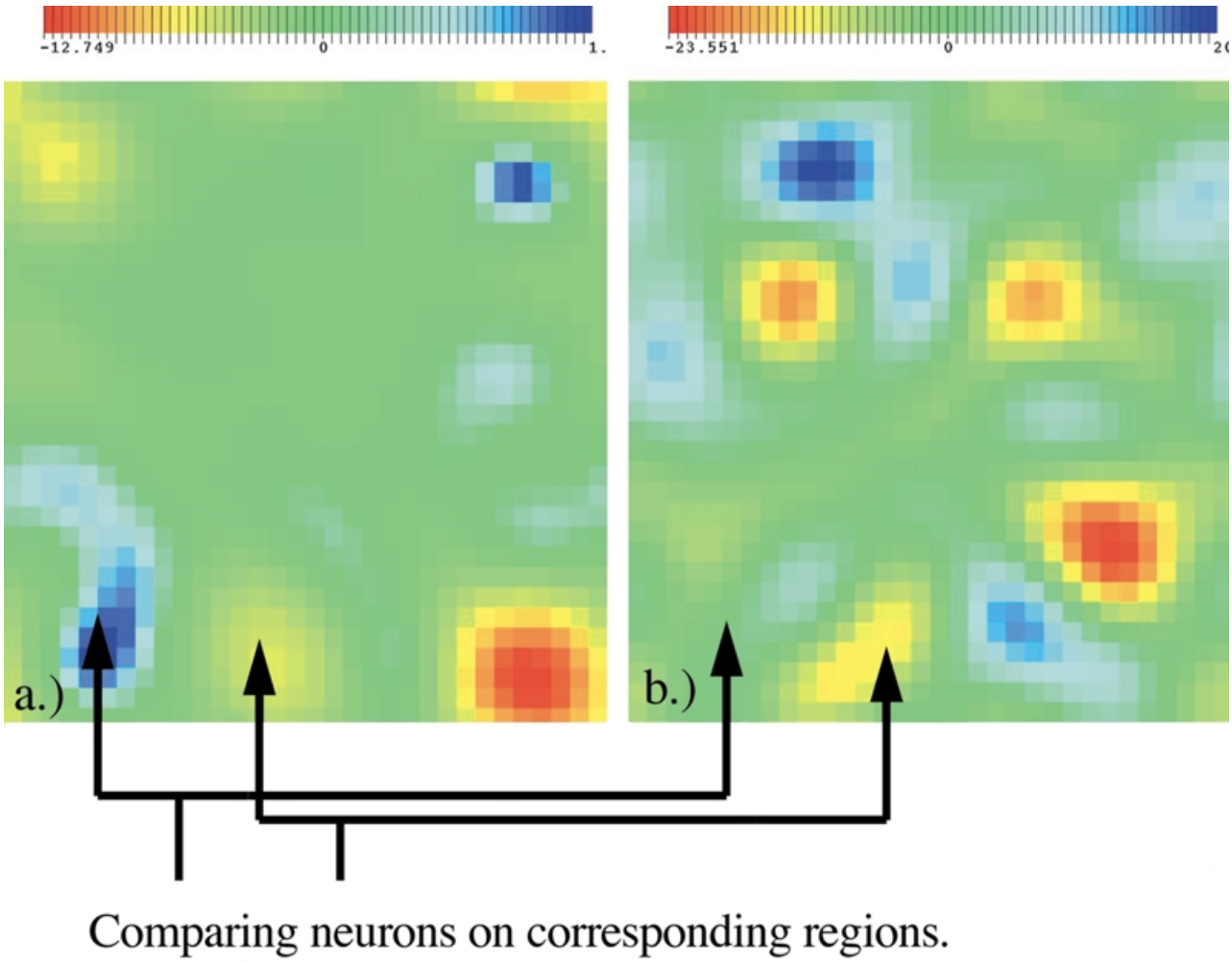

Figure 4 Calculation of electrostatic complementarity scores.[25]

Since the methodology developed here (intended to predict the BA of peptides to MHC class I molecules) can also be applied to the characterization of MHC class I molecule binding motifs, we describe the assessment of its potential applicability in two parts corresponding to these two purposes.

As illustrated in Table 2, peptides binding to MHC class I molecules can be classified as high-, moderate- and low-BA peptides. A further category comprises those peptides showing BA equal or close to zero (large dissociation constants), but for which well-defined experimental data are seldom reported. Accordingly, the analysis performed here employing only those peptides with high BA can be directed to predict motifs with high BA for MHC class I molecules. The whole data set in Table 1 is intended to predict the binding ability of the peptides.

After computing the matrix of frequencies (Fig. 7) using step (a) in Fig. 2a, each element of the matrix being the frequency of appearance of each amino acid type at each of the nine positions in the object set of 9-mer peptides, a population of 1000 chromosomes (strings of the type shown in Fig. 1) is generated randomly to start the evolutionary learning process. With a crossover rate of 50% and a mutation rate of 3%, the GA was performed for 500 generations (Fig. 7b). The evolutionary learning process is depicted in Fig. 7c.

This process results in the similarity matrix shown in Table 3, each element of which is computed according to equations 6–8. Using the values of the frequency matrix and Table 3 (the optimal similarity matrix), the BM, shown in Table 4, is computed according to equations 2 and 3. The final correlation of computed affinity values (using scores of the optimal BM in Tables 4 and 5) to the corresponding experimental ones is illustrated in Fig. 8.

Scores for amino acids in the optimal BM represent the contribution of every amino acid to the overall BA of the peptide. Therefore, they can be ranked in decreasing order for each position and information can be obtained on the significance of each amino acid at a particular position. This operation leads to the elucidation of primary structural characteristics for the peptides, related to their binding affinity. Information on primary and secondary anchors can be easily extracted and compared with other peptides that have BA to the same or other MHC molecules. Defining primary anchors as those amino acids having a score larger than 1.0 in the BM, and secondary scores as those having a value fluctuating

Table 1 Object set of HLA-A24 binding peptides

Sequence	Experimental binding activity (nM IC50)	Electrostatic attraction normalized	Hydrophobic attraction normalized	Length
KMVELVHFLL	5.0	0.47	0.57	11
KASEYLQLV	151.5	0.59	0.50	9
KVAELVHFL	54.3	0.30	0.42	10
KVAELVHF	68.4	0.55	0.52	9
YLWWVNNQSL	26.2	0.30	0.35	10
VLYGPDAPTI	454.5	0.16	0.34	10
VVLGVVFGI	14.3	0.54	0.48	9
VMAGVGSPYV	277.8	0.23	0.45	10
YMIMVKCWMI	83.3	0.58	0.61	10
YLQLVFGIEV	50.0	0.76	0.45	10
QLVFGIELMEV	7.9	0.23	0.47	11
VLGPDTPI	200.0	0.85	0.62	9
VLIQRNPQL	22.7	0.43	0.55	9
LVFGIEVVEV	357.1	0.27	0.81	10
LVFGIELMEV	76.9	1.00	1.00	10
YLWWVNGQSL	33.3	0.40	0.54	10
CLTSTVQLV	147.1	0.30	0.41	9
KMELVHFLL	22.7	0.40	0.65	10
YIFATCLGL	185.2	0.68	0.48	9
IMIGVLVGV	68.5	0.16	0.53	9
SIISAVVGI	69.4	0.24	0.28	9
KMVELVHFL	9.8	0.39	0.85	9
FLWGPRALV	31.3	0.43	0.76	9
GIMIGVLVGV	56.8	0.19	0.61	10
GLACHQLCA	416.7	0.36	0.31	9
KIWEELSML	166.7	0.19	0.82	9
HLFIYATCLGL	55.6	0.52	0.32	11
YLSGANLNL	27.8	0.25	0.18	9
QLFEDNYAL	17.2	0.41	0.22	9
FLWGPRAL	238.1	0.32	0.45	9
LLTFWNPPT	178.6	0.84	0.47	9
ALCRWGLLL	100.0	0.31	0.45	9

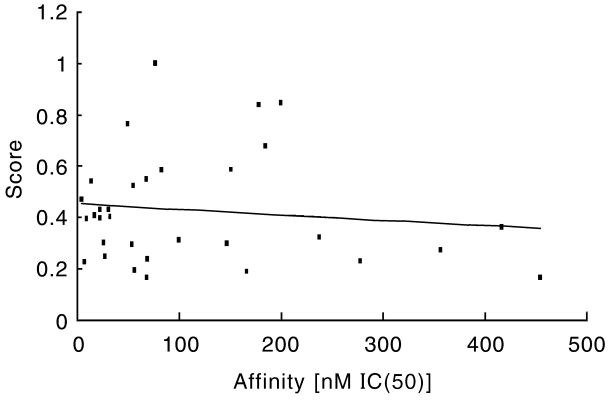

Figure 5 Correlations of electrostatic interaction scores with binding affinities.

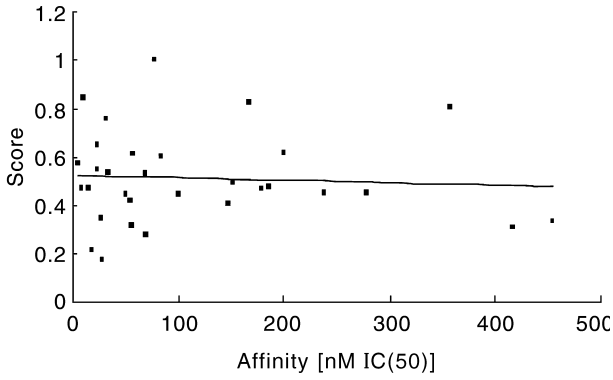

Figure 6 Correlations of hydrophobic bonding scores with binding affinities.

between 0.5 and 1.0, we find that at P2, the amino acids Phe, Tyr and Trp can be catalogued as primary anchors, as can amino acids Ile, Leu and Phe at position P9 (Fig. 9). Although Trp at P9 has a score lower than Tyr and Val at this position, these results are comparable to those reported by Brusic *et al.*,[6] who used a larger training set and a more complex learning algorithm. These results are similarly comparable to those obtained by Maier *et al.*,[26] except for M at position P2, which has a negative score in the BM of Table 5. However, the results are identical at the C-terminus.

The primary and secondary anchors, taking into account our definition, are summarized in Fig. 9; the binding matrix itself is a summary map of the effects influencing the binding capacity of 9-mer peptides to HLA-A24.

Binding affinity profile for HLA-A2.1 binding peptides

To further assess the potential applicability of the methodology, it was applied to an object set consisting of 9-mer peptides binding to the MHC class I molecule HLA-A2.1 (22

Table 2 Set of HLA-A24 binding 9-mers used as object set

Sequence	Binding capacity (nM IC50)	Sequence	Binding capacity (nM IC50)	Sequence	Binding capacity (nM IC50)
LYAAVTNFL	1.8	LYQTFGRKL	430.0	NFLQAAYRL	431.0
EYVLLLFLL	4.1	VYGDTLEKL	580.0	TWPLLPHVI	46.0
LYSILSPFL	12.0	CYSIEPLDL	699.0	AWQNGLLPF	841.0
VYCKTVLEL	18.0	PYAVCDKCL	744.0	SFHNLHLLF	1.2
AYSLTLQGL	46.0	EYVNARHCL	951.0	EYTNIPISL	60.0
CYSLYGTTL	103.0	LYKTFGRKL	951.0	PYKRIEELL	150.0
TYLPTNASL	155.0	KYADKIYSI	0.57	PWTHKVGNF	177.0
IYQEPFKNL	178.0	SWLSLLVPF	18.0	AYINADSSI	192.0
TYSTYGKFL	258.0	SWWTSNFL	18.0	SFLLSHGLI	812.0
LYNLLIRCL	314.0	NWKPIVQFL	27.0	GWSPQAQGI	2875.0

(a)

Amino acid	Amino acid position within the peptide 1st	2nd	3rd	4th	5th	6th	7th	8th	9th
A	3	0	3	1	3	1	1	0	0
R	0	0	0	1	0	1	3	1	0
N	2	0	1	5	0	1	2	2	0
D	0	0	0	2	0	2	0	1	0
C	2	0	1	0	1	0	0	3	0
Q	0	0	3	1	0	0	2	0	0
E	3	0	0	1	1	1	2	1	0
G	0	0	1	0	1	4	2	1	0
H	0	0	1	1	0	2	2	0	0
I	1	0	1	2	3	2	1	0	4
L	5	0	4	6	6	6	4	4	21
K	1	0	3	1	2	0	3	3	0
M	0	0	0	0	0	0	0	0	0
F	0	3	0	0	2	1	1	5	4
P	3	0	1	2	1	3	1	2	0
S	4	0	5	1	2	1	1	4	0
T	3	0	2	4	4	1	1	1	0
W	0	6	1	0	0	0	0	0	0
Y	0	20	0	0	2	0	2	0	0
V	2	0	2	1	1	3	1	1	0

(b)

GA	Conditions
Population	1000
Generation	500
Crossover	50%
Mutation	3%

(c)

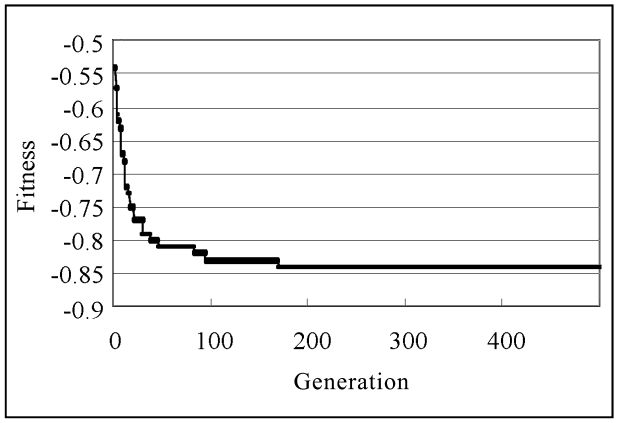

Figure 7 (a) Frequencies of every amino acid at each position: (Table 2 peptides). (b) Genetic algorithm (GA) conditions. (c) GA evolution (training set: Table 2).

peptides), shown in Table 5. The GA, executed under the same conditions as in the case of HLA-A24, leads to the BM illustrated in Table 5. The parameters for the GA are similar to those used previously (Fig. 7b). Figure 10 illustrates the correlation between experimental BA and computed ones. A correlation coefficient of 0.84 was obtained in this case.

Analysis of the information contained in Table 6 also leads to the determination of primary and secondary anchors in peptides binding to HLA-A2.1, according to their scores in the BM.

Parker *et al*. have performed an extensive analysis of HLA-A2.1 complexes based on experimental peptide binding data, deriving a table of coefficients representing the contribution of each amino acid in the binding to HLA-A2.1.[5] Since the methods of Parker *et al*. share some similarities with the BM proposed here, and because their analysis was performed over extensive numbers and types of peptides, it constitutes a suitable source for comparing the results yielded by our methodology and the conclusions derived from it.

Performing a correlation analysis on the experimental values for the peptides reported in the work of Parker *et al*., and the BA values calculated using the BM scores of Table 6, we found a correlation (Fig. 10). This correlation is comparable to that obtained by Parker *et al*. with their own predicted values.[5] Furthermore, an analysis of the scores of the BM shown in Table 6 enabled the extraction and generalization of the structural characteristics that peptides with high BA to MHC class I molecule HLA-A2.1 must possess.

First, as in the case of HLA-A24, primary anchor peptides can be found easily by inspection of the BM. Thus, primary anchor amino acids at P2 are Leu and Val (scores 2.22 and 1.02, respectively; Table 6), and at P9 are Phe and Val (scores 1.03 and 1.47, respectively; Table 6). Because the methodology

Table 3 Optimal similarity matrix generated by artificial evolution

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	1.00	0.05	0.96	0.99	0.09	0.23	0.11	0.06	0.71	0.77	0.71	0.96	0.24	0.71	0.95	0.95	0.41	0.74	0.71	0.70
R	0.05	1.00	0.09	0.06	0.96	0.82	0.94	0.99	0.34	0.28	0.34	0.09	0.81	0.34	0.00	0.00	0.64	0.31	0.34	0.35
N	0.96	0.09	1.00	0.97	0.12	0.27	0.15	0.10	0.75	0.80	0.74	1.00	0.27	0.75	0.91	0.91	0.44	0.78	0.74	0.74
D	0.99	0.06	0.97	1.00	0.10	0.24	0.12	0.07	0.72	0.77	0.72	0.97	0.24	0.72	0.94	0.94	0.41	0.75	0.71	0.71
C	0.09	0.96	0.12	0.10	1.00	0.86	0.98	0.98	0.38	0.32	0.38	0.13	0.85	0.38	0.04	0.04	0.68	0.34	0.38	0.38
Q	0.23	0.82	0.27	0.24	0.86	1.00	0.88	0.84	0.52	0.46	0.52	0.27	0.99	0.52	0.18	0.18	0.82	0.48	0.52	0.52
E	0.11	0.94	0.15	0.12	0.98	0.88	1.00	0.95	0.40	0.34	0.40	0.15	0.87	0.40	0.06	0.06	0.70	0.37	0.40	0.41
G	0.06	0.99	0.10	0.07	0.98	0.84	0.95	1.00	0.35	0.30	0.36	0.10	0.83	0.35	0.01	0.01	0.66	0.32	0.36	0.36
H	0.71	0.34	0.75	0.72	0.38	0.52	0.40	0.35	1.00	0.94	1.00	0.75	0.53	1.00	0.66	0.66	0.69	0.97	1.00	0.99
I	0.77	0.28	0.80	0.77	0.32	0.46	0.34	0.30	0.94	1.00	0.94	0.81	0.47	0.94	0.72	0.72	0.64	0.98	0.94	0.94
L	0.71	0.34	0.74	0.72	0.38	0.52	0.40	0.36	1.00	0.94	1.00	0.75	0.53	1.00	0.66	0.66	0.70	0.96	1.00	1.00
K	0.96	0.09	1.00	0.97	0.13	0.27	0.15	0.10	0.75	0.81	0.75	1.00	0.28	0.75	0.91	0.91	0.45	0.78	0.75	0.74
M	0.24	0.81	0.27	0.24	0.85	0.99	0.87	0.83	0.53	0.47	0.53	0.28	1.00	0.53	0.19	0.19	0.83	0.49	0.53	0.53
F	0.71	0.34	0.75	0.72	0.38	0.52	0.40	0.35	1.00	0.94	1.00	0.75	0.53	1.00	0.66	0.66	0.70	0.96	1.00	1.00
P	0.95	0.00	0.91	0.94	0.04	0.18	0.06	0.01	0.66	0.72	0.66	0.91	0.19	0.66	1.00	1.00	0.36	0.69	0.66	0.65
S	0.95	0.00	0.91	0.94	0.04	0.18	0.06	0.01	0.66	0.72	0.66	0.91	0.19	0.66	1.00	1.00	0.36	0.69	0.66	0.65
T	0.41	0.64	0.44	0.41	0.68	0.82	0.70	0.66	0.69	0.64	0.70	0.45	0.83	0.70	0.36	0.36	1.00	0.66	0.70	0.70
W	0.74	0.31	0.78	0.75	0.34	0.48	0.37	0.32	0.97	0.98	0.96	0.78	0.49	0.96	0.69	0.69	0.66	1.00	0.96	0.96
Y	0.71	0.34	0.74	0.71	0.38	0.52	0.40	0.36	1.00	0.94	1.00	0.75	0.53	1.00	0.66	0.66	0.70	0.96	1.00	1.00
V	0.70	0.35	0.74	0.71	0.38	0.52	0.41	0.36	0.99	0.94	1.00	0.74	0.53	1.00	0.65	0.65	0.70	0.96	1.00	1.00

proposed here takes into account the contribution of every amino acid to the overall BA of the peptide, contributions by auxiliary anchor amino acids are also automatically taken into account. All contributions can be determined from the BM by setting ranges for primary and secondary or auxiliary anchors.

The prediction ability of the automatically obtained model was derived using peptides reported from CEA (carcynoembryonic antigen) reported by Kawashima *et al*. (Table 7) that bind to HLA-21.1 but which are not included in the training set used to obtain the model (Table 5). The predictive ability of the model is remarkable, as can be seen in Fig. 11, which is a plot of the correlation of values predicted by the model and the experimental values reported in the literature.[5] The correlation of predicted values to experimental values is high in this case, which shows the high predictive ability of the methodology.

Receptor–ligand interaction analysis and binding affinity profiles

An analysis of the interaction interface for complexes with peptides of several degrees of affinity (both experimental and computed) was performed for sets of peptides included in the object set for HLA-A2.1 (Table 8). After the prediction of the 3-D structure of the peptides using GAX,[18] TINKER[20] and AMBER,[19] the molecules were docked into the receptor groove using MIAX.[11–14] After the extraction of the interaction interface, we computed the electrostatic and hydrophobic potentials associated with the atoms of the receptor and the ligand. This process for the first peptide in Table 8 is illustrated in Fig. 12, where the docked structure is shown using the backbone of the receptor and ligand. Finally, the binding scores were computed. For the peptide KMVELVHFL, the electrostatic and hydrophobic potentials are shown in Fig. 13. Putative electrostatic and hydrophobic clusters are depicted on the interaction surface, showing the complementarity in the receptor and ligand interaction.[14] The results from repeating the calculation for the other three peptides in Table 8 and plotting the values against the experimental activities of the peptides are shown in Fig. 14. It is evident from these plots that the binding scores (hydrophobic and electrostatic) correlate well to the experimental affinities for the peptides. However, the analysis shows that the correlation is less apparent between the experimental affinity and the scores based on amino acid BM (Fig. 15). Therefore, the docking process and the subsequent analysis of the interaction surface becomes relevant, since the nature of the interaction, as well as the main factors contributing to the bonding, become apparent at the atomic level.

Discussion

We have presented here a new method to express and compute peptide BA or activities with MHC class I molecules. One of the features that characterizes and differentiates this methodology from others is that the contribution of each amino acid to the overall affinity of the peptide is not computed directly from the correlation of the experimental values. An evolutionary process is used to compute similarities among amino acids at positions determined within the peptide sequence from the BA point of view. This new

Table 4 Binding matrix for HLA-A24 binding 9-mer peptides

Amino acid	Amino acid position within the peptide 1st	2nd	3rd	4th	5th	6th	7th	8th	9th
A	0.58	– 0.13	0.58	0.48	0.25	0.13	– 0.27	0.42	– 0.13
R	– 1.87	– 1.37	– 1.85	– 1.82	– 1.78	– 1.7	– 1.58	– 1.79	– 1.33
N	0.64	– 0.01	0.64	0.66	0.35	0.28	0.04	0.58	– 0.02
D	0.55	– 0.11	0.57	0.52	0.24	0.18	– 0.22	0.46	– 0.11
C	– 1.58	– 1.24	– 1.62	– 1.63	– 1.55	– 1.5	– 1.32	– 1.53	– 1.21
Q	– 0.95	– 0.79	– 0.87	– 0.90	– 0.86	– 0.88	– 0.49	– 0.98	– 0.77
E	– 1.44	– 1.17	– 1.52	– 1.50	– 1.43	– 1.38	– 1.15	– 1.46	– 1.14
G	– 1.78	– 1.32	– 1.76	– 1.76	– 1.69	– 1.57	– 1.47	– 1.71	– 1.29
H	0.77	0.77	0.77	0.82	0.98	**1.08**	**1.21**	0.87	0.72
I	0.77	0.63	0.77	0.82	0.93	0.95	0.98	0.80	0.97
L	0.83	0.77	0.80	0.88	**1.05**	**1.14**	**1.24**	0.92	**2.70**
K	0.63	0.00	0.67	0.61	0.38	0.27	0.07	0.60	– 0.01
M	– 0.91	– 0.76	– 0.86	– 0.88	– 0.82	– 0.84	– 0.48	– 0.94	– 0.75
F	0.77	**1.03**	0.76	0.81	1.00	**1.07**	**1.21**	0.93	**1.10**
P	0.40	– 0.30	0.38	0.26	– 0.02	– 0.11	– 0.74	0.23	– 0.29
S	0.41	– 0.30	0.42	0.25	– 0.01	– 0.14	– 0.74	0.26	– 0.29
T	– 0.11	– 0.21	– 0.17	– 0.05	0.06	– 0.09	0.22	– 0.22	– 0.22
W	0.76	**1.22**	0.78	0.80	0.93	0.98	**1.06**	0.83	0.64
Y	0.77	**2.51**	0.75	0.80	1.00	**1.05**	**1.22**	0.86	0.72
Z	0.79	0.77	0.77	0.81	0.98	**1.09**	**1.21**	0.87	0.71

Table 5 Object set of HLA-A2.1 binding 9-mer peptides

Sequence	Binding capacity (nM IC50)	Sequence	Binding capacity (nM IC 50)
VVLGVVFGI	14.3	YMIMVKCWM	217.4
QLFEDNYAL	17.2	ALIHHNTHL	238.1
RLLQETELV	20.8	ILDEAYVMA	238.1
VLIQRNPQL	22.7	GLACHQLCA	416.7
KIFGSLAFL	33.3	KMVELVHFL	9.8
SIISAVVGI	69.4	KASEYLQLV	151.5
ILHNGAYSL	74.6	KIWEELSML	166.7
ALCRWGLLL	100.0	FLWGPRALI	238.1
CLTSTVQLV	147.1	FLWGRPALV	31.3
HLYQGCQVV	147.1	KVAELVHFL	68.4
QLMPYGCLL	217.4	YIFATCLGL	185.2

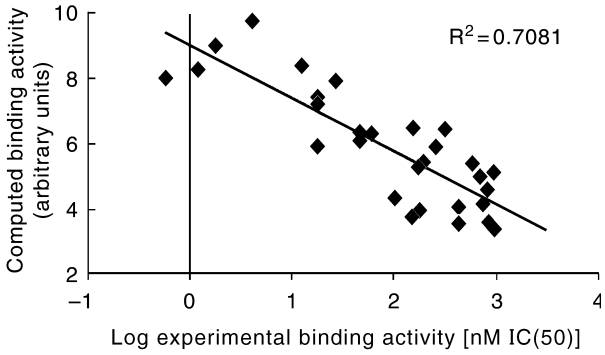

Figure 8 Relationship between computed versus experimental affinities for HLA-A24 binding peptides.

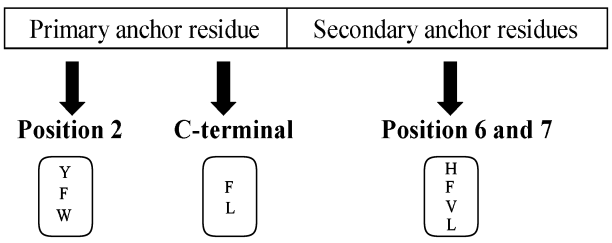

Figure 9 Anchor amino acids for HLA-A24 binding 9-mer peptides.

element in determining the BM has the effect of reducing the amount of data in the training set that is necessary for a good learning process (dictated by a rigorous statistical analysis).[6] This is demonstrated not only by the remarkable performance in predicting the BA of peptides to the MHC class I molecules HLA-A24 and HLA-A2.1 using the two models, but by the fact that abundant data on the similarity of amino acids are used to derive the model (since the number of physicochemical coefficients characterizing each amino acid exceed 400 coefficients and all of them are used recursively during the artificial evolution process).

Calculation of the BM allows the direct prediction of the BA of peptides to particular MHC class I molecules, taking into consideration the primary sequence of the amino acid residues only. However, this methodology (and by extension similar BM-based methodologies) can, as a result of the averaging nature of the relationship sought by the evolutionary algorithm, have some intrinsic limitations, as illustrated by the four MAGE2 peptides considered here. To overcome this problem, we went a step further and suggested a process to evaluate affinities based on a meticulous analysis of the receptor–ligand interaction surfaces, extracted from plausible artificially docked complexes. To our knowledge, a methodology of this nature has not previously been considered.

Table 6 Binding matrix for HLA-A2.1 binding 9-mer peptides

Amino acid	Amino acid position within the peptide								
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th
A	0.45	– 0.62	– 0.31	– 0.54	0.60	0.53	– 0.33	– 0.40	– 0.44
R	0.32	– 0.78	– 0.46	– 0.63	0.48	0.50	– 0.51	– 0.54	– 0.68
N	1.04	0.6	0.93	0.66	**1.12**	**1.36**	0.91	**1.01**	0.81
D	0.10	0.89	– 0.61	– 0.84	0.29	0.30	– 0.67	– 0.69	– 0.81
C	– 1.85	– 1.65	– 2.09	– 2.03	– 1.69	– 0.82	– 1.98	– 1.78	– 1.71
Q	0.16	0.54	0.56	**1.01**	– 0.23	– 0.49	0.79	0.49	0.43
E	– 0.82	0.28	0.13	0.84	– 0.87	– 1.30	0.13	– 0.15	0.08
G	0.80	– 0.17	0.36	0.33	**1.14**	**1.01**	0.14	0.48	– 0.02
H	– 2.00	– 0.08	– 0.57	– 0.03	– 1.89	– 2.09	– 0.62	– 0.68	– 0.35
I	– 0.87	0.63	0.16	0.51	– 1.06	– 1.41	0.00	– 0.23	0.29
L	**1.06**	**2.22**	**1.20**	0.78	**1.05**	**1.09**	**1.29**	**1.57**	2.05
K	– 0.20	– 1.20	– 1.22	– 1.34	– 0.45	– 0.10	– 1.20	– 1.16	– 1.18
M	0.19	0.79	0.68	0.95	– 0.09	– 0.36	0.78	0.65	0.6
F	**1.10**	0.91	**1.22**	0.82	0.85	0.79	**1.21**	**1.36**	**1.03**
P	– 1.65	– 1.58	– 1.96	– 1.90	– 1.41	– 0.76	– 1.85	– 1.71	– 1.63
S	0.82	0.80	**1.03**	0.99	0.59	0.38	**1.1**	**1.07**	0.85
T	– 0.24	– 1.09	– 0.98	– 1.17	– 0.12	0.06	– 0.98	– 1.00	– 1.05
W	0.79	– 0.09	0.54	0.22	**1.14**	0.98	0.22	0.48	0.07
Y	– 0.40	0.36	0.30	0.69	– 0.61	– 1.02	0.38	0.03	0.19
V	1.21	1.02	1.10	0.71	1.16	1.37	1.18	1.23	1.47

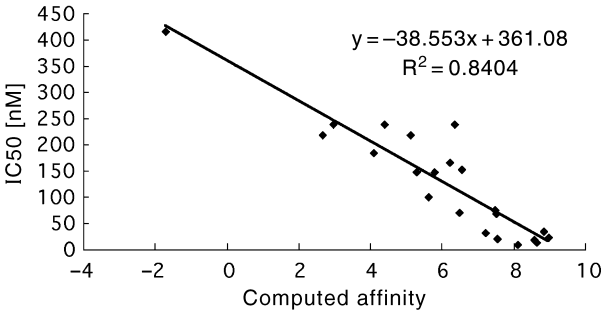

Figure 10 Relationship between computed versus experimental affinities for HLA-A2.1 binding peptides.

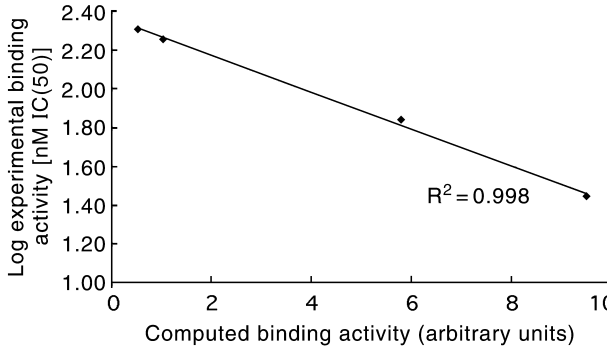

Figure 11 Correlation of binding matrix (BM) scores and experimental affinities for HLA-CEA.

Table 7 Sequences from HLA-CEA used in evaluating affinities to HLA-21.1

Sequence	Experimental binding activity	Log experimental binding activity	Computed binding activity (arbitrary units)
YLSGANLNL	27.8	1.44	9.49
IMIGVLVGV	68.5	1.84	5.79
LLTFWNPPT	178.6	2.25	1.05
VLYGPDTPI	200.0	2.30	0.55

Table 8 Peptides used in complex formation assessment

Sequence	Experimental binding activity	Log experimental binding activity	Computed binding activity HLA-A2.1BM
KMVELVHFL	9.8	0.99	7.47
KASEYLQLV	151.5	2.18	5.36
KIWEELSML	166.7	2.22	5.83
FLWGPRALI	238.1	2.38	7.79

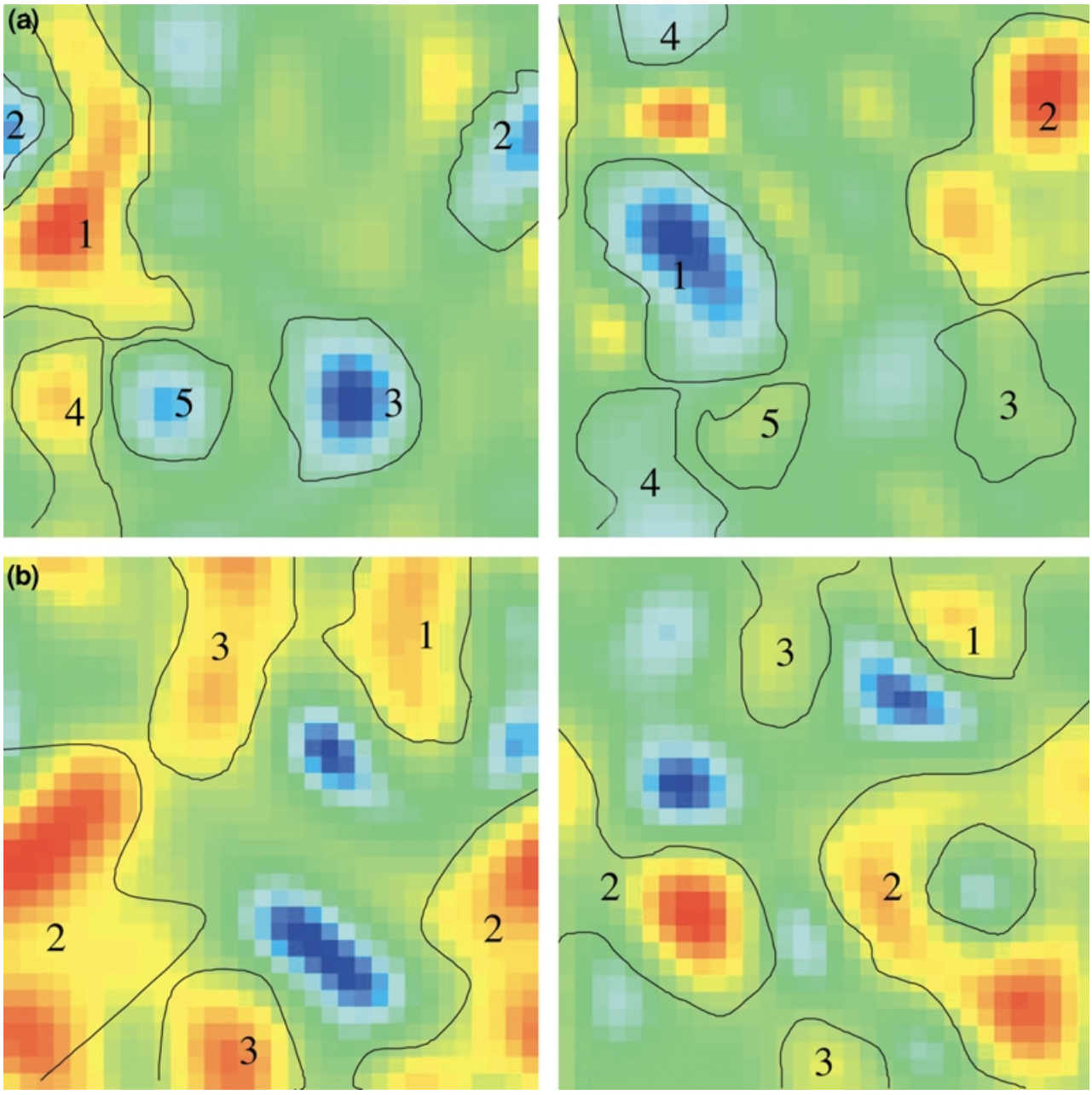

Figure 13 (a) Electrostatic potential on the interaction interface. Left panel: receptor; right panel: ligand; red, negative regions; blue, positive regions; green, neutral. (b) Hydrophobic potential on the interaction interface Left panel: receptor, right panel: ligand; red, hydrophobic regions; blue, hydrophilic regions.

Automatic systems for docking peptides onto receptor grooves are still the exception rather than the rule, due in part to the high flexibility that characterizes these molecular systems. Therefore, the relevance of the system MIAX, developed by Del Carpio *et al.*,[11–14] becomes apparent as it takes into account the flexibility of the ligand and the receptor during the process of docking. However, the process is cost-demanding because of the complexity of the system and has to be limited to a small number of structures.

This problem is further complicated by the fact that, hitherto, more than 200 types of HLA class I and class II molecules have been identified.[27,28] Consequently, a binding profile matrix such as the one proposed here remains quite valuable as it can assist in the reduction of the number of computer experiments (thereby the number of laboratory experiments) that must be performed in order to characterize the antigenic properties of series of peptides and antigens.

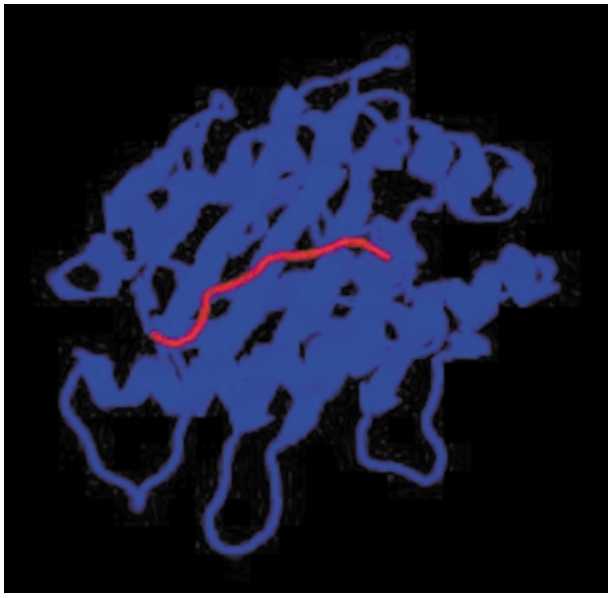

Figure 12 Predicted complex KMVELVHFL-HLA-A2.1.

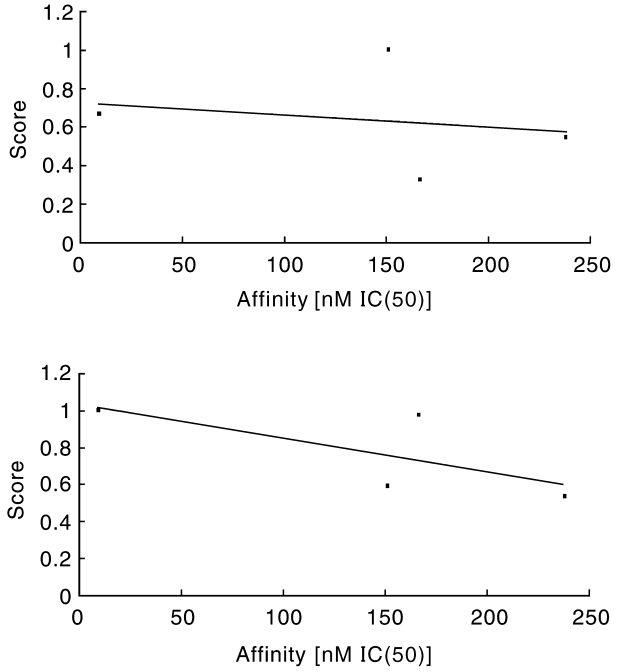

Figure 14 (a) Correlations of electrostatic interaction scores with binding affinities for 9-mers (MAGE2). (b) Correlations of hydrophobic interaction scores with binding affinities for 9-mers (MAGE2).

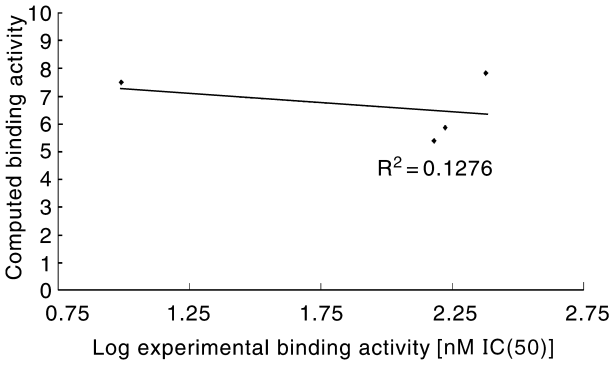

Figure 15 Correlation of binding matrix (BM) scores and experimental affinities for MAGE2.

References

1 Brusic V, Rudy G, Honeyman M, Hammer J, Harrison L. Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network. *Bioinformatics.* 1998; **14**: 121–30.

2 Engelhardt VH, Beier DC, Chen Y, Henderson RA, Huczko E, Slinglaff CL. Sturcture of peptides associated with MHC class I molecules. *Curr. Opin. Immunol.* 1994; **6**: 13–23.

3 Cresswell P. Assembly, transport, and function of MHC class II molecules. *Ann. Rev. Immunol.* 1994; **12**: 259–93.

4 Chicz RM, Urban RG, Gorga JC, Vignali AA, Lane WS, Strominger JL. Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. *J. Exp. Med.* 1993; **178**: 27–47.

5 Parker KC, Bednarek MA, Coligan JE. Scheme for ranking potential HLA-binding peptides based on independent binding of individual peptide side-chains. *J. Immunol.* 1994; **152**: 163–75.

6 Brusic V, Schoenbach C, Takiguchi M, Ciesielski V, Harrion C. Application of genetic search in derivation of matrix models of peptide binding to MHC molecules. *ISMB* 1997.

7 Kondo A, Sidney J, Southwood S *et al.* Prominent roles of secondary anchor in peptide binding to HLA-A24 human class molecules. *J. Immunol.* 1995; **155**: 4307–12.

8 Nokihara K, Yamaguchi M, Schönbach C, Ibe M, Shiga H, Takiguchi M. Highly efficient multiple peptide synthesis and characterization of anchor and non-anchor residues on peptide binding to HLA class I molecules (HLA-B*3501). In PTP Kaumaya and RS Hodges (eds). *Peptides: Chemistry, Structure and Biology*. Mayflower Scientific, 1996; 820–2.

9 Goldberg DE. *Genetic Agorithms in Search, Optimization, and Machine Learning*. Reading, Massachusets: Addison-Wesley, 1989.

10 Gribskov M, McLachlan AD, Eisenberg D. Profile analysis: Detection of distantly related proteins. *Proc. Natl Acad. Sci. USA* 1987; **84**: 4355–8.

11 Del Carpio CA, Yoshimori A. MIAX. A novel system for assessment of macromolecular interaction in condensed phases. (1) Description of the interaction model and simulation algorithm. *Genome Inform.* 1999; **10**: 3–12.

12 Del Carpio CA, Yoshimori A. MIAX. A system for assessment of macromolecular interaction in condensed phases. (3) A parallel hybrid GA for flexible protein docking. *Genome Inform.* 2000; **11**: 205–14.

13 Yoshimori A, Del Carpio CA. Automatic epitope recognition in proteins oriented to the system for macromolecular interaction assessment MIAX. *Genome Inform.* 2001; **12**: 113–22.

14 Del Carpio CA, Yoshimori A. MIAX: A novel paradigm for modeling bio–macromolecular interactions and complex formation in condensed phases. *Proteins. Stucture, Function Genetic* 2002 (in press).

15 Saper MA, Bjorkman PJ, Wiley D. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J. Mol. Biol.* 1991; **219**: 277–319.
16 Grantham R. Amino acid difference formular to help explain protein evolution. *Science*. 1974; **185**: 862–4.
17 Ogiwara A, Uchiyama J, Takagi T, Kanahisa M. Construction and analysis of a profile library characterizing groups of structurally known proteins. *Protein Sci.* 1996; **5**: 1991–9.
18 Del Carpio CA. A parallel genetic algorithm for polypeptide three dimensional structure prediction. A transputer implementation. *J. Chem. Inf. Comput. Sci.* 1996; **36**: 258–69.
19 Kollman P, Pealman DA, Case DA *et al.* 'AMBER', a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to stimulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* 1995; **91**: 1–42.
20 Ponder JW, Richards FM. An efficient Newton-like method for molecular mechanics energy minimization of large molecules. *J. Comput. Chem.* 1987; **8**: 1016–24 (On-line) http://dasher.wustl.edu/tinker/, 1995: 02/02/09.
21 Pappu RV, Hart RK, Ponder JW. Analysis and application of potential energy smoothing for gobal optimization. *J. Phys. Chem. B* 1998; **102**: 9725–42.
22 Dudek M, Ramnarayan K, Ponder JW. Protein structure prediction using a combination of sequence homology and global energy minimization. II. Energy functions. *J. Comput. Chem.* 1998; **19**: 548–73.
23 Del Carpio CA, Takahashi Y, Sasaki S. A new approach to the automatic identification of candidates for ligand receptor sites in proteins: (I) Search for pocket regions. *J. Mol. Graphics* 1993; **11**: 23–9.
24 Kohonen T. The Self-Organizing Map. *Proceeds of the IEEE* 1990; **78**: 1464–80.
25 Katsuta Y, Yoshimori A, Del Carpio CA. Scoring protein-substrate complex decoys based on electrostatic and hydrophobic complementarities: evaluation of hormone–receptor interactions. *Genome Inform.* 2001; **12**: 356–7.
26 Maier R, Falk K, Roetzchke O *et al.* Peteide motifs of HLA-A3, A24, and B7 molecules as determined by pool sequencing. *Immunogenetics* 1994; **40**: 306–8.
27 Kawashima I, Hudson SJ, Tsai V *et al.* The multi-epitope approach for immunotherapy for cancer: Identification of several CTL epitopes from various tumour-associated antigens expressed on solid epithelial tumors. *Human Immunology* 1998; **59**: 1–14.
28 Mallios RR. Class II MHC quantitative binding motifs derived from a large molecular database with a versatile iterative stepwise discriminant analysis meta-alogrithm. *Bioinformatics* 1999; **15**: 432–9.