Journal of Computer-Aided Molecular Design, **16:** 229–243, 2002.
KLUWER/ESCOM

Recovery of known T-cell epitopes by computational scanning of a viral genome

Antoine Logean[a,b] & Didier Rognan[a,*]
[a]*Bioinformatic Group, Laboratoire de Pharmacochimie de la Communication Cellulaire, UMR CNRS 7081, 74 route du Rhin, B.P.24, F-67401 Illkirch, France;* [b]*Department of Applied Biosciences, Swiss Federal Institute of Technology, Winterthurerstrasse 190, CH 8057 Zürich, Switzerland*

Received 24 January 2002; accepted 4 June 2002

Key words: antigen, epitope prediction, free energy scoring, homology modelling, major histocompatibility complex, threading

Summary

A new computational method (EpiDock) is proposed for predicting peptide binding to class I MHC proteins, from the amino acid sequence of any protein of immunological interest. Starting from the primary structure of the target protein, individual three-dimensional structures of all possible MHC-peptide (8-, 9- and 10-mers) complexes are obtained by homology modelling. A free energy scoring function (Fresno) is then used to predict the absolute binding free energy of all possible peptides to the class I MHC restriction protein. Assuming that immunodominant epitopes are usually found among the top MHC binders, the method can thus be applied to predict the location of immunogenic peptides on the sequence of the protein target. When applied to the prediction of HLA-A*0201-restricted T-cell epitopes from the Hepatitis B virus, EpiDock was able to recover 92% of known high affinity binders and 80% of known epitopes within a filtered subset of all possible nonapeptides corresponding to about one tenth of the full theoretical list.

The proposed method is fully automated and fast enough to scan a viral genome in less than an hour on a parallel computing architecture. As it requires very few starting experimental data, EpiDock can be used: (i) to predict potential T-cell epitopes from viral genomes (ii) to roughly predict still unknown peptide binding motifs for novel class I MHC alleles.

Introduction

Class I Major Histocompatibility Complex (MHC)-encoded proteins form a family of highly polymorphic glycoproteins whose function is to selectively bind and present antigenic peptides (epitopes) to cytotoxic T lymphocytes (CTLs), thus allowing the immediate detection of intracellular pathogens [1–3]. By opposition to B-cell epitopes that are located on the outer surface of circulating proteins, T-cell epitopes are usually parts of the protein core. Hence foreign proteins are first degraded into short peptides (about 15 amino acids long) by the proteasome [4], then transported to the endoplasmic reticulum by an ATP-dependent transporter (TAP) [5] and finally loaded into recently-synthesized class I MHC proteins [3]. Remarkably, the targets involved in epitope processing show a selectivity towards specific peptide patterns that increases along the processing machinery. Thus, the proteasome is relatively permissive for peptide cleavage [6] whereas the MHC proteins show an exquisite selectivity for a specific peptide binding motif [7]. It is therefore logical that high MHC binding affinity often correlates with immunogenicity [8–10] and more specifically with the stability of MHC-peptide complexes [11]. Prediction of MHC-binding properties is therefore a necessary step towards the rational design of peptide vaccines aimed at boosting the immune response against a foreign antigen. Since experimental testing of all overlapping peptides spanning a whole

*To whom correspondence should be addressed. E-mail: didier.rognan@pharma.u-strasbg.fr

protein sequence is usually out of reach, considerable effort has gone into the development of computer-based methods [12] that predict either proteasome, TAP or MHC preferences for a set of specific peptides. This study will only focus on MHC binding prediction since programs able to quantitatively predict MHC binding affinity should remarkably downsize the number of potential epitope candidates to test experimentally. Up to now, algorithms aimed at predicting MHC binders can be divided in two main groups (sequence-based or structure-based methods) depending on the nature of the experimental information that is required.

Sequence-based methods rely on the primary sequence of peptides known to experimentally bind to a target MHC protein. This information can be easily encoded into either a binding motif [13, 14], a position-dependent matrix [15–17], or an artificial neuronal network (ANN) [15–17]. Using databases derived from naturally-bound peptides [7] or synthetic peptide libraries [18, 19], each amino acid of the peptide is usually scored using a matrix taking into account the relative contribution of any amino acid at every peptide position, and finally summing up the contribution of all peptide positions. This approach has two main drawbacks: (i) it assumes that different peptide positions contribute in an additive manner to the overall binding affinity and neglect the experimentally-demonstrated interplay between different peptide side chains [20], (ii) its real predictive power is directly dependent on the amount of experimental data used to interpolate MHC binding properties. Thus, MHC alleles for which rather few experimental data are available are unsuitable for sequence-based prediction methods.

The second group of prediction methods uses the three-dimensional (3-D) structure of existing MHC-peptide complexes. The current release of Protein Data Bank [21] comprises a total of 63 class I MHC-peptide complexes. These crystallographic structures enable a better understanding of the structural principles governing peptide recognition by class I MHC molecules [22]. Most of the peptides selected by class I MHC molecules are from 8 to 10 amino acids long and present anchoring backbone atoms at both termini, and MHC-binding side chains at positions 2 and at the C-terminus [23, 24]. Auxiliary anchors at P_1 and P_3 (P_n: peptide position n) usually fine tune peptide recognition [25, 26]. Each anchoring side chain interacts with one of the 6 polymorphic MHC pockets [24, 27], whose location has been conserved along evolution but whose physicochemical properties are highly variable and thus ensure allele specificity [28].

One of the first structure-based computational approach, based on molecular dynamics (MD) simulation of MHC-peptide complexes [29] allowed a crude discrimination of binders from non-binders but was not suitable for high-throughput predictions. A more recent approach involved threading of the peptides through an X-ray template and evaluation of their binding by statistical pairwise potentials [30, 31]. However, it does not allow the direct prediction of binding affinity values. Last, methods using computational combinatorial ligand design (CCLD) [32] for placing amino acids in specificity pockets, or based on three-dimensional quantitative structure-affinity relationship (3-D QSAR) of MHC-peptide complexes [33] have also been developed. Excepted for the previously-described threading methods, all approaches developed up to now are not suitable for the systematic high-throughput scanning of simple genomes because they are either restricted to MHC alleles for which a large body of experimental information is already available (all sequence-based and QSAR methods) or still are much too slow (MD and CCLD-based approaches). We propose here a new structure based approach (EpiDock) similar in its spirit to a previously described threading approach [30, 31] but able to quantitatively predict binding affinities and fast enough to ensure complete scanning of entire genomes for potential high-affinity binders to any class I MHC allele. Homology models of class I MHC alleles in complex with all possible peptides (from 8 to 10 amino acids long) are first constructed by searching three separate 3D databases, quickly energy-minimized and then scored using a tailor-made free energy scoring function (Fresno)[34, 35]. When applied to the identification of HLA-A*0201-restricted T-cell epitopes from the Hepatitis B virus (HBV) genome, EpiDock was able to recover 80% of experimentally-determined T-cell epitopes peptides encoded by that viral genome.

Material and methods

Description of the program

EpiDock is a 1600 lines-containing perl script that automatically prepare the necessary input files, launches sequentially four independent programs: a molecular editor (e.g. SYBYL) [35], a molecular mechanics refinement program (e.g. Amber) [37], a routine for

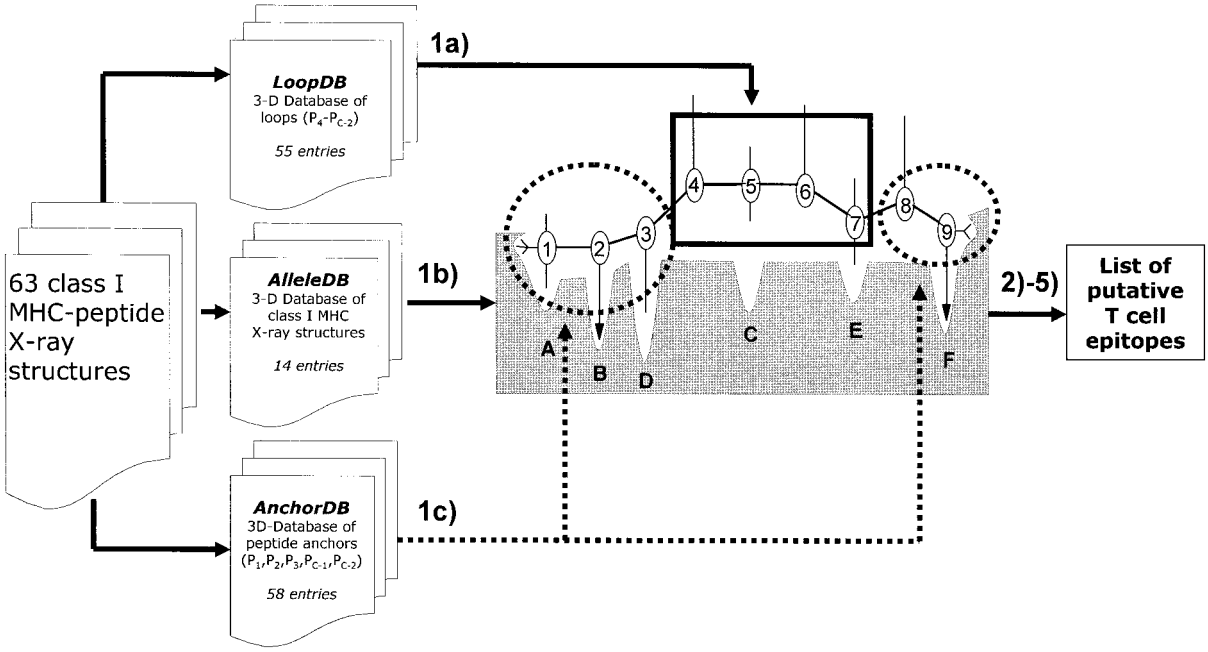

Figure 1. Flowchart of the methods used to predict putative T-cell epitopes: (1) Automated construction of all overlapping peptides (8-, 9-, 10-mers) from the amino acid sequence of a target protein, and knowledge-based homology modelling of all possible MHC-peptide complexes; (2) energy minimisation with Amber6.0 using the Amber95 force-field [55]; (3) computation of the peptide desolvation energy with DelPhi [56]; (4) calculation of the absolute binding free energy (ΔG_{bind}) of every peptide ligand using Fresno [34]; (5) ranking the list of all possible peptides by ΔG_{bind} values.

computing desolvation energies (e.g. DelPhi) [56], and a free energy scoring function (e.g. Fresno) [34], and output the results in from of an ascii list where peptides are ranked by decreasing affinity order for the selected MHC allele.

A flowchart of all computational methods used by EpiDock is displayed in Figure 1. Starting from the primary sequence of any protein of immunological interest, a list of all possible peptides (from 8-mers to 10-mers) is first generated and modelled in the binding groove of the target MHC allele. Automated build-up of all possible MHC-peptide complexes is performed by searching three-dimensional databases (AlleleDB, AnchorDB, LoopDB) for defined amino acid patterns. After energy refinement of all possible MHC-peptide complexes, all peptides are sorted according to their predicted absolute binding free energy computed by an empirical scoring function. Every peptide ranked above a defined binding free energy (affinity) threshold is then considered as a potential T-cell epitope.

Each above-described step of the current method will be now developed in more detail.

Preparation of 3-D databases

The first database AlleleDB is an archive of PDB files which stores 14 X-ray structures of class I MHC alleles currently available in the Protein Data Bank [21]. Alleles for which several entries are available (e.g. HLA-A*0201) have only been stored once, selecting the structure determined with the best resolution. After extracting the corresponding crystal coordinates (antigen binding domain, only) from the Protein Data Bank, hydrogen atoms were automatically added using the SYBYL package [36]. All class I alleles were fitted together on their backbone atoms, in order to defined a common reference frame for further modelling. A second database, AnchorDB (same PDB format as AlleleDB) stores for each crystallized MHC subtype, 3-D amino acid coordinates of anchor positions (P_1, P_2, P_3, P_{C-1} and P_C) as determined in all X-ray structures available for that MHC allele. For example, 23 PDB structures are available for HLA-A*0201, defining a total of 34 AnchorDB entries for A*0201-binding anchoring positions (Table 1). Taking all 14 AlleleDB entries together afforded a total of 58 AnchorDB entries out of the 100 (5×20) possible (Table 2) The PDB entry with the highest resolution

Table 1. AnchorDB entries for HLA-A*0201-binding peptides.

PDB code	Bound peptide[a]
1duy	**LFG**YPV**YV**
1ao7	**LLF**GYPV**YF**
1b0g	**ALW**GFFP**VL**
1bd2	**LLF**GYPV**YF**
1hhg	**TLT**SCNT**SV**
1hhi	**GIL**GFVF**TL**
1hhj, 1akj	**ILK**EPVH**GV**
1hhk	**LLF**GYPV**YV**
1duz, 1im3	**LLF**GYPV**YV**
1i7r	**FAP**GFFP**YL**
1i7t	**ALW**GVFP**VL**
1i7u	**ALW**GFVP**VL**
1ilf	**FLK**EPVH**GV**
1ily	**YLK**EPVH**GV**
1qse	**LLF**GYPR**YV**
1qsf	**LLF**GYPV**AV**
1qr1	**IIS**AVVG**IL**
1qrn	**LLF**GYAV**YV**
1qr1	**IIS**AVVG**IL**
1hhh	**FLP**SDFFP**SV**
1i4f	**GVY**DGRE**ATV**
1jf1	**ELA**GIGI**LTV**
2clr	**MLL**SVPLL**LG**

[a]Anchor positions entered in AnchorDB for HLA-A*0201-binding peptides are displayed in bold face.

was used as template to store a single set of backbone coordinates for anchor positions, as well as to select a single copy of the rotameric state for anchor side chains present several times at the same position (e.g. Leu3 in PDB entries 1hhi and 2clr, Table 1).

At each anchor position, missing amino acids (40 in total) were generated by homology modelling to the closest entry already existing. The rotameric state of the new side chains was directly derived from that of the template while conserving side chain dihedral angles. For example a Thr at P_2 was modelled by homology to Ile2 present in the 1hhi entry (Table 1).

To build the middle part (P_4 to P_{C-2}) of the peptide that bulges out of the binding groove, a third database, LoopDB, was built from loops present in 55 non redundant peptide-class I MHC X-ray structures (Table 3). All loops, of a length of three, four or five residues, share a similar distance between Cα atoms of anchor residues (P_3, P_{C-1}) delimiting the loop window. The LoopDB database is written in TRIPOS protein database format using the mkprodat utility

Table 2. AnchorDB entries from a set of 54 class I MHC-peptide X-ray structures stored in the Protein Data Bank.

Anchor	P_1	P_2	P_3	P_{C-1}	P_C
Residue	G	G	G	G	G
	A	A	A	A	A
	V	V	V	V	V
	I	I	I	I	I
	L	L	L	L	L
	M	P	P	P	M
	F	M	F	Y	F
	Y	Y	Y	S	Y
	S	Q	W	T	
	T	R	S	N	
	Q		T	D	
	E		D	E	
	K		K	K	
	R				

and is searchable by the Loopsearch module of the SYBYL package.

High-throughput construction of MHC-peptide complexes

For a protein of n amino acids, EpiDock will generate all possible n-7 octamers, n-8 nonamers and n-9 decamers overlapping the complete amino acid sequence of the target protein. Then, a combinatorial approach is used to (i) search the AlleleDB database for defining 3-D coordinates of the restriction MHC protein; (ii) to merge into the selected MHC protein, peptide anchor residues starting from the backbone core (P_1-P_3, P_{C-1}-P_C) and adding the side chains from the AnchorDB database, (iii) to complete the peptide building by generating the loop between P_4 and P_{C-2} positions using a previously-described knowledge based loop search procedure [29] that searches the LoopDB database for a loop presenting the same length and distance between the Cα atoms (P_3, P_{C-2}) delimiting the loop window. The loop sequence presenting the highest homology to the target loop was further selected for insertion, missing side chains added using the SYBYL biopolymer dictionary; and all hydrogen atoms were finally added to both the MHC protein and the bound peptide. Special caution was given to polar hydrogen atoms that were added in order to optimise intra and inter-molecular interactions. N-terminal hydrogens of the peptide were notably added in a conservative

Table 3. List of 55 peptides (13 octamers, 36 nonamers and 6 decamers) extracted from a set of 63 class I MHC- peptides X-ray structures used to build the LoopDB database.

Peptide length	MHC allele	PDB entry	Peptide sequence[a]
8-mers	HLA-A*0201	1duy	LFG**YPV**YV
	HLA-B*3501	1a1n	VPL**RPM**TY
	HLA-B*0801	1agb	GGR**KKY**KL
		1agc	GGK**KKY**QL
		1agd	GGK**KKY**KL
		1age	GGK**KKY**RL
		1agf	GGK**KRY**KL
	HLA-B*5l0l	1e28	TAF**TIP**SI
	H-2K[b]	2vaa,2mha	RGY**VYQ**GL
		1vac	SII**NFE**KL
		1osz	RGY**LYQ**GL
		1fo0	INF**DFN**TI
		2ckb	EQY**KFY**SV
9-mers	HLA-A*0201	lhhg	TLT**SCNT**SV
		1hhi	GIL**GFVF**TL
		1hhj, lakj	ILK**EPVH**GV
		1hhk	LLF**GYPV**YV
		1b0g	ALW**GFFP**VL
		1im3, lduz	LLF**GYPV**YV
		1a07	LLF**GYPV**YF
		1bd2	LLF**GYPV**YF
		1qrn	LLF**GYAV**YV
		1i7r	FAP**GFFP**YL
		1i7t	ALW**GVFP**VL
		1i7u	ALW**GFVP**VL
		1jht	ALG**IGIL**TV
		1i1f	FLK**EPVH**GV
		1i1y	YLK**EPVH**GV
		1qr1	IIS**AVVG**IL
		1qse	LLF**GYPR**YV
		1qsf	LLF**GYPV**AV
	HLA-B*2705	1hsa	ARA**AAAA**AA
	HLA-B*3501	1a9b	LPP**LDIT**PY
		1age	LPP**LDIT**PY
	HLA-B*5101	1e27	LPP**VVAK**EI
	HLA-B*5301	1alo	KPI**VQYD**NF
		1alm	TPY**DINQ**ML
	HLA-Cw3	1efx	GAV**DPLL**AL
	HLA-Cw4	1qqd, 1im9	QYD**DAVY**KL
	HLA-E	1mhe	VMA**PRTV**LL
	H-2K[b]	2vab	FAP**GNYP**AL
		1vad	SRD**HSRT**PM
	H-2D[b]	Ihoc	ASN**ENME**TM
		1ce6	FAP**GVFP**YM
		1q1f	FAP**SNYP**AL
		1bz9	FAP**GVFP**YM
	H-2L[d]	1d9	YPN**VNIH**NF
		1ldb	APA**AAAA**AM
		2ldp	QLS**PFPF**DL
10-mers	HLA-A*O201	1hhh	FLP**SDFFP**SV
		1i4f	GVY**DGREH**TV
		1jf1	ELA**GIGIL**TV
		2clr	MLL**SVPLL**LG
	HLA-Aw68	1tmc	EVA**PPEYH**RK
	H-2D[d]	1bii	RGP**GRAFV**TI

[a]Residues in bold face corresponds to bulging parts stored in LoopDB.

manner to ensure optimal hydrogen-bonding to the conserved Tyr7 and Tyr171 side chains [23].

Energy refinement of MHC-peptide complexes

Once built, all MHC-peptide complexes were relaxed as previously described [35] using the AMBER6 program [37]. Briefly, all complexes were first refined by 1000 steps of descent method followed by a maximum of 1000 steps conjugate gradient minimization, unless the rms gradient of the potential energy converged to a threshold of 0.25 kcal mol $Å^{-1}$. Energy refinement of all complexes was performed in vacuum using a distance-dependent dielectric function ($\varepsilon = 4r$), and a twin cut-off (10.0, 15.0 Å) to calculate non-bonded interactions. Using the MPI parallel version of the AMBER6 program on 32 processors of a SGI Origin3800 enables the refinement of about 2000 complexes per hour.

Scoring the peptides according to their binding free energy

We previously developed a method for predicting the binding free energy of peptides to class I MHC proteins using a fast free energy scoring function [34]. The current version of EpiDock uses a slightly modified version of this function (Equation 1), more adapted to high throughput predictions.

$$\Delta G_{bind} = K + \alpha\mathbf{HB} + \beta\mathbf{LIPO} + \gamma\mathbf{ROT} + \delta\mathbf{BP} + \varepsilon\mathbf{DESOLV} + \phi\mathbf{CC} \quad (1)$$

HB (H-bonding), LIPO (lipophilic interactions), ROT (rotational entropy loss), BP (buried polar-apolar contacts) and DESOL V (peptide desolvation energy) represent scores directly computed from the 3-D model of a MHC-peptide complex, whereas K,α-β are regression coefficients previously derived for defined MHC alleles [34, 35]. In the present study, an additional term (CC: close contacts) was added for taking into account possible mismatches between a MHC pocket and anchor side chains that are generated with a fully automated procedure. The CC score computes all close contacts shorter than 2.5 Å between the main two anchoring side chains (P_2, P_C) and their complementary pockets B and F. Multiplying the CC score by a ϕ coefficient of $+2$ gave the best results for discriminating binders from non binders (data not shown).

The previously-described affinity threshold of 500 nM (equivalent to a binding free energy of -35.9 kJ mol^{-1}) [8–10] was used to discriminate immunogenic from non-immunogenic epitopes in the list of all possible octa-, nona- and decamers docked to the target MHC protein.

*Prediction of HLA-A*0201-restricted T-cell epitopes encoded by the HBV genome*

The HBV genome was scanned for HLA-A*0201-restricted T-cell. The prediction was here limited to nonameric peptides from the sequence of three HBV proteins: the DNA polymerase (Swiss-Prot code: DPOL_HPBVJ), the core protein (Swiss-Prot code: VMSA_HPBVJ) and the envelope protein (Swiss-Prot code: CORA_HPBVJ).

*Prediction of the allele-specific peptide motif for two class I MHC alleles (HLA-A*0201, HLA-B*2705)*

Allele specific peptide motifs were predicted by computing the binding free energy of a combinatorial library of 180 MHC-bound polyalanine-derived peptides for which each peptide position of the canonical nonapeptide was singly substituted by any of the 20 naturally-occurring amino acids.

Results

*Recovering HLA-A*0201-restricted T-cell epitopes encoded by the HBV genome*

The binding affinity and immunogenic properties of several HBV peptides, in the context of HLA-A*0201 restriction, has already been reported [9]. Recovering the HLA-A*0201 dependent CTL epitopes by scanning the entire HBV genome was thus an attractive validation test for the present algorithm. Hence, the HBV genome which encodes only three viral proteins is simple enough to be entirely scanned by EpiDock. Moreover, we already reported a HLA-A*0201-specific free energy scoring function [34]. The HBV genome encodes three main proteins, a DNA polymerase (843 residues), an envelope protein (389 residues) and a core protein (183 residues). For each primary sequence, all possible nonamers were defined, docked in the binding groove of HLA-A*0201, and scored as previously described (Figure 2). In order to evaluate the prediction accuracy of the algorithm, peptides were classified in four groups: (1) good binders (predicted $\Delta G_{bind} < -41.6$ kJ mol^{-1}, or predicted $IC_{50} < 50$ nM), (2) intermediate binders (predicted ΔG_{bind} between -41.6 and -35.9 kJ mol^{-1}, or predicted IC_{50} in the 50–500 nM range), (3) weak binders (predicted ΔG_{bind} between -35.9 and -24.5 kJ mol^{-1}, or predicted IC_{50} in the 500–50,000 nM range) and (4) non-binders (predicted $\Delta G_{bind} > -24.5$ kJ mol^{-1}, or predicted $IC_{50} > 50,000$ nM). A binding threshold of 500 nM, proposed to define a reasonable limit to discriminate between immunogenic and non-immunogenic peptides [9] was selected for identifying potential epitopes.

Two peptides from the core protein (Table 4) have been reported to bind to HLA-A*0201 with a high affinity [9]. They are both predicted to bind to the target MHC with an affinity lower than 500 nM (Figures 2A and 2B). All 12 envelope peptides, reported to bind to HLA-A*0201 were also ranked in the pool of good/intermediate binders (Figure 2B). Last, 8 out of the 10 known binders from the DNA polymerase are indeed predicted to bind with an IC_{50} lower than 500 nM ($\Delta G_{bind} < -35.9$ kJ mol^{-1}). A significant fraction of HBV peptides (3–8%) are predicted to bind with very high affinity ($IC_{50} < 50$ nM) to HLA-A*0201 (Table 5). Altogether, 21 out of the 24 HBV peptides known to bind to HLA-A*0201 could be ranked among the top 10% scorers as predicted by EpiDock. Using as selection criterion a predicted bind-

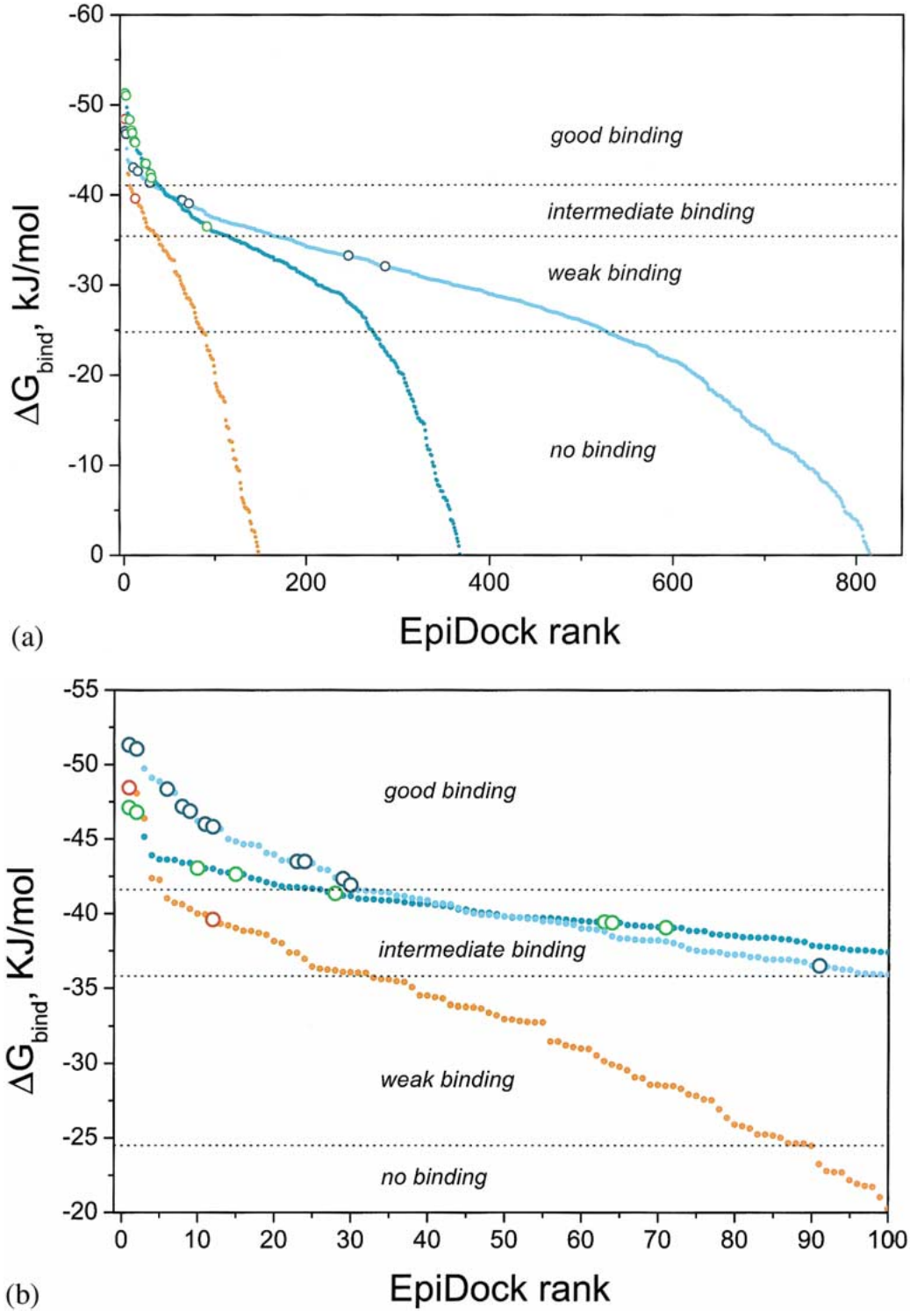

Figure 2. EpiDock ranking of all possible nonapeptides derived from the HBV genome (orange, core protein; greenblue, envelope protein; cyan, polymerase). Peptides known to bind to HLA-A*0201 [9] are displayed by open circles (red, core protein; green, envelope protein; blue, polymerase) Affinity thresholds used to distinguish good, intermediate, weak binders from non-binding peptides [9] are displayed by dotted lines. (A) full ranking of all possible HBV peptides, (B) close-up to the top 100 scorers.

Table 4. List of HBV nonapeptides reported to bind to HLA-A*0201 with an affinity better than 27 000 nM [9]. Peptide sequences able to elicit a CTL response in either A*0201/K^b transgenic mice or A2 patients are underlined.

Protein[a]	Peptide number[b]	EpiDock Rank	Peptide sequence	Pred. ΔG_{bind}[c]	Exp. IC_{50}[d]
CORE	118–127	1	YLVSFGVWI	−48.45	1.9
CORE	100–109	12	LWWFHISCL	−39.62	208
ENV	377–386	1	PLLPIFCL	−51.32	160
ENV	360–369	2	MMWYWGPSL	−51.05	12
ENV	250–259	6	LLLCLIFLL	−48.37	26
ENV	177–186	8	VLQAGFFLL	−47.19	238
ENV	249–258	9	ILLLCLIFL	−46.88	143
ENV	371–380	11	ILSPFMPLL	−46.02	45
ENV	335–344	12	WLSLLVPFV	−45.84	7
ENV	183–192	23	FLLTKILTI	−43.51	7.1
ENV	251–260	24	LLCLIFLLV	−43.50	101
ENV	204–213	29	FLGGTPVCL	−42.35	238
ENV	259–268	30	VLLDYQGML	−41.93	47
ENV	62–71	91	GLLGWSPQA	−36.52	5.8
POL	502–511	1	HLYSHPIIL	−47.12	38
POL	573–582	2	FLLSLGIHL	−46.81	10
POL	770–779	9	WILRGTSFV	−43.06	278
POL	525–534	14	LLAQFTSAI	−42.66	0.52
POL	418–427	28	LLSSNLSWL	−41.37	1087
POL	61–70	63	GLYSSTVPV	−39.46	33
POL	814–823	64	SLYADSPSV	−39.41	14
POL	500–509	70	KLHLYSHPI	−39.08	17
POL	422–431	231	NLSWLSLDV	−33.29	385
POL	453–462	270	GLSRYVARL	−32.09	42

[a]CORE, HBV core protein (Swiss-Prot code: CORA_HPBVJ); POL: DNA polymerase (Swiss-Prot code: DPOL_HPBVJ); ENV, envelope protein (Swiss-Prot code: VMSA_HPBVJ)
[b]SwissProt residue numbering.
[c]Predicted binding free energy, kJ mol^{-1} (this study).
[e]Experimentally-determined IC_{50} value, nM [9].

ing free energy threshold of −35.9 kJ mol^{-1} (IC_{50} < 500 nM) would even afford better predictions with 22 well predicted nonameric peptides. This selection protocol would enable the identification of 80% of all known HLA-A*0201-restricted CTL epitopes encoded by the HBV genome.

*Recovering peptide binding motifs/or HLA-A*0201 and HLA-B*2705 alleles*

EpiDock was next challenged in its ability to predict HLA-A*0201 and HLA-B*2705 peptide binding motifs. As the number of theoretically possible nonapeptides (20^9) far exceeds a manageable level for any prediction algorithm, the current study was limited to the relative contribution of every natural amino acid at each position of a 180 singly-substituted polyalanine peptide library (Figures 3–4, Tables 6 and 7). The computed HLA-A*0201 motif (Figure 4) shows, for the two major anchor position (P_2, P_9), a marked preference for hydrophobic residues (Leu, Met and lie) in agreement with the binding motif determined by pool sequencing of naturally-bound peptides [28] and binding studies on synthetic peptides [25]. The positive contribution of aromatic residues (F, Y, W) as well as the negative impact of negatively-charged amino acids (D, E) at secondary anchor positions (P_1 and P_3) and of aromatic side chains at P_9 is also in remarkable agreement with known binding data [18]. The main discrepancies between predicted and experi-

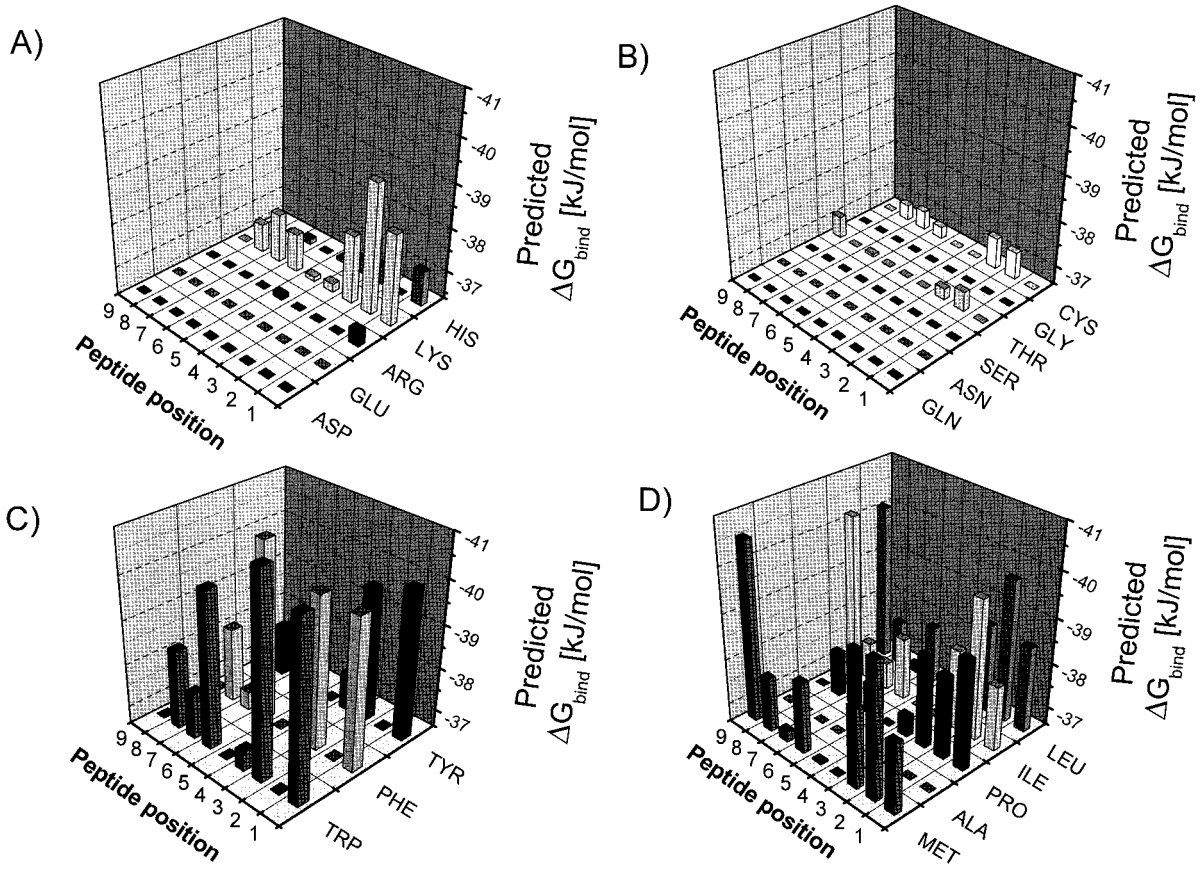

Figure 3. Three-dimensional plots representing the prediction of the allele specific peptide motif for HLA-A*0201. (A) Charged residues, (B) non-charged polar residues, (C) aromatic residues, (D) hydrophobic residues.

mental binding motifs concerns (i) the experimentally-determined negative role of basic side chains (H, K, R) at P_3 [25] which is missed in our prediction, (ii) the binding contribution of Pro at P_1 and Lys at P_2 that are clearly overestimated.

The predicted motif for the second allele investigated herein (HLA-B*2705, Figure 4 and Table 7) shows a unique preference for Arg at P_2, as well as positively-charged (Arg, Lys) or hydrophobic/aromatic side chains (Phe, lie, Met, Lys, Arg) at P_9. Computed preferences for these two main anchors thus nicely fit the experimental binding motif [38] as well as structure-affinity/stability data [39, 40], with the exception of Tyr9 contribution (clearly underestimated) whereas Arg4 and Arg8 roles are overestimated by EpiDock. The important contribution of aromatic residues at secondary anchoring positions (P_1, P_3) are more in agreement with binding data [39, 40] than with the natural peptide binding motif [38] that is biased, especially at the N-terminus, by special preferences for proteasome cleavage and TAP processing. Interestingly, the predicted negative contributions of polar (C, N, Q) and negatively-charged (E, D) side chains at P_1 and P_2, as well as the detrimental effect of Pro/Trp at P_9 is also in agreement with experimental data [39, 40].

Discussion

In the present study, we describe a novel approach for the high-throughput scanning of entire genomes for HLA-restricted T-cell epitopes. A threading approach has been considered to model MHC-peptide X-ray structures as we think it cumulates several advantages. First, it is a structure-based approach that allows the modelling of any class I MHC-peptide complex, at the condition that the peptide is between 8 and 10 amino acids long. Our choice to store structural information

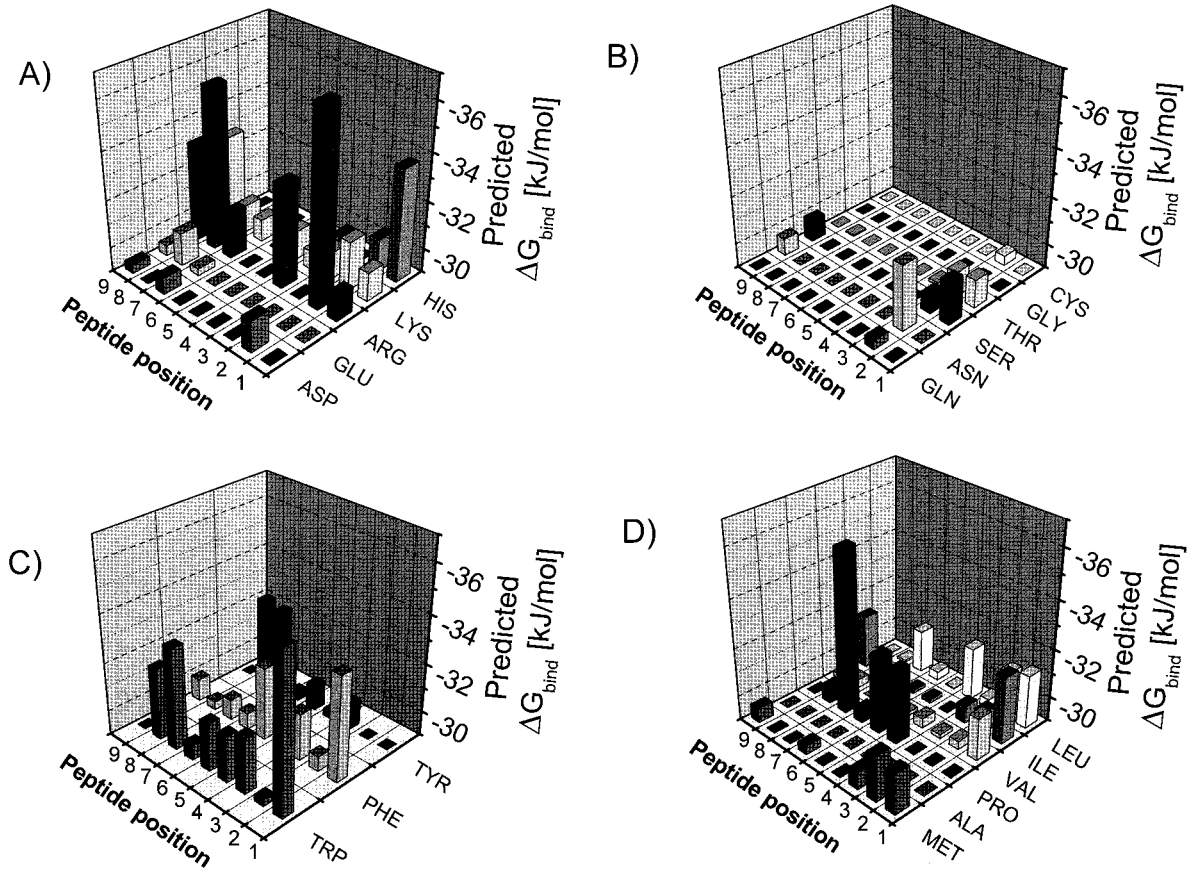

Figure 4. Three-dimensional plots representing the prediction of the allele specific peptide motif for HLA-A*0201. (A) Charged residues, (B) non-charged polar residues, (C) aromatic residues, (D) hydrophobic residues.

about class I MHC-peptide complexes in three separate 3-D databases (AlleleDB, AnchorDB, LoopDB) has been justified by the availability of 63 PDB entries that exactly depict all 3-D information necessary to our threading protocol. Modelling the target protein is a very straightforward procedure due to the very high sequence identity (about 80%) found within class I MHC heavy chains, and the availability of several templates for homology modelling [22, 41]. The averaged root-mean square (rms) deviation of backbone coordinates of the antigen-binding domain is 1.24 Å when considering all class I PDB entries. Thus, modelling any of all possible class I MHC allele from the 14 available X-ray structures present in the AlleleDB database can be fully automated without altering the quality of the homology models. The anchoring positions of the bound-peptide are also easy to model. All available X-ray structures depict a canonical binding mode in which peptide backbone atoms at positions 1, 2, 3, C-1, C-2 (C being the C-terminal residue) are strongly-H-bonded to conserved MHC residues, whereas anchoring side chains (usually at P_2, P_9) provides an allele specific recognition by interacting with specificity pockets [24, 27]. The AnchorDB database will thus store all the 3-D coordinates of backbone anchors (P_1, P_2, P_3, P_{C-1}, P_{C-2}) necessary to initiate peptide building. Corresponding side chains are added depending on their availability in the AnchorDB database (Table 2). For 58 out of the possible 100 side chains, coordinates are simply generated according to the existing X-ray template. When a side chain is not available at a precise position, it is modelled by homology to the closest possible X-ray template. Thus our procedure enables the reliable modelling of nearly half of the bound peptide in a consistent manner, compatible with all available MHC-peptide X-ray structures. Modelling the middle part of the bound-peptide that may either zig-zag [26] or bulge [42] out of the binding groove is less trivial. This part is considered as a loop in EpiDock as it contains mostly T-cell receptor-

Table 5. Prediction of potential HLA-A*0201 T-cell epitopes for the three HBV proteins.

HBV Protein	Core	Envelope	Polymerase
Residues	183	389	843
Possible nonamers	175	381	835
Known binders[a]	2	12	10
Predicted good binders[b]	6 (3,4%)	31 (8,1%)	25 (3%)
Predicted intermediate binders[c]	26 (14,9%)	69 (18,1%)	108 (12,9%)
Predicted weak binders[d]	57 (32,6%)	175 (46%)	232 (27,8%)
Predicted non-binders[e]	86 (49,1 %)	106 (27,8)	469 (56,3%)
EpiDock ranking of known binders	1,12	1,2,6,8,9,11, 12,23,24, 29,30,91	1,2,9,14,28, 63,64,70, 231,270
% of known binders in the top 1 0% scorers	100% (2/2)	92% (11/12)	80% (8/10)
% of known binders predicted as good/intermediate binders	100% (2/2)	100% (12/12)	80% (8/10)
Known CTL epitopes	0	4	6
% of known epitopes predicted as immunogenic[f]	–	100% (4/4)	66% (4/6)

[a] Experimentally-determined IC_{50} lower than 27 000 nM [9].
[b] $\Delta G_{bind} < -41.6$ kJ mol^{-1} ($IC_{50} < 50$ nM).
[c] $-41.6 < \Delta G_{bind} < -35.9$ kJ mol^{-1} ($50 < IC_{50} < 500$ nM).
[d] $-35.9 < \Delta G_{bind} < -24,5$ kJ mol^{-1} ($500 < IC_{50} < 50\,000$ nM range).
[e] $\Delta G_{bind} > -24.5$ kJ mol^{-1} ($IC_{50} > 50\,000$ nM).
[f] $\Delta G_{bind} < -35.9$ kJ mol^{-1} ($IC_{50} < 500$ nM).

anchoring residues [43, 44] that do not participate to MHC binding. We have chosen a knowledge-based procedure to model this part of the peptide, that proved to be able to reproduce within 1 Å rms deviations experimentally-determined loop structures [34]. Altogether, the threading protocol enables the fast and fully-automated building of both the MHC and the bound-peptide from primary amino acid sequences, a necessary condition for scanning entire genomes.

A second advantage of EpiDock is that its predictive power is relatively independent on the amount of previously known experimental data. Sequence-based position-scoring matrices as well as artificial neural networks can only be used when a large body of experimental binding data supports the predictive power of the method. This is not the case for the proposed EpiDock approach for which a minimum of 5 experimentally-determined binding free energies is sufficient indeed to derive a predictive empirical scoring function [34]. In the context of the human MHC project [45] aimed at mapping the peptide specificity of the whole MHC gene, EpiDock may be applied to the prediction of binding motifs for poorly-studied HLA alleles.

Two possible applications have been chosen to explore the accuracy of our approach. First, the full genome of the Hepatitis B virus (HBV) was scanned for potential T-cell epitopes restricted by the well-known HLA-A*0201 allele. As EpiDock basically sort all possible peptides according to their predicted binding free energies, we first looked at the ranking of all HBV peptides previously-demonstrated to bind to the target HLA protein [9]. Out of the 24 binders spread over the sequence of three proteins, 22 were indeed predicted as good/intermediate binders sharing a predicted affinity better than 500 nM (Table 5). The accuracy of EpiDock is thus analogous to that of neural nets specifically trained over 500 peptides towards a unique HLA target [17] and significantly higher than a previously described threading approach [30, 31]. The selected 22 peptides are included in a total list of 265 nonamers representing 18% of all possible HBV nonapeptides. Reducing the size of the hitlist can be achieved without significant loss in accuracy by selecting the top 10% EpiDock scorers (141 peptides in total) that still contains 87.5% of all known binders (21 out of 24 nonamers). To really predict potential T-cell epitopes, we assumed that immunogenic peptides are mostly found among the high-affinity binders. Al-

Table 6. Predicted peptide motif of the HLA-A*0201 allele. The relative contribution to the absolute binding free energy of every amino acid at each position (P_1 to P_9) of a canonical singly-substituted polyalanine peptide is displayed (reference alanine: relative contribution of 100). The best score for each position is indicated in bold face.

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9
Y	110	35	109	103	100	104	**104**	104	14
F	110	87	110	**105**	100	**112**	102	105	50
W	**112**	100	**113**	102	100	110	104	**106**	11
L	106	**109**	106	101	100	104	101	103	110
I	105	**109**	106	101	100	104	102	103	110
V	104	105	104	101	100	102	102	102	107
P	107	106	106	102	101	103	104	103	98
A	100	100	100	100	100	100	100	100	100
M	105	**109**	106	100	101	105	101	104	**111**
C	99	102	103	100	100	101	102	102	99
G	100	100	100	100	100	100	100	100	100
T	101	102	101	100	100	101	101	100	102
S	99	99	99	100	100	100	100	100	99
N	97	95	98	99	100	98	98	97	96
Q	99	91	99	99	100	98	99	99	82
H	102	96	101	101	100	99	100	100	81
K	106	108	104	101	100	102	103	102	87
R	101	83	99	99	100	97	99	98	27
E	100	92	100	99	100	100	99	99	82
D	97	96	98	99	99	99	99	97	96

Table 7. Predicted peptide motif of the HLA-B*2705 allele. The relative contribution to the absolute binding free energy of every amino acid at each position (P_1 to P_9) of a canonical singly-substituted polyalanine peptide is displayed (reference alanine: relative contribution of 100). The best score for each position is indicated in bold face.

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9
Y	86	100	104	101	104	101	112	113	66
F	115	102	107	106	111	**103**	103	101	103
W	**123**	101	**108**	107	107	102	**115**	110	83
L	108	107	101	108	100	101	106	101	96
I	110	103	103	100	101	101	99	100	108
V	106	101	101	102	101	100	100	101	101
P	81	94	91	111	**111**	102	123	103	88
A	100	100	100	100	100	100	100	100	100
M	106	107	103	99	100	101	98	98	102
C	93	102	100	99	98	100	97	96	94
G	100	100	100	100	100	100	100	100	100
T	104	104	101	95	96	99	95	100	98
S	106	103	95	94	97	99	99	99	103
N	99	110	94	100	92	98	92	99	103
Q	93	102	92	94	90	99	95	89	98
H	117	105	98	103	98	99	96	96	96
K	104	107	104	103	104	100	103	104	113
R	104	**128**	99	**115**	92	99	107	**124**	**114**
E	94	94	97	91	95	97	101	105	102
D	96	105	97	96	95	97	103	97	102

though this assumption is sometimes not always true [46], it is generally believed that using an affinity threshold of 500 nM would enable the selection of a vast majority of T-cell epitopes [8–10]. When applied to HLA A*0201-restricted HBV epitopes, EpiDock was able to recover 8 out of the 10 previously described T-cell epitopes (Table 5). Examining the real predictive power of EpiDock (Table 8) in discriminating true from false positives or true from false negatives [15], the proposed method shows a sensitivity and specificity rather similar to neural networks or polynomial methods calibrated on large sets of experimental data [15]. As expected, the ability to identify most of known binders within the smallest possible list (positive prediction ratio) is significantly lower (6%) than that obtained from matrix-based methods (ca. 40–50%). However, by filtering out about 90% of all possible peptides, our approach affords a feasible validation of the epitope prediction by multiple solid-phase synthesis [47] of predicted high affinity binders and further high-throughput experimental testing for MHC binding [48] and T-cell response [49].

A second potential application of EpiDock is the prediction of still unknown peptide binding motifs. Again, two well-known alleles (HLA-A*0201, HLA-B*2705) have been used for a validation purpose. In both cases, the computed binding motifs were in good agreement with experimentally-determined binding data concerning the preference for main anchor residues at P_2 and P_9. Some discrepancies still exist between computed and experimental binding motifs, notably at secondary anchor positions (P_1, P_3) that could be explained by (i) bias towards proteasome cleavage and TAP-processing preferences observed in *in vivo* binding motifs [50], (ii) known discrepancies already observed in the binding motifs determined by either epitope stabilisation or refolding assays [39, 40], (iii) the failure of EpiDock to mirror the protein adaptation to its ligand (a unique set of coordinates is stored for every MHC allele).

Table 8. Measure of goodness [15] of EpiDock assuming a 500 nM cut-off to discriminate binders from non-binders.

HBV protein	Core	Envelope	Polymerase
Total number of peptides (T)	175	381	835
True positives (TP)	2	12	8
True negatives (TN)	143	281	700
False positives (FP)	30	88	125
False negatives (FN)	0	0	2
Sensitivity, % (TP/TP+FN)[a]	100	100	80
Specificity, % (TN/TN+FP)[b]	83	76	85
Positive prediction, % (TP/TP+FP)[c]	6	12	6
Negative prediction, % (TN/TN+FN)[d]	100	100	99
Accuracy, % (TP+TN/T)[e]	83	77	85

[a]Sensitivity: Percentage of all binders correctly identified.
[b]Specificity: Percentage of all non-binders correctly identified.
[c]Positive prediction: Probability that a predicted binder is a true binder.
[d]Negative prediction: Probability that a predicted non-binder is a true non-binder.
[e]Accuracy: Proportion of all predictions that are correct.

The current version of EpiDock is adapted to our hardware/software architecture but could fit nearly any platform at the condition that four software independent from the EpiDock code are available: (1) a molecular editor tool (e.g. Sybyl, InsightII, etc.) for performing '*in silico*' amino acid mutations, (2) a program for refining protein-peptide complexes (e.g. Amber, Discover, etc.), (3) a routine for calculating desolvation energies (e.g. DelPhi, Amsol, etc.), (4) a free energy scoring function (e.g. Fresno, Chemscore, Score, Ludi, etc.). However, changing the main EpiDock executable to mirror any different software environment (e.g. InsightII/Discover/Amsol/Chemscore) is relatively straightforward. Due to the heterogeneous nature of EpiDock components, it is not currently available as a web-server but only as a standalone program available upon request to the authors. Future improvements of EpiDock includes (i) the coupling to a companion module (Epicom) that predicts proteasome cleavage and TAP-binding from available databases [6, 51–54]. The combined used of these two modules should allow to significantly increase EpiDock predictivity by reducing the list of potential T-cell epitopes required for synthesis and experimental testing.

Conclusion

The present method, based on the high-throughput quantitative prediction of binding free energies, is fast enough to scan in a few hours a complete viral genome for class I MHC high affinity binding peptides and thus MHC-restricted T-cell epitopes. The main advantages of our approach over sequence-based positional matrices or neural networks is that it can be applied to class I MHC alleles for which very few experimental data are known. Since the primary sequence of class I MHC proteins is highly conserved, building a 3D model of any class I MHC allele is a rather straightforward procedure. Thus, EpiDock can be easily applied to any class I MHC allele for which no X-ray structure is currently available. By systematically building all possible peptides with a length ranging from 8 to 10 residues, EpiDock should enable the prediction of most high-affinity binders. A clear drawback of the current method is that it does not predict T-cell epitopes *per se*, but only high-affinity binding peptides. Assuming that most immunogenic peptides bind with a high affinity to their MHC target is however supported by several binding studies. Using a predicted binding affinity of 500 nM as a cut-off, we have shown on the example of the HBV genome that EpiDock really miss very few immunogenic peptides. Even if the number of potentially T-cell epitopes is still overestimated, the list is restricted a number of peptides (50–100) that can be easily synthesized by automated multiple peptide synthesizers and further tested for binding affinity and T-cell response. To further restrict the list of potential candidates, EpiDock will be next coupled to a complementary algorithm for predicting proteasome cleavage and TAP-binding.

Acknowledgements

This work is supported by the Schweizerischer Nationalfonds zur Förderung der wissenschaftlichen Forschung (Project 31-57307.99) and grants from the Fondation pour la Recherche Médicale (Paris, France) to DR. The Centre Informatique National de l'Enseignement Supérieur (CINES, Montpellier, France) is acknowledged for allocation of computing time on the Origin3800 supercomputer.

References

1. Heemels, M.T. and Ploegh, H.L., Annu. Rev. Biochem., 64 (1995) 643.
2. Pamer, E. and Cresswell, P., Annu. Rev. Immunol., 16 (1998) 323.
3. Cresswell, P. and Lanzavecchia, A., Curr. Opin. Immunol., 13 (2001) 11.
4. Rock, K.L., Gramm, C., Rothstein, L., Clark, K., Stein, R., Dick, L., Hwang, D. and Goldberg, A.L., Cell, 78 (1994) 761.
5. Neefjes, J.J., Momburg, F. and Hammerling, G.J., Science, 261 (1993) 769.
6. Daniel, S., Brusic, V., Caillat-Zucman, S., Petrovsky, N., Harrison, L., Riganelli, D., Sinigaglia, F., Gallazzi, F., Hammer, J. and van Endert, P.M., J. Immunol., 161 (1998) 617.
7. Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A. and Stevanovic, S., Immunogenetics, 50 (1999) 213.
8. Kubo, R.T., Sette, A., Grey, H.M., Appella, E., Sakaguchi, K., Zhu, N.Z., Arnott, D., Sherman, N., Shabanowitz, J., Michel, H., Bodnar, W.M., Davis, T.A. and Hunt, D.F., J. Immunol., 152 (1994) 3913
9. Sette, A., Vitiello, A., Reherman, B., Fowler, P., Nayersina, R., Kast, W.M., Melief, C.J., Oseroff, C., Yuan, L. and Ruppert, J.J., Immunol., 153 (1994) 5586.
10. Sette, A., Sidney, J., del Guercio, M.F., Southwood, S., Ruppert, J., Dahlberg, C., Grey, H.M. and Kubo, R.T., Mol. Immunol., 31 (1994) 813.
11. van der Burg, S.H., Visseren, M.J., Brandt, R.M., Kast, W.M. and Melief, C.J., J. Immunol., 156 (1996) 3308.
12. Schirle, M., Weinschenk, T. and Stevanovic, S., J. Immunol. Meth., 257 (2001) 1.
13. Meister, G.E., Roberts, C.G., Berzofsky, J.A. and De Groot, A.S., Vaccine, 13 (1995) 581.
14. D'Amaro, J., Houbiers, J.G., Drijfhout, J.W., Brandt, R.M., Schipper, R., Bavinck, J.N., Melief, C.J. and Kast, W.M., Hum. Immunol., 43 (1995) 13.
15. Gulukota, K., Sidney, J., Sette, A. and DeLisi, C., J. Mol. Biol., 267 (1997) 1258.
16. Milik, M., Sauer, D., Brunmark, A.P., Yuan, L., Vitiello, A., Jackson, M.R., Peterson, P.A., Skolnick, J. and Glass, C.A., Nat. Biotechnol., 16 (1998) 753.
17. Adams, H.P. and Koziol, J.A., J. Immunol. Meth., 185 (1995) 181.
18. Parker, K.C., Bednarek, M.A. and Coligan, J.E., J. Immunol., 152 (1994) 163.
19. Stryhn, A., Pedersen, L.O., Romme, T., Holm, C.B., Holm, A. and Buus, S., Eur. J. Immunol., 26 (1996) 1911.
20. Fremont, D.H., Stura, E.A., Matsumura, M., Peterson, P.A. and Wilson, I.A., Proc. Natl. Acad. Sci. USA, 92 (1995) 2479.
21. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E., Nucleic Acids Res., 28 (2000) 235.
22. Batalia, M.A. and Collins, E.J., Biopolymers, 43 (1997) 281.
23. Madden, D.R., Gorga, J.C., Strominger, J.L. and Wiley, D.C., Cell, 70 (1992) 1035.
24. Guo, H., Madden, D., Silver, M., Jardetzky, T., Gorga, J., Strominger, J. and Wiley, D., Proc. Natl. Acad. Sci. USA, 90 (1993) 8053.
25. Ruppert, J., Sidney, J., Celis, E., Kubo, R., Grey, H. and Sette, A., Cell, 10 (1993) 929.
26. Madden, D.R., Garboczi, D.N. and Wiley, D.C., Cell, 75 (1993) 693.
27. Saper, M.A., Bjorkman, P.J. and Wiley, D.C., J. Mol. Biol., 75 (1991) 693.
28. Falk, K., Rotzschke, 0., Stevanovic, S., Jung, G. and Rammensee, H.-G., Nature, 351 (1991) 290.
29. Rognan, D., Scapozza, L., Folkers, G. and Daser, A., Biochemistry, 33 (1994) 11476.
30. Altuvia, Y., Sette, A., Sidney, J., Southwood, S. and Margalit, H., Hum. Immunol., 58 (1997) 1.
31. Schueler-Furman, O., Altuvia, Y., Sette, A. and Margalit, H., Protein Sci., 9 (2000) 1838.
32. Zeng, J., Treutlein, H.R. and Rudy, G.B., J. Comput. Aid. Mol. Des., 15 (2001) 573.
33. Doytchinova, I.A. and Flower, D.R., J. Med. Chem., 44 (2001) 3572.
34. Rognan, D., Lauemoller, S.L., Holm, A., Buus, S. and Tschinke, V., J. Med. Chem, 42 (1999) 4650.
35. Logean, A., Sette, A. and Rognan, D., Bioorg. Med. Chem. Lett., 11 (2001) 675.
36. TRIPOS Assoc., I., release 6.72.
37. Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham III, D.E., Ross, W.S., Simmerling, C.L., Darden, T.A., Merz, K.M., Stanton, R.V., Cheng, A.L., Vincent, J.J., Crowley, M., Tsui, V., Radmer, R.J., Duan, Y., Pitera, J., Massova, I., Seibel, G.L., Singh, V.C., Weiner, P.K. and Kollman, P.A. (1999) Amber 6.0, University of California, San Francisco.
38. Jardetzky, T.S., Lane, W.S., Robinson, R.A., Madden, D.R. and Wiley, D.C., Nature, 353 (1991) 326.
39. Lamas, J.R., Paradela, A., Roncal, F. and Lopez de Castro, J.A., Arthritis Rheumatism, 42 (1999) 1975.
40. Dedier, S., Reinelt, S., Reitinger, T., Folkers, G. and Rognan, D., J. Biol. Chem., 275 (2000) 27055.
41. Fairchild, P.J., J. Pept. Sci., 4 (1998) 182.
42. Guo, H.C., Jardetzky, T.S., Garrett, T.P.J., Lane, W.S., Strominger, J.L. and Wiley, D.C., Nature, 360 (1992) 364.
43. Garcia, K.C., Degano, M., Pease, L.R., Huang, M., Peterson, P.A., Teyton, L. and Wilson, I.A., Science, 279 (1998) 1666.
44. Garboczi, D.N., Ghosh, P., Utz, U., Fan, Q.R., Biddison, W.E. and Wiley, D.C., Nature, 384 (1996) 134.
45. Buus, S., Curr. Opin. Immunol., 11 (1999) 209.
46. Lamas, J.R., Brooks, J .M., Galocha, B., Rickinson, A.B. and Lopez de Castro, J.A., Int. Immunol., 10 (1998) 259.
47. Luu, T., Pham, S. and Deshpande, S., Int. J. Pept. Protein Res., 47 (1996) 91.
48. Dedier, S., Reinelt, S., Rion, S., Folkers, G. and Rognan, D., J. Immunol. Meth., 255 (2001) 57.
49. McKinney, D.M., Skvoretz, R., Qin, M., Ishioka, G. and Sette, A., J. Immunol. Meth., 237 (2000) 105.
50. Davenport, N.P., Smith, K.J., Barouch, D., Reid, S.W., Bodnar, W.M., Willis, A.M., Hunt, D.F. and A.V.S., H., J. Exp. Med., 185 (1997) 367.

51. Holzhutter, H.G., Frommel, C. and Kloetzel, P.M., J. Mol. Biol., 286 (1999) 1251.
52. Kesimir, C., Nussbaum, A.K., Schild, H., Detours, V., Brunak, S. Protein Eng., 15 (2002) 287; http://www.cbs.dtu.dk/services/NetChop/
53. Nussbaum, A.K., Kuttler, C., Hadeler, K.P., Rammensee, H.G. and Schild, H. Immunogenetics, 53 (2001) 87.
54. Brusic, V., van Endert, P., Zeleznikow, J., Daniel, S., Hammer, J. and Petrovsky, N., In Silico Biol., 1 (1999) 109.
55. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Jr., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A., J. Am. Chem. Soc., 117 (1995) 5179.
56. Nicholls, A., Sharp, K.A. and Honig, B., Proteins Struct. Funct. Genet., 11 (1991) 281.