Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Prediction Model**

Ranking of binding and nonbinding peptides to MHC class I molecules using inverse folding approach: Implications for vaccine design

Satarudra Prakash Singh[1, 2] and Bhartendu Nath Mishra[2, *]

[1]Amity Institute of Biotechnology, Amity University Uttar Pradesh, Gomti Nagar, Lucknow-226010, India; [2]Department of Biotechnology, Institute of Engineering & Technology, U.P. Technical University, Sitapur Road, Lucknow-226021, India;
Bhartendu Nath Mishra* - Email: profbnmishra@gmail.com; * Corresponding author

received September 21, 2008, accepted September 30, 2008; published October 24, 2008

Abstract:
T cell recognition of the peptide–MHC complex initiates a cascade of immunological events necessary for immune responses. Accurate T-cell epitope prediction is an important part of the vaccine designing. Development of predictive algorithms based on sequence profile requires a very large number of experimental binding peptide data to major histocompatibility complex (MHC) molecules. Here we used inverse folding approach to study the peptide specificity of MHC class I molecule with the aim of obtaining a better differentiation between binding and nonbinding peptides. Overlapping peptides, spanning the entire protein sequence, are threaded through the backbone coordinates of a known peptide fold in the MHC groove, and their interaction energies are evaluated using statistical pairwise contact potentials. We used the Miyazawa & Jernigan and Betancourt & Thirumalai tables for pairwise contact potentials, and two distance criteria (Nearest atom < 4.0 Å & C-beta< 7.0 Å) for ranking the peptides in an ascending order according to their energy values, and in most cases, known antigenic peptides are highly ranked. The predictions from threading improved when used multiple templates and average scoring scheme. In general, when structural information about a protein-peptide complex is available, the current application of the threading approach can be used to screen a large library of peptides for selection of the best binders to the target protein. The proposed scheme may significantly reduce the number of peptides to be tested in wet laboratory for epitope based vaccine design.

Keywords: epitope; MHC; threading; template; contact potential

Abbreviations: CTL - cytotoxic T lymphocytes; MHC - major histocompatibility complex; MJ- Miyazawa and Jernigan; BT- Betancourt and Thirumalai; PDB- Protein databank; MPID-MHC peptide interaction database

Background:
Development of epitope-based vaccines critically requires identification of regions in non-self and mutated proteins which are recognized by cytotoxic T lymphocyte (CTLs). The recognition of such regions by CTLs is a multistep processes where binding of peptides to MHC class I molecule is an important step and further transport of peptide–MHC complex to the antigen presenting cell surface **[1]**. Much of the information has accumulated regarding the specific binding of peptides to MHC class I molecules. A number of computational methods have been developed for the prediction of MHC binding peptide according to the data and computational approaches they apply i.e. sequence and structure based. Sequence based approaches includes motif, quantitative matrix and machine learning models have been successful applied in the discovery of novel T-cell epitopes involved in the cancer immunity **[2, 3]**. Although sequence based approaches are well established, but they require large sets of peptides that were tested experimentally and not feasible in situations where insufficient experimental binding data are available **[4, 5]**. Availability of crystallographically solved MHC-peptide complexes provides the opportunities for inverse folding (threading) approach which do not rely on previously binding data but aim to take account of the contributions of individual amino acids along the peptide that prompt them to fit into the groove of MHC allele using structural considerations **[6, 7, 8]**.

In this paper, an approach developed to address the inverse protein folding problem is applied to prediction of potential binding peptides to a specific MHC molecule and their interaction energies **[9]** are evaluated using statistical pairwise contact potentials, MJ **[10]** and BT **[11].**The number of conformations the peptide can adopt in the binding groove is limited and defined by the peptide-MHC structure that imposes physical constraints on the peptide **[12]**. The residues were considered to be in contact or not according to two different distance criteria **[6, 13]**. We also investigated whether using multiple template structures and taking the average improves the predictions or not. After these analysis, we found that using

ISSN 0973-2063 (online) 0973-2063 (print)
Bioinformation 3(2): 72-82 (2008)

BT potential with any two atoms are closer than 4 Å and taking multiple peptide conformations into consideration improves the threading procedure in discriminating between binding and nonbinding peptides. Hence, the compatibility of the peptide sequence with the space in the binding groove has an important role in molecular recognition which implies that the peptide conformation should be taken into consideration to improve the predictions of threading methods.

Methodology:

Template structures

The available data in the PDB are redundant and hence we created a non-redundant set from those entries with the best resolution for the related structural complexes having identical sequence information **[14]**. The non-redundant dataset consists of fifty four class I MHC-Peptide complexes (Table 1 in supplementary material). All the complexes chosen for the study were characterized using IMGT/3Dstructure-DB Structural Query tool **[15]** and MHC-Peptide Interaction Database (MPID) **[16]** including eleven 8-mer peptide-H2-Kb, seventeen 9-mer peptide-HLA-A*0201, twelve 9-mer peptide-H2-Db, four 10-mer peptide-HLA-A*0201 complexes. The MHC non-binding peptide data set for the selected alleles were retrieved from AntiJen database **[17]**, which covered a large range of IC_{50} value from 5000-440000 (Table 2 in supplementary material). The interface of peptide–MHC complexes is defined using the parameters, Interface Area and Gap Index **[18]**. Interface area for class I MHC-peptide complexes was defined as the change in their solvent accessible surface area (delta ASA) when going from a monomeric MHC molecule to a dimeric MHC-peptide complex state whereas, Gap index is used as means to evaluate the complementarity of interacting surfaces. The gap index is calculated using the formula, Gap Index = Gap Volume / delta ASA.

Threading with a contact potential matrix

In this method, binding affinity of a peptide is predicted by the total energy of interaction with contact residues. The contacts of the peptide in the available template co-crystal structure are determined according to two different criteria 1), ß-carbon atoms are closer than 7 Å **[6]**; and 2) any two atoms are closer than 4 Å **[13]**. Then, the amino-acid sequence of the query peptide is threaded onto the coordinates of the peptide in the template using MODPROPEP web server **[19]**. The contacts are assumed to be conserved, and the total interaction energy is obtained by summing the interaction energy values of peptide residues using a contact potential matrix. The contacting residues are determined for the conformation in the known structure, and therefore are only approximate for different sequences threaded. Energy values for amino acid-to-amino acid interactions are taken from the table of statistical pairwise contact potentials derived by MJ and BT **[10, 11]**. The experimental binding energies are correlated with binding affinity (IC_{50}) using the expression, $\Delta G\text{exp} = -RT \ln(IC_{50})$ where R is the gas constant and T the absolute temperature **[20]**. The predicted contact energies are given in dimensionless units of RT.

Results and discussion:

The peptide sequences in the test dataset (Table 1 and 2, see supplementary material) were threaded onto the crystal structures of the MHC class I peptide complexes. Different statistical potential matrices (MJ & BT) were used to obtain an estimate of the binding affinity of the threaded sequences, with the goal of ranking the binding and nonbinding sequences in the selected data set (see Methodology). We applied the method of Altuvia and colleagues **[21]** to score and rank the binding affinities of peptides to MHC class I molecules. Table 3 and 4 (under supplementary material)gives the ranking of peptides according to the binding affinities predicted by MJ and BT threading algorithm and using the 1VAC, 1LEG (H2-Kb/8); 1INQ, 1JPG (H2-Db/9); 1HHI, 1AO7 (HLA-A*0201/9); 1I4F, 2CLR (HLA-A*0201/10) complex structure as the template for two different distance criteria (Nearest atom < 4.0 Å & C-beta < 7.0 Å) to define the contacting residues. Although it is reasonable to use the same distance criterion as in the parameterization of the statistical contact potentials, we have applied both distance criteria to enable a direct comparison of the results.

Here, we found that the nearest atom < 4.0 Å distances criterion to determine the contacting residues gives a better prediction compared to C-beta < 7.0 Å distances (Table 3 and 4 in supplementary material). Surprisingly, although it still ranks high, the template structure's own peptide does *not* have the highest score, indicating that this force field may not have adequate precision. Overall, there is a tendency that the nonbinding peptides are ranked lower than the binding ones, but it is not possible to differentiate the binder and non binder using these rankings.

The pair wise potential is used to estimate the binding energies of peptide sequences threaded upon the different structural template. MJ pair-wise contact potential table puts much emphasis on hydrophobic interaction for the MHC alleles that contain various pockets of hydrophobic characters. Although most peptide are relatively buried within the binding groove of the MHC molecule, one can not assume that hydrophobic interaction are the mainly one that will tell binding from nonbinding peptides apart. So we have used the table of BT that has modified table of MJ by changing the reference state from solvent to a defined single solvent like molecule, the amino acid threonine and improved the ranking of template. However, in some cases (HLA-A*0201-10/1I4F), the template structure's own peptide has a very bad score, and is predicted to have a binding affinity even lower than nonbinding peptides. The results of threading are very much dependent on the template structure used, as a peptide ranks high if its binding scheme is similar to the template peptide. Hence, using multiple templates potentially should provide a better fit for the binding peptides. Therefore, this crude force field is not accurate enough to distinguish the subtle differences between the various peptide sequences.

ISSN 0973-2063 (online) 0973-2063 (print)
Bioinformation 3(2): 72-82 (2008)

For the other sequences, it is not possible to differentiate binding and nonbinding peptides based on energy using a single template; however, some binders have lower scores using one template and have high scores in the other. Using multiple templates provides more possible conformations accessible in the binding groove than the binding sequences can possibly assume. Therefore, taking the average of results from the two templates improves the results as seen in Table 5 (supplementary material). The non binders are ranked lower than the binder, but once again, the binding and nonbinding peptides are not separated significant. In another test to justify the use of the threading method, we evaluated their performances using the rank analysis of binding peptide in the source protein sequences derived from the overlapping peptides. The BT potentials generally rank the template structure's own peptide high among all possible 8, 9 and 10-mers in the source protein (Table 6 in supplementary material).

Conclusion:

Threading methodology employing two different statistical contact potentials (MJ and BT) and distance criteria (Nearest atom < 4.0 Å and C-beta < 7.0 Å) were applied to MHC class I molecules with a test set consisting of both its natural binding peptide and nonbinding peptide sequences The aim was to find which force field gives better predictions to rank and differentiate between the two groups in the test dataset, and hence determine which factors are important in the peptide recognition in MHC class I molecules. We found that using a BT force field, nearest atom < 4.0 Å distance criteria and the average of results from multiple template structures gives better predictions. Nevertheless, we could not obtain results that could separate the binders from the nonbinders in the test dataset even when we used multiple templates. This leads to the idea that shapes, rather than certain amino acids, are recognized by the MHC. Although the MHC also adapts to bind different sequences, the binding groove restricts the conformations accessible to the bound peptide.

The affinity of the peptide is thus affected by how well it can fit into the volume defined by the binding groove. This finding suggests that the "fitness" of a given peptide to the conformations accessible in the bound form is an important determinant of its binding affinity. This also indicates that the force field precisely defines the energy of the peptide when the exact conformation is available. Thus the inverse folding approach is advantageous for MHC alleles that lack binding data but have solved structure in complex with peptide, or alternatively, a structural model of the complex based on known structures. In this postgenomic era, the approach is potentially useful for screening a library of potential binding sequences to the newly discovered proteins to develop epitope based vaccines.

Acknowledgment:

We are grateful to Mr. Feroz Khan, Central Institute of Medicinal and Aromatic Plant, Lucknow and Dr. Mohammad Israil Ansari, Amity Institute of Biotechnology, Amity University, Lucknow campus for their critical reading of the manuscript and valuable suggestions. The authors are also thankful to U.P Technical University, Lucknow and Amity University Uttar Pradesh, Lucknow for their laboratory support during the research work.

References:

[01] A. Sette *et al., Tissue Antigens,* 59: 443 (2002) [PMID: 17090577]

[02] C. Parker *et al., J Immunol.*, 152: 163 (1994) [PMID: 8254189]

[03] Sette *et al., Curr Opin Investig Drugs*, 3: 132 (2002) [PMID: 12054064]

[04] V. Brusic *et al., Methods*, 34: 436 (2004) [PMID: 15542369]

[05] P. Kangueane and M. K. Sakharkar, *Bioinformation*, 1: 21 (2005) [PMID: 17597847]

[06] Y. Altuvia *et al.*, *J. Mol. Biol.,* 249: 244 (1995) [PMID: 7540211]

[07] O. Schueler-Furman *et al., Fold Des*., 3: 549 (1998) [PMID: 9889166]

[08] O. Schueler-Furman *et al., Protein Sci.,* 9: 1838 (2000) [PMID: 11045629].

[09] U. Sezerman, *et al., Protein Sci.,* 5: 1272 (1996) [PMID: 8819160]

[10] S. Miyazawa and R. L. Jernigan, *J. Mol. Biol.,* 256: 623 (1996) [PMID: 8604144]

[11] M. R. Betancourt and D. Thirumalai, *Prot. Sci*., 8: 361 (1999) [PMID: 10048329]

[12] N. Kurt, *et al*., *Biophys. J*., 85**:** 853 (2003) [PMID: 12885633]

[13] D. R. Madden *et al*., *Cell,* 75: 693(1993) [PMID: 7694806]

[14] H. M Berman *et al.*, *Nucleic Acids Res.,* 28: 235 (2000) [PMID: 10592235]

[15] Q. Kaas, *et al., Nucleic Acids Res.*, 32: D208 (2004) [PMID: 14681396]

[16] J. C Tong *et al. Appl Bioinformatics*, 5: 111 (2006) [PMID: 16722775]

[17] C. P Toseland *et al.*, *Immunome Research,* 6: 1 *(*2005) [PMID: 16305757]

[18] P. E. Adrian *et al.*, *BMC Structural Biology,* 2: 2 (2002) [PMID: 12010576]

[19] N. Kumar and D. Mohanty, *Nucleic Acids Res*., 35: W549 (2007) [PMID: 17478500]

[20] P. U. Guan, *et al.*, *Nucleic Acids Res*., 31: 3621 (2003) [PMID: 16539539]

[21] Y. Altuvia. and H. Margalit, *Methods*, 34: 454 (2004) [PMID: 15542371]

Edited by P. Kangueane

Citation: Singh *et al.*, Bioinformation 3(2): 72-82 (2008)

Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Prediction Model**

Supplementary material

MHC Class I allele/peptide length	PDB ID	Peptide sequence	Peptide source/ Uniport accession no.	Resolution (A^0)	Gap Index (A^0)	Interface area (A^{02})	Total Hydrogen bonds in pMHC
H2-Kb/8	1KJ3	KVITFIDL	Naturally processed	2.3	0.6	905.3	14
	1FO0	INFDFNTI	pBM1 peptide [Mouse]	2.50	0.9	902.8	18
	1VAC	SIINFEKL	P01012	2.5	0.8	876.9	14
	1S7S	ALYNFATM	P07399	1.99	0.8	877.5	10
	1S7T	AVFNFATM	Lcmv-Peptide	2.30	0.8	852.3	11
	1G6R	SIYRYYGL	Syir protein	2.80	1.2	865.2	13
	1N59	AVYNFATM	P07399	2.95	0.9	864.7	9
	1LK2	GNYSFYAL	Synthetic	1.35	1.1	776.2	15
	1OSZ	RGYLYQGL	P11212	2.1	0.8	909.5	18
	1BQH	RGYVYQGL	P11212	2.80	0.8	882.6	12
	1LEG	EQYKFYSV	Q62425	1.75	0.9	880.6	15
H2-Db/9	1JUF	SSVIGVWYL	AAB81863	2.0	0.7	865.1	16
	1INQ	SSVVGVWYL	AAB81863	2.20	0.6	870.2	16
	1S7W	KALYNFATM	P07399	2.40	0.5	993.0	18
	1S7X	KAVFNFATM	P07399	2.41	0.5	950.0	18
	1FFO	AAVYNFATM	P07399	2.65	0.6	878.4	17
	1BZ9	FAPGVFPYM	P04857	2.8	0.8	1164.2	10
	1FFN	KAVYNFATM	P07399	2.70	0.7	946.8	18
	1FFP	SAVYNFATM	SRC2066(IEDB)	2.60	0.4	905.2	21
	1JPG	FQPQNGQFI	M20869	2.20	0.7	947.8	19
	1FG2	KAVYNFATC	P07399	2.75	0.6	928.4	17
	1CE6	FAPGNYPAL	SV nucleoprotein	2.90	0.9	867.7	15
	1QLF	FAPSNYPAL	P04857	2.65	0.7	879.5	13
HLA-A*0201/9	1QRN	LLFGYAVYV	Tax peptide P6A	2.8	1.1	871.5	10
	1AO7	LLFGYPVYV	Q82235	2.6	1.2	883.3	10
	1QSF	LLFGYPVAV	Tax peptide	2.80	1.2	828.6	10
	1HHI	GILGFVFTL	Q66PA1	2.5	0.5	842.0	9
	1QEW	FLWGPRALV	P43357	2.20	1	843.2	12
	1QSE	LLFGYPRYV	Tax peptide	2.80	1.3	873.0	11
	1B0G	ALWGFFPVL	Self peptide P1049]	2.5	0.5	860.2	12
	1I7T	ALWGVFPVL	Self peptide P1049	2.8	0.7	847.9	9
	1I7U	ALWGFVPVL	Self peptide [Human]	1.8	0.7	845.2	11
	1QR1	IISAVVGIL	P04626	2.4	0.9	827.1	9
	1EEY	ILSALVGIV	P04626	2.25	1.1	787.1	11
	1JHT	ALGIGILTV	NP_005502	2.15	0.8	781.0	12
	1I7R	FAPGFFPYL	P04857	2.20	0.9	902.5	11
	1I1F	FLKEPVHGV	HIV- reverse transcriptase	2.80	0.9	850.9	11
	1AKJ	ILKEPVHGV	HIV-reverse transcriptase	2.65	0.9	857.2	13
	1I1Y	YLKEPVHGV	HIV-reverse transcriptase	2.20	0.9	877.9	13
	1HHG	TLTSCNTSV	HIV-1 gp 120	2.6	1.4	765.8	12
HLA-A*0201/10	2CLR	MLLSVPLLLG	Calreticulin peptide	2.0	1.0	896.5	10
	1HHH	FLPSDFFPSV	HBV Nucleocapsid protein	3.0	0.6	918.4	11
	1JF1	ELAGIGILTV	melan-A [Homo sapiens]	1.85	0.6	870.4	11
	1I4F	GVYDGREHTV	P43358	1.40	1.1	820.1	15

Table1: MHC-peptide complexes used in the study.

lSSN 0973-2063 (online) 0973-2063 (print)
Bioinformation 3(2): 72-82 (2008)

MHC Class I allele/peptide length	Peptide ID	Peptide sequence	Category	Swiss Prot Accession no.	IC_{50} (nM)
H2-Kb/8	1	NTVVFDAL	SYNTHETIC	P23700	155000
	2	DDEEYVIL	SYNTHETIC	P31682	124000
	3	QPQNYLRL	SYNTHETIC	O86164	51667
	4	ANEGYDAL	SYNTHETIC	P31681	11923
	5	IIFLFILL	VIRAL	P03140	7500
	6	MWYWGPSL	VIRAL	P03140	7500
	7	LMSGFRQM	SYNTHETIC	Q9Z7H7	6889
	8	CLIFLLVL	VIRAL	P03140	6000
	9	FIIFLFIL	VIRAL	P03140	5000
H2-Db/9	1	AEDTNVSLI	SYNTHETIC	P23700	440000
	2	GFKSNFNKI	SYNTHETIC	Q9Z7H7	440000
	3	QLPPNSLLI	BACTERIAL	Q9Z7P3	293333
	4	VENPGGYCL	VIRAL	P09991	33500
	5	TAGANPMDL	SYNTHETIC	P31681	22000
	6	TGKLNLENL	SYNTHETIC	Q9Z6X8	17600
	7	SGVENPGGY	VIRAL	P09991	13155
	8	KAVYNFATC	VIRAL	P09991	5429
	9	LLVFNYPGI	SYNTHETIC	Q9Z6X9	5167
HLA-A*0201/9	1	VVHFFKNIV	SELF PEPTIDE	P02686	50000
	2	KIFGSLAFL	SELF PEPTIDE	P04626	27000
	3	TLPRARRRV	CANCER	Q01726	25000
	4	SLLMWITQC	SELF PEPTIDE	P78358	21070
	5	FLFGSLAFL	SELF PEPTIDE	P04626	19000
	6	FLYAALLLA	SELF PEPTIDE	P06905	17177
	7	NGMLIMCNA	CANCER	Q01718	16667
	8	SLYITVAVL	VIRAL	P05889	16667
	9	RLCVQSTHV	VIRAL	P03129	16666
	10	FLFESLAFL	SELF PEPTIDE	P04626	15000
	11	HLSLRGLPV	VIRAL	P20977	12500
	12	ALVARAAVL	CANCER	Q01726	11111
	13	SLCFLGAIA	CANCER	Q01726	10000
	14	HLEGKVILV	VIRAL	P03368	8333
	15	VALVGLFVL	SELF PEPTIDE	P40126	8333
	16	RVMAPRALL	CANCER	P43355	7667
	17	VCMTVDSLV	SELF PEPTIDE	P40126	7143
	18	ILLGIFFLC	CANCER	Q01726	5556
	19	KLPQLCTEL	VIRAL	P03126	5556

lSSN 0973-2063 (online) 0973-2063 (print)
Bioinformation 3(2): 72-82 (2008)

	20	ALGLVCVQM	CANCER	P43355	5000
	21	FLHLTLIVL	CANCER	Q01726	5000
	22	HLESLFTAV	VIRAL	P03156	5000
	23	LLGCAANWI	VIRAL	P12900	5000
HLA-A*0201/10	1	AAGIGILTVI	CANCER	Q16655	5555
	2	ELCCQHLWQI	ALLERGEN	P04721	5555
	3	FLPRHRDTGI	SELF PEPTIDE	P02686	5000

Table 2: MHC non binding peptide dataset used in the study.

MHC Class I allele/peptide length	Nearest atom < 4.0 Å				C-beta< 7.0 Å			
	PDB ID/Peptide ID	MJ	PDB ID/Peptide ID	BT	PDB ID/Peptide ID	MJ	PDB ID/Peptide ID	BT
H2-Kb/8	5	-196.3	6	-6.4	5	-90.5	3	-0.2
	9	-195.8	7	-6.3	9	-89.3	1FO0	-1.2
	8	-187.2	1FO0	-5.5	8	-84.6	6	-1.2
	1KJ3	-157.3	1S7S	-4.8	1FO0	-68.8	1OSZ	0.0
	1FO0	-157.0	1LEG	-4.8	6	-67.8	1BQH	0.0
	7	-154.9	5	-4.7	7	-66.2	7	0.2
	1VAC*	-154.8	9	-4.4	1KJ3	-65.7	2	0.3
	1S7S	-154.3	1KJ3	-4.3	1S7S	-65.2	1LK2	0.5
	1S7T	-153.6	1G6R	-4.2	1S7T	-65.1	1VAC*	0.6
	1G6R	-150.2	1N59	-4.2	1VAC*	-64.7	1S7S	0.6
	6	-149.2	1LK2	-4.2	1	-64.4	1S7T	0.6
	1N59	-147.5	1S7T	-4.0	1G6R	-64.3	4	0.6
	1	-146.0	1VAC*	-3.7	2	-63.0	1KJ3	0.7
	1LK2	-144.8	1OSZ	-3.7	1OSZ	-62.9	1N59	0.8
	1OSZ	-137.3	1BQH	-3.6	1N59	-62.4	1G6R	0.9
	2	-136.7	3	-3.4	1LK2	-62.2	1LEG	0.9
	1BQH	-135.4	8	-3.4	1BQH	-61.1	5	0.9
	1LEG	-133.4	1	-3.3	3	-57.2	1	1.1
	3	-131.7	4	-3.0	4	-56.7	8	1.3
	4	-127.8	2	-2.8	1LEG	-54.5	9	1.3
H2-Db/9	9	-166.6	1BZ9	-9.3	9	-101.2	1BZ9	-9.3
	3	-158.1	1CE6	-7.1	3	-100.4	1S7W	-7.0
	1BZ9	-155.6	1QLF	-7.0	1INQ*	-97.0	1FFN	-6.7
	1JUF	-153.8	1S7W	-7.0	1S7W	-91.1	1FG2	-2.6
	1JPG	-153.0	9	-7.0	1S7X	-90.1	8	-2.6
	1INQ*	-151.5	4	-6.9	1FFO	-87.9	4	-2.6
	1CE6	-148.6	1FFN	-6.7	1JUF	-87.8	1INQ*	-2.3
	4	-148.6	1JPG	-6.5	1FFN	-87.1	1JUF	-2.1
	1QLF	-148.1	1S7X	-6.2	1BZ9	-87.1	1CE6	-2.1
	1	-140.2	3	-6.1	1FFP	-87.0	1JPG	-2.0
	1FFO	-138.4	1FG2	-5.9	1FG2	-85.5	1S7X	-1.8
	1S7X	-137.3	8	-5.9	8	-85.5	3	-1.8
	1S7W	-135.6	1FFO	-5.4	4	-85.2	1QLF	-1.7
	1FFP	-134.6	1INQ*	-5.2	1	-84.0	1FFO	-1.4
	5	-134.5	1JUF	-5.2	1CE6	-83.4	1FFP	-1.2
	6	-134.5	1FFP	-4.1	1JPG	-83.4	1	-1.2
	1FFN	-132.6	1	-4.0	1QLF	-82.7	9	-1.0
	1FG2	-128.6	5	-3.9	6	-81.4	2	-0.8
	8	-128.6	7	-3.3	2	-80.0	6	-0.7
	2	-128.1	2	-2.5	5	-79.8	7	-0.5
	7	-117.6	6	-2.5	7	-73.6	5	0.7
HLA-	18	-209.2	1AO7	-10.6	18	-110.7	16	-7.4

lSSN 0973-2063 (online) 0973-2063 (print)
Bioinformation 3(2): 72-82 (2008)

A*0201/9	10	-201.6	1QRN	-10.3	21	-108.0	14	-4.4
	5	-200.9	6	-9.9	10	-107.8	6	-3.7
	21	-200.2	1QEW	-9.8	5	-107.1	10	-3.6
	1QRN	-195.7	18	-9.7	6	-104.4	1QRN	-3.5
	1AO7	-195.1	1QSF	-9.6	2	-100.0	1AO7	-3.5
	6	-193.9	10	-9.6	15	-99.0	18	-3.5
	15	-192.0	1B0G	-9.4	1B0G	-98.8	1QSF	-3.3
	1QSF	-190.6	1I7T	-9.2	14	-98.7	2	-3.0
	1HHI*	-190.1	1I7U	-9.1	1QRN	-97.9	21	-3.0
	1QEW	-184.1	1QSE	-8.9	1HHI*	-97.9	1QEW	-2.8
	1QSE	-182.5	1I7R	-8.7	1I7T	-97.8	1B0G	-2.8
	1B0G	-181.1	5	-8.6	1AO7	-96.7	22	-2.7
	2	-179.7	21	-8.5	1I7U	-96.1	1I7T	-2.6
	1I7T	-179.3	14	-8.2	1QR1	-95.9	5	-2.6
	1I7U	-179.1	2	-7.3	8	-95.8	11	-2.6
	1QR1	-177.4	19	-7.3	1EEY	-95.2	1I7U	-2.5
	14	-176.3	15	-7.1	1QEW	-94.5	1I1Y	-2.4
	1EEY	-174.5	1HHI*	-6.9	1QSF	-93.7	1HHI*	-2.3
	12	-173.1	22	-6.7	1JHT	-93.4	8	-2.3
	8	-172.5	1I1F	-6.3	12	-91.6	1JHT	-2.2
	23	-171.3	20	-6.1	1QSE	-90.2	1I7R	-2.1
	16	-170.2	11	-5.8	16	-89.4	1I1F	-2.1
	1JHT	-169.8	1I1Y	-5.5	23	-89.3	15	-2.1
	17	-168.0	12	-5.5	20	-88.9	20	-2.1
	1I7R	-167.9	1AKJ	-5.4	22	-88.5	1AKJ	-2.0
	11	-165.9	8	-5.3	1I7R	-87.5	19	-1.9
	20	-165.7	17	-5.1	4	-87.2	1QSE	-1.8
	1	-163.5	23	-5.0	13	-87.2	12	-1.8
	4	-162.3	9	-4.7	11	-87.1	1EEY	-1.1
	1I1F	-161.3	1JHT	-4.5	1I1F	-85.6	9	-1.1
	13	-159.9	1EEY	-4.3	1AKJ	-85.0	13	-1.0
	22	-159.0	1QR1	-4.1	17	-84.8	7	-0.9
	1AKJ	-158.8	1	-4.0	1	-83.9	1QR1	-0.7
	19	-155.6	3	-3.8	19	-83.5	4	-0.6
	9	-154.5	7	-3.4	1I1Y	-83.4	3	-0.5
	1I1Y	-152.9	13	-3.4	9	-81.1	23	-0.5
	1HHG	-146.1	4	-3.3	1HHG	-76.1	17	-0.1
	7	-143.0	16	-2.5	3	-74.4	1HHG	-0.0
	3	-142.7	1HHG	-2.4	7	-74.3	1	0.3
HLA-A*0201/10	2CLR	-176.8	3	-7.8	2CLR	-121.8	2CLR	-4.0
	1HHH	-158.6	2CLR	-7.3	1JF1	-108.0	1HHH	-2.6
	2	-157.0	1HHH	-6.7	2	-107.9	1JF1	-2.4
	3	-155.7	2	-5.6	1HHH	-102.2	2	-1.9
	1JF1	-148.9	1I4F*	-4.6	1	-101.7	3	-1.6
	1	-145.9	1JF1	-4.4	3	-95.8	1I4F*	-1.5
	1I4F*	-135.0	1	-1.9	1I4F*	-87.9	1	-0.0

Table 3 : Ranking of MHC binding and non peptides according to their predicted binding affinity by threading using a scoring matrix (MJ & BT) and two distance criteria (Nearest atom < 4.0 Å & C-beta< 7.0 Å). * Structure used as template.

MHC Class I allele/peptide length	Nearest atom < 4.0 A^0				C-beta< 7.0 A^0			
	PDB ID/Peptide	MJ	PDB ID/Peptide	BT	PDB ID/Peptide	MJ	PDB ID/Peptide	BT

lSSN 0973-2063 (online) 0973-2063 (print)
Bioinformation 3(2): 72-82 (2008)

	ID		ID		ID		ID	
H2-Kb/8	9	-185.0	1LEG*	-5.5	5	-105.3	6	-2.1
	5	-184.0	6	-6.8	9	-104.4	1FO0	-1.9
	8	-175.1	1FO0	-4.8	8	-100.1	1VAC	-0.6
	1FO0	-151.6	2	-4.0	1S7S	-81.0	1S7S	-0.6
	7	-145.6	1S7S	-3.6	1VAC	-79.9	1S7T	-0.5
	1VAC	-145.5	1LK2	-3.5	1KJ3	-79.3	3	-0.5
	1KJ3	-144.9	1OSZ	-3.5	1FO0	-79.1	1N59	-0.3
	1S7T	-144.8	3	-3.5	1S7T	-79.0	2	-0.2
	1S7S	-144.0	1BQH	-3.4	7	-79.0	4	-0.2
	6	-142.2	7	-3.3	6	-78.8	1KJ3	-0.1
	1N59	-137.3	1N59	-3.0	1G6R	-77.2	1LK2	-0.1
	1G6R	-136.3	9	-2.5	1N59	-76.3	1	-0.1
	1	-133.9	1G6R	-2.4	1	-75.7	5	-0.0
	1LK2	-131.6	4	-2.4	1LK2	-72.4	1OSZ	0.1
	1OSZ	-127.2	1KJ3	-2.2	2	-71.8	1BQH	0.1
	1BQH	-124.5	1VAC	-2.2	1OSZ	-71.5	1LEG*	0.2
	1LEG*	-123.9	5	-2.1	1BQH	-69.7	7	0.2
	3	-117.8	1S7T	-1.7	3	-65.9	9	0.2
	2	-116.9	1	-1.7	4	-65.9	8	0.3
	4	-112.6	8	-0.8	1LEG*	-64.9	1G6R	0.5
H2-Db/9	9	-190.3	1BZ9	-8.7	3	-112.8	1S7W	-3.5
	3	-177.7	1S7W	-7.6	9	-109.9	1FFN	-3.2
	1BZ9	-177.5	1FFN	-7.2	1JUF	-102.6	1BZ9	-3.1
	1JUF	-167.9	1S7X	-6.5	1INQ	-101.5	1FG2	-2.9
	1INQ	-166.2	8	-6.0	1BZ9	-96.6	8	-2.9
	1JPG*	-165.7	4	-5.4	1FFN	-96.6	4	-2.6
	1CE6	-165.3	3	-5.2	1S7W	-95.3	1S7X	-2.5
	1QLF	-165.0	9	-5.2	1S7X	-94.2	3	-2.4
	4	-160.8	1FFO	-4.1	1FFO	-93.1	1CE6	-2.3
	2	-156.3	1JUF	-4.0	2	-93.1	1QLF	-2.3
	1S7W	-156.0	1INQ	-4.0	1JPG*	-93.0	1JUF	-2.2
	1FFO	-156.0	1FFP	-3.7	4	-92.5	1INQ	-2.2
	1S7X	-155.2	1	-3.6	1FFP	-92.3	1JPG*	-2.2
	1	-153.4	6	-3.3	1CE6	-91.6	1FFO	-2.0
	1FFP	-153.2	2	-3.2	1QLF	-91.3	1FFP	-1.9
	1FFN	-152.0	7	-3.0	1	-90.6	9	-1.9
	6	-151.4	1FG2	-2.9	1FG2	-90.0	2	-1.6
	5	-149.1	5	-2.7	8	-90.0	1	-1.3
	1FG2	-147.3	1CE6	-2.3	5	-88.1	6	-1.1
	8	-147.3	1QLF	-2.3	6	-87.6	7	-1.0
	7	-133.9	1JPG*	0.1	7	-79.1	5	-0.0
HLA-A*0201/9	5	-191.5	1QRN	-10.1	18	-113.5	10	-4.7
	10	-191.5	1AO7*	-10.1	21	-110.4	1QRN	-4.6
	18	-188.1	1QEW	-9.8	10	-108.6	1AO7*	-4.6
	1QRN	-181.4	5	-9.4	5	-107.9	18	-4.5
	1AO7*	-181.0	10	-9.4	6	-	1QSF	-4.3

lSSN 0973-2063 (online) 0973-2063 (print)
Bioinformation 3(2): 72-82 (2008)

						106.5		
	21	-179.3	6	-9.2	15	-	6	-4.0
						104.0		
	1QEW	-177.2	1QSE	-9.0	1HHI	-	21	-3.9
						101.2		
	6	-176.3	2	-8.8	2	-	5	-3.7
						100.0		
	1QSF	-175.5	1QSF	-8.7	1QRN	-99.9	14	-3.7
	1QSE	-175.2	16	-8.1	14	-99.4	1HHI	-3.6
	15	-166.4	1B0G	-8.0	1B0G	-99.0	1B0G	-3.6
	1QR1	-166.2	1I7U	-8.0	1AO7*	-98.6	1QEW	-3.5
	2	-164.7	18	-8.0	1I7T	-98.0	1I7T	-3.4
	1B0G	-164.2	21	-7.9	1QR1	-96.3	1I1F	-3.4
	23	-164.0	1I7T	-7.8	1I7U	-96.2	1I7U	-3.3
	1EEY	-163.7	14	-7.5	1JHT	-95.9	15	-3.3
	1I7U	-163.4	19	-7.1	1QSF	-95.6	1AKJ	-3.2
	1I7T	-163.2	1I7R	-7.0	8	-95.6	1I7R	-3.1
	1HHI	-162.5	12	-6.3	1EEY	-95.2	1I1Y	-3.1
	12	-160.8	8	-6.2	1QEW	-94.9	1JHT	-2.9
	17	-157.6	23	-6.1	12	-92.3	2	-2.9
	16	-157.4	9	-6.0	1QSE	-90.7	1QSE	-2.8
	14	-157.3	1I1F	-5.9	20	-90.5	20	-2.8
	8	-156.5	1I1Y	-5.8	16	-90.2	12	-2.7
	1I7R	-152.1	1HHI	-5.7	1I7R	-89.9	16	-2.7
	1	-151.6	11	-5.7	23	-89.2	11	-2.6
	1I1F	-150.2	15	-5.6	11	-88.9	8	-2.4
	1AKJ	-147.3	22	-5.3	13	-87.7	22	-2.0
	4	-145.5	1AKJ	-5.0	4	-87.6	7	-1.9
	13	-145.4	20	-4.9	22	-87.4	13	-1.7
	1JHT	-144.4	3	-4.8	1I1F	-86.7	19	-1.7
	11	-142.3	1EEY	-4.4	17	-86.4	1EEY	-1.6
	20	-141.8	1JHT	-4.4	1AKJ	-86.1	4	-1.4
	9	-141.7	13	-4.4	1	-85.1	1QR1	-1.3
	22	-141.1	17	-4.4	1I1Y	-83.9	3	-1.1
	1I1Y	-139.9	4	-4.1	19	-82.8	17	-1.1
	19	-136.5	1QR1	-4.0	9	-80.5	23	-1.1
	1HHG	-132.5	1	-3.9	7	-78.1	9	-1.0
	3	-131.6	1HHG	-2.6	1HHG	-75.7	1	-0.6
	7	-117.7	7	-2.4	3	-74.4	1HHG	-0.1
HLA-A*0201/10	2CLR*	-191.1	2CLR *	-9.7	1	-83.4	2CLR*	-4.3
	1HHH	-159.1	1HHH	-8.0	1HHH	-84.0	1JF1	-2.7
	2	-151.9	2	-7.0	1I4F	-75.6	2	-1.7
	3	-150.7	3	-5.4	1JF1	-95.2	1HHH	-1.5
	1JF1	-150.1	1JF1	-4.9	2	-89.1	1I4F	-1.0
	1	-138.9	1I4F	-4.2	2CLR*	-	1	-0.6
						106.6		
	1I4F	-129.4	1	-3.0	3	-83.2	3	-0.5

Table 4: Ranking of MHC binding and non binding peptides according to their predicted binding affinity by threading using a scoring matrix (MJ & BT) and two distance criteria (Nearest atom < 4.0 A^0 & C-beta< 7.0 A^0). * Structure used as template.

MHC Class I allele/peptide length	Nearest atom < 4.0 A^0				C-beta< 7.0 A^0			
	PDB ID/Peptide ID	MJ	PDB ID/Peptide ID	BT	PDB ID/Peptide ID	MJ	PDB ID/Peptide ID	BT

lSSN 0973-2063 (online) 0973-2063 (print)
Bioinformation 3(2): 72-82 (2008)

H2-Kb/8	5	-196.3	6	-6.60	5	-97.90	6	-1.65
	9	-195.8	1LEG*	-5.15	9	-96.85	1FO0	-1.55
	8	-187.2	1FO0	-5.15	8	-92.35	3	-0.35
	1KJ3	-157.3	7	-4.80	1FO0	-73.95	1VAC*	0.0
	1FO0	-157.0	1S7S	-4.20	6	-73.30	1S7S	0.0
	7	-154.9	1LK2	-3.85	1S7S	-73.10	1S7T	0.05
	1VAC*	-154.8	1OSZ	-3.60	7	-72.60	2	0.05
	1S7S	-154.3	1N59	-3.60	1KJ3	-72.50	1OSZ	0.05
	1S7T	-153.6	1BQH	-3.50	1VAC*	-72.30	1BQH	0.05
	1G6R	-150.2	3	-3.45	1S7T	-72.05	4	0.20
	6	-149.2	9	-3.45	1G6R	-70.75	1LK2	0.20
	1N59	-147.5	2	-3.40	1	-70.05	7	0.20
	1	-146.0	5	-3.40	1N59	-69.35	1N59	0.25
	1LK2	-144.8	1G6R	-3.30	2	-67.40	1KJ3	0.30
	1OSZ	-137.3	1KJ3	-3.25	1LK2	-67.30	5	0.45
	2	-136.7	1VAC*	-2.95	1OSZ	-67.20	1	0.50
	1BQH	-135.4	1S7T	-2.85	1BQH	-65.40	1LEG*	0.55
	1LEG*	-133.4	4	-2.70	3	-61.55	1G6R	0.70
	3	-131.7	1	-2.50	4	-61.30	9	0.75
	4	-127.8	8	-2.10	1LEG*	-59.70	8	0.80
H2-Db/9	9	-178.45	1BZ9	-9	3	-106.6	1BZ9	-6.2
	3	-167.9	1S7W	-7.3	9	-105.55	1S7W	-5.25
	1BZ9	-166.55	1FFN	-6.95	1INQ*	-99.25	1FFN	-4.95
	1JUF	-160.85	1S7X	-6.35	1JUF	-95.2	1FG2	-2.75
	1JPG*	-159.35	4	-6.15	1S7W	-93.2	8	-2.75
	1INQ*	-158.85	9	-6.1	1S7X	-92.15	4	-2.6
	1CE6	-156.95	8	-5.95	1BZ9	-91.85	1INQ*	-2.25
	1QLF	-156.55	3	-5.65	1FFN	-91.85	1CE6	-2.2
	4	-154.7	1FFO	-4.75	1FFO	-90.5	1S7X	-2.15
	1FFO	-147.2	1CE6	-4.7	1FFP	-89.65	1JUF	-2.15
	1	-146.8	1QLF	-4.65	4	-88.85	3	-2.1
	1S7X	-146.25	1JUF	-4.6	1JPG*	-88.2	1JPG*	-2.1
	1S7W	-145.8	1INQ*	-4.6	1FG2	-87.75	1QLF	-2
	1FFP	-143.9	1FG2	-4.4	8	-87.75	1FFO	-1.7
	6	-142.95	1FFP	-3.9	1CE6	-87.5	1FFP	-1.55
	1FFN	-142.3	1	-3.8	1	-87.3	9	-1.45
	2	-142.2	5	-3.3	1QLF	-87	1	-1.25
	5	-141.8	1JPG*	-3.2	2	-86.55	2	-1.2
	1FG2	-137.95	7	-3.15	6	-84.5	6	-0.9
	8	-137.95	6	-2.9	5	-83.95	7	-0.75
	7	-125.75	2	-2.85	7	-76.35	5	0.35
HLA-A*0201/9	18	-198.65	1AO7*	-10.35	18	-112.1	16	-5.05
	10	-196.55	1QRN	-10.2	21	-109.2	10	-4.15
	5	-196.2	1QEW	-9.8	10	-108.2	1QRN	-4.05
	21	-189.75	6	-9.55	5	-107.5	1AO7*	-4.05
	1QRN	-188.55	10	-9.5	6	-105.45	14	-4.05
	1AO7*	-188.05	1QSF	-9.15	15	-101.5	18	-4
	6	-185.1	5	-9	2	-100	6	-3.85
	1QSF	-183.05	1QSE	-8.95	1HHI*	-99.55	1QSF	-3.8
	1QEW	-180.65	18	-8.85	14	-99.05	21	-3.45
	15	-179.2	1B0G	-8.7	1QRN	-98.9	1B0G	-3.2
	1QSE	-178.85	1I7U	-8.55	1B0G	-98.9	5	-3.15
	1HHI*	-176.3	1I7T	-8.5	1I7T	-97.9	1QEW	-3.15
	1B0G	-172.65	21	-8.2	1AO7*	-97.65	1I7T	-3
	2	-172.2	2	-8.05	1I7U	-96.15	1HHI*	-2.95
	1QR1	-171.8	14	-7.85	1QR1	-96.1	2	-2.95
	1I7U	-171.25	1I7R	-7.85	8	-95.7	1I7U	-2.9
	1I7T	-171.25	19	-7.2	1EEY	-95.2	1I1F	-2.75

lSSN 0973-2063 (online) 0973-2063 (print)
Bioinformation 3(2): 72-82 (2008)

	1EEY	-169.1	15	-6.35	1QEW	-94.7	1I1Y	-2.75
	23	-167.65	1HHI*	-6.3	1JHT	-94.65	15	-2.7
	12	-166.95	1I1F	-6.1	1QSF	-94.65	1AKJ	-2.6
	14	-166.8	22	-6	12	-91.95	1I7R	-2.6
	8	-164.5	12	-5.9	1QSE	-90.45	11	-2.6
	16	-163.8	8	-5.75	16	-89.8	1JHT	-2.55
	17	-162.8	11	-5.75	20	-89.7	20	-2.45
	1I7R	-160	1I1Y	-5.65	23	-89.25	8	-2.35
	1	-157.55	23	-5.55	1I7R	-88.7	22	-2.35
	1JHT	-157.1	20	-5.5	11	-88	1QSE	-2.3
	1I1F	-155.75	9	-5.35	22	-87.95	12	-2.25
	11	-154.1	16	-5.3	13	-87.45	19	-1.8
	4	-153.9	1AKJ	-5.2	4	-87.4	7	-1.4
	20	-153.75	17	-4.75	1I1F	-86.15	13	-1.35
	1AKJ	-153.05	1JHT	-4.45	17	-85.6	1EEY	-1.35
	13	-152.65	1EEY	-4.35	1AKJ	-85.55	9	-1.05
	22	-150.05	3	-4.3	1	-84.5	4	-1
	9	-148.1	1QR1	-4.05	1I1Y	-83.65	1QR1	-1
	1I1Y	-146.4	1	-3.95	19	-83.15	3	-0.8
	19	-146.05	13	-3.9	9	-80.8	23	-0.8
	1HHG	-139.3	4	-3.7	7	-76.2	17	-0.6
	3	-137.15	7	-2.9	1HHG	-75.9	1	-0.15
	7	-130.35	1HHG	-2.5	3	-74.4	1HHG	-0.05
HLA-A*0201/10	2CLR*	-183.95	2CLR*	-8.5	2CLR*	-121.8	2CLR*	-4.15
	1HHH	-158.85	1HHH	-7.35	1JF1	-108	1JF1	-2.55
	2	-154.45	3	-6.6	2	-107.9	1HHH	-2.05
	3	-153.2	2	-6.3	1HHH	-102.2	2	-1.8
	1JF1	-149.5	1JF1	-4.65	1	-101.7	1I4F*	-1.25
	1	-142.4	1I4F*	-4.4	3	-95.8	3	-1.05
	1I4F*	-132.2	1	-2.45	1I4F*	-87.9	1	-0.3

Table 5 : Ranking of MHC binding and non binding peptides according to their average predicted binding affinity by threading using a scoring matrix (MJ & BT) and two distance criteria (Nearest atom < 4.0 A^0 & C-beta< 7.0 A^0). * Structure used as template.

MHC Class I allele/peptide length	Nearest atom < 4.0 A^0				C-beta< 7.0 A^0(Rank)			
	PDB ID	MJ	PDB ID	BT	PDB ID	MJ	PDB ID	BT
H2-Kb/8	1VAC	13	1LEG	4	1LEG	42	1LEG	40
	1LEG	29	1BQH	59	1VAC	66	1BQH	126
	1BQH	96	1VAC	87	1BQH	92	1VAC	165
H2-Db/9	1INQ	17	1INQ	31	1INQ	6	1INQ	19
	1JPG	79	1JPG	143	1JPG	135	1JPG	89
HLA-A*0201/9	1AO7	1	1AO7	1	1HHI	1	1AO7	4
	1HHI	1	1QEW	5	1QEW	6	1QEW	22
	1QEW	2	1HHI	22	1AO7	8	1HHI	35
	1QR1	5	1QR1	430	1QR1	14	1QR1	325
HLA-A*0201/10	1I4F	104	1I4F	105	1I4F	147	1I4F	142

Table 6: Ranking of MHC binding peptides according to their predicted binding affinity by threading using own template in their source protein sequence for scoring matrices (MJ & BT) and two distance criteria (Nearest atom < 4.0 A^0 & C-beta< 7.0 A^0).

lSSN 0973-2063 (online) 0973-2063 (print)
Bioinformation 3(2): 72-82 (2008)