Pergamon

SCIENCE DIRECT®

Bioorganic & Medicinal Chemistry 11 (2003) 2307–2311

BIOORGANIC & MEDICINAL CHEMISTRY

A Comparative Molecular Similarity Indices (CoMSIA) Study of Peptide Binding to the HLA-A3 Superfamily

Pingping Guan, Irini A. Doytchinova and Darren R. Flower*

Edward Jenner Institute for Vaccine Research, Compton, Berkshire RG20 7NN, UK

Received 24 September 2002; accepted 4 February 2003

Abstract—Epitope identification is the basis of modern vaccine design. The present paper studied the supermotif of the HLA-A3 superfamily, using comparative molecular similarity indices analysis (CoMSIA). Four alleles with high phenotype frequencies were used: A*1101, A*0301, A*3101 and A*6801. Five physicochemical properties—steric bulk, electrostatic potential, local hydrophobicity, hydrogen-bond donor and acceptor abilities—were considered and 'all fields' models were produced for each of the alleles. The models have a moderate level of predictivity and there is a good correlation between the data. A revised HLA-A3 supermotif was defined based on the comparison of favoured and disfavoured properties for each position of the MHC bound peptide. The present study demonstrated that CoMSIA is an effective tool for studying peptide–MHC interactions.

Introduction

Human class I major histocompatibility complex (MHC) molecules are membrane glycoproteins that bind to degraded intracellular antigen fragments, and present them to cytotoxic T lymphocytes (CTL) in order to elicit an immune response.[1–3]

The three-dimensional structures of HLA-A2,[4] Aw68[5] and B27[6] are used as the prototypes for all MHC class I molecules. The major part of the class I molecule is formed by a transmembrane heavy chain of 44 kDa folded into three domains: α1, α2 and α3.[7] These domains have been defined by sequence analysis and do not correspond to the two structural domains apparent in the extracellular part of MHC crystal structures. We use this nomenclature to remain consistent with the immunological literature. α1 and α2 form the peptide binding domain which contains the peptide binding groove and the site of interaction with T cell receptors.[8,9] The α3 domain is involved in interactions with CD8.[10] Native class I MHCs form a complex with a small polypeptide chain of 12 kDa, called β_2-microglobulin. β_2-microglobulin is not encoded by an MHC gene.[11] Sequence analysis has shown that the peptide domains α1 and α2 are polymorphic, while the α3 domain, the transmembrance and cytoplasmic regions are more conserved. Twenty residues are found to be the most variable.[12] Study of the HLA-A2 structure shows that most of the variable residues in the domain form contacts with the peptide, giving the MHC molecule a broad specificity range, allowing it to bind a wide variety of peptides.[4]

Recently, it was proposed that HLA class I molecules can be grouped into supertypes on the basis of similar peptide-binding characteristics.[13] The peptides binding to MHC molecules are called epitopes and the specific arrangement of certain residues along the epitope is called a motif. Four HLA supertypes have been described: A2, A3, B7 and B44.[14] The HLA-A3 supertype covers 44% of the human population and includes five main alleles: A*0301, A*1101, A*3101, A*3301 and A*6801.[14] The common binding motif (or supermotif) for this supertype is characterised by the presence of a hydroxyl containing (Ser or Thr) or hydrophobic (Leu, Ile, Val or Met) residue at position 2 and a positively charged amino acid (Arg or Lys) at the C-terminus.[14]

In the present study, we analyse the nature of peptide binding to the A3 supertype using three-dimensional quantitative structure–activity relationship (3D-QSAR) analysis. Comparative molecular similarity indices analysis (CoMSIA)[15] was used. In CoMSIA, changes in ligand affinities are directly related to changes in molecular properties.[16] This method is good at describing the intermolecular interactions (steric, electrostatic, hydrophobic, hydrogen bond formation) present at the molecular

*Corresponding author. Tel.: +44-1635-577968; fax: +44-1635-577908; e-mail: darren.flower@jenner.ac.uk

0968-0896/03/$ - see front matter © 2003 Published by Elsevier Science Ltd.
doi:10.1016/S0968-0896(03)00109-3

binding site. The method has been used to study the ligand–receptor interactions before and has proved to be of good predictivity.[17,18] In this paper, we apply the method to MHC-peptide interactions. The HLA-A3 superfamily was chosen since it is a relatively large family and has high frequency (>40%) in the general population. The alleles within the family have overlapping specificities. Four members of the A3 family A*1101, A*0301, A*3101 and A*6801 were included in the study. Previously, we applied this method to peptides binding to the HLA-A*0201 allele and found it had a good predictive and explanatory ability.[19] More recently, we extended its application to peptides binding to the HLA-A2 supertype and defined a more precise A2-super-motif.[20] Now we use the CoMSIA technique to define the position-dependent amino acid preferences for peptides binding to HLA-A3 alleles.

Results

Four sets of peptides, one set per allele, were collected from our JenPep[21] database (http://www.jenner.ac.uk/JenPep). Because they were the most common, we used only nonamers in this study of the A3 family. The number of peptides available for alleles A*1101, A*0301, A*3101 and A*6801 were 59, 69, 30 and 39, respectively. Among the selected peptides, some bound to more than one allele. The correlation coefficients between the affinity data for the common peptides range from 0.168 for A*3101/A*6801 ($n=22$) to 0.661 for A*0301/A*1101 ($n=50$). The pIC_{50} ranges are from 3.3 to 3.5 log units.

The binding affinity values (IC_{50}) we used were originally assessed by a competition assay based on the inhibition of the binding of a radiolabeled standard peptide to solubilized MHC molecules.[22] The datasets and their binding affinities used in this study are available upon request.

The all-fields models for the four HLA-A3 alleles are presented in Table 1. The model of A*3101 had the highest predictability ($q^2=0.700$). The q^2 value of the models for A*6801, A*1101 and A*0301 was 0.430, 0.496 and 0.486 respectively. Overall the predictivity of the models was acceptable. They produced 56–90% of their affinity predictions with residuals <0.5 log units and the percentage of poorly predicted peptides (residuals >1.0) was between 0 and 18%. Neither over-prediction nor under-prediction was observed. There are several badly predicted peptides but they are not treated as outliers.

The values of r^2 were greater than 0.9 for the four models, indicating a good explanation of the variance of the data. As the affinity range for the separate alleles was slightly different, the ratio of the SEP to affinity range and SEE to affinity range were used to assess the fitness and predictivity of the models. This ratio should generally be close to 10% for good QSAR models[25] and as a rule the ratio SEP/affinity range is higher than the ratio SEE/affinity range. The present models had ratios from 16.5 to 18.6% for SEP/affinity range and from 2.8 to 8.4% for SEE/affinity range (Table 1).

Five contour maps were generated for each allele, representing the five physicochemical properties: steric

Table 1. CoMSIA models for alleles A*1101, A*0301, A*3101 and A*6801

		A*1101		A*0301		A*3101		A*6801	
Number of peptides		59		69		30		39	
Grid spacing (Å)		2		2		1.5		2	
α[a]		0.6		0.6		0.6		0.5	
Column filtering (kcal/mol)		0.5		0.5		0.5		0.5	
q^2_{LOO}[b]		0.496		0.486		0.700		0.430	
Number of components		8		6		4		5	
SEP[c]		0.588		0.629		0.551		0.674	
SEP/affinity range (%)		17.3		18.1		16.5		19.1	
q^2_{CV5}[d]		0.416		0.424		0.640		0.349	
r^2		0.972		0.959		0.921		0.950	
SEE[e]		0.141		0.177		0.282		0.119	
SEE/affinity range (%)		4.1		5.1		8.4		3.4	
$r^2_{bootstrap}$		0.989		0.971		0.950		0.961	
F ratio		167.666		241.818		73.177		126.217	
Fraction	Steric	0.114		0.104		0.071		0.126	
	Electrostatic	0.234		0.277		0.254		0.251	
	Hydrophobic	0.250		0.260		0.225		0.257	
	H donor	0.261		0.226		0.364		0.241	
	H acceptor	0.141		0.133		0.087		0.125	
$\|res\| \leq 0.5$		40	67.80%	42	60.87%	27	90%	22	56.41%
$0.5 < \|res.\| \leq 1.0$		15	25.42%	20	27.03%	3	10%	10	25.64%
$\|res.\| > 1.0$		4	6.78%	7	9.46%	0	0%	7	17.95%
Mean |residual|		0.443		0.585		0.179		0.516	
Standard deviation		0.343		0.500		0.188		0.377	

[a]Attenuation factor.
[b]q^2 Factor obtained after leave-one-out cross-validation.
[c]Standard error of prediction.
[d]q^2 Obtained by cross-validation in five groups.
[e]Standard error of estimate.

bulk, electrostatic potential, local hydrophobicity, hydrogen-bond donor and acceptor abilities. The maps were produced using non-cross-validated PLS and were displayed in Figures 1–3. A property that is favoured for two or more alleles without being disfavoured for any of them was considered as a favoured property for the supertype. A property disfavoured for two or more alleles was defined as being disfavoured for the supertype.

Comparing the fractional values of the different properties in Table 1, steric complementary was not the major property involved in peptide–MHC binding. Figure 1a showed that the contribution of steric bulk was very different among the alleles. Steric bulk was favoured at positions 5 and 8. Negative electrostatic potential was favoured at positions 3 and 8 (Fig. 1b). Hydrophobicity was favoured at positions 3 and 7 and disfavoured at position 6 (Fig. 2a). Hydrogen-bond donor groups in the ligand (hydrogen-bond acceptors on the receptor) were favoured at positions 1, 3, 4 and 6 and disfavoured at position 5 (Fig. 2b). As was evident from Figure 3, hydrogen acceptors in the ligand (hydrogen-bond donors on the receptor) were well accepted at positions 1 and 6, but they were disfavoured at position 5.

Discussion

The motif of HLA-A3 superfamily includes main anchor positions 2 and 9.[26] Peptides bound to members of the A3 family usually had a positively charged residue—Arginine or Lysine—at the C terminus, and a variety of hydrophobic residues at position 2. In the present study it was found that steric bulk was favoured at position 2 for A*0301 and A*3101 but disfavoured in A*1101 and A*6801 models. The study of crystal structures of

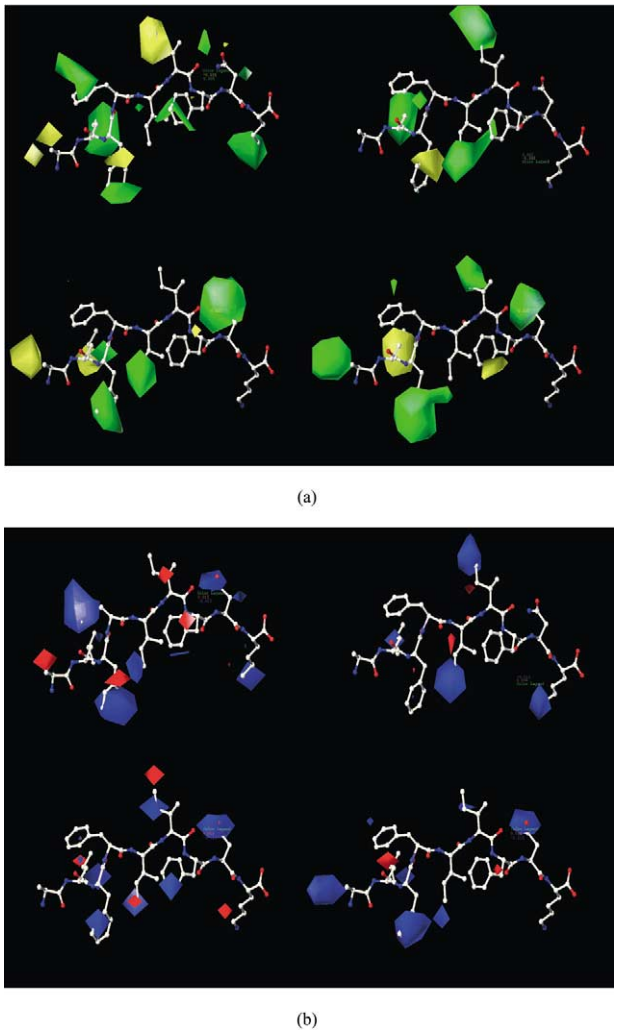

Figure 1. CoMSIA stdev*coeff contour maps. Peptide ALFFIIFNK is shown inside the fields. The peptide is positioned with the N-terminus and position 1 to the left. (a) Steric bulk. Green and yellow areas indicate where steric bulk will increase or decrease the affinity, respectively. Upper left: A*0301 allele. Upper right: A*3101 allele. Lower left, A*1101 allele. Lower right, A*6801 allele. (b) Electrostatic potential. Blue and red areas indicate where negative electrostatic potential will increase or decrease the affinity, respectively. Upper left: A*0301 allele. Upper right: A*3101 allele. Lower left: A*1101 allele. Lower right: A*6801 allele.

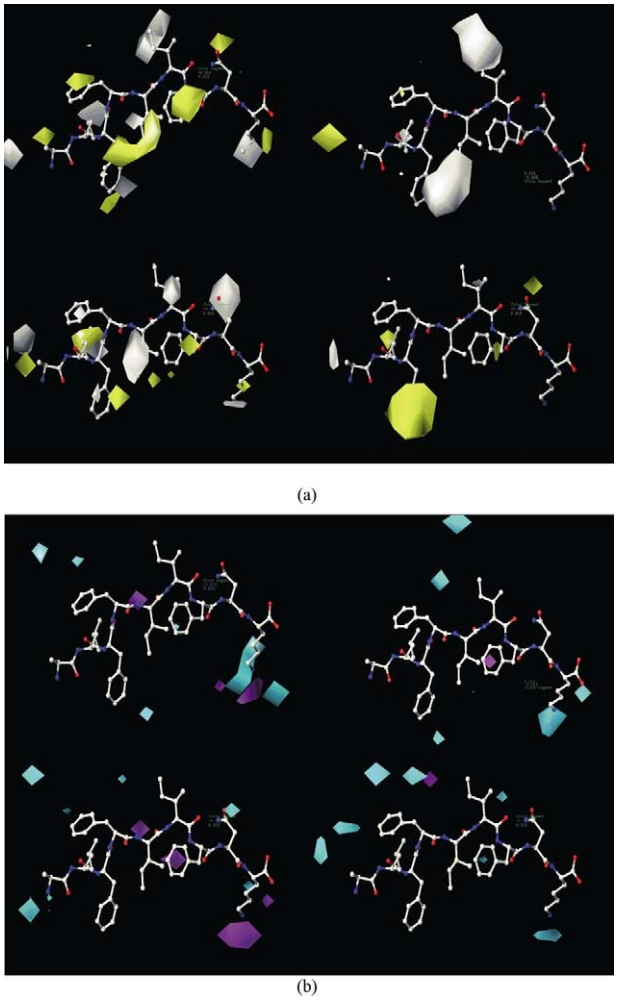

Figure 2. CoMSIA stdev*coeff contour maps. Peptide ALFFIIFNK is shown inside the fields. The peptide is positioned with the N-terminus and position 1 to the left. (a) Local hydrophobicity. Yellow and white areas indicate where hydrophobic amino acid residues will increase or decrease the affinity, respectively. Upper left: A*0301 allele. Upper right: A*3101 allele. Lower left, A*1101 allele. Lower right, A*6801 allele. (b) Hydrogen-bond donor ability. Cyan and purple areas indicate where hydrogen-bond donor groups on the ligand will increase or decrease the affinity, respectively. Upper left: A*0301 allele. Upper right: A*3101 allele. Lower left: A*1101 allele. Lower right: A*6801 allele.

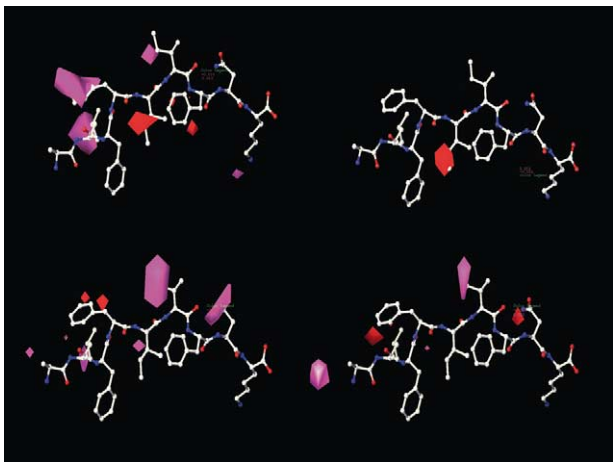

Figure 3. Hydrogen-bond acceptor ability. Peptide ALFFIIFNK is shown inside the fields. The peptide is positioned with the N-terminus and position 1 to the left. Magenta and red areas indicate where hydrogen-bond acceptor groups on the ligand will increase or decrease the affinity, respectively. Upper left: A*0301 allele. Upper right: A*3101 allele. Lower left: A*1101 allele. Lower right: A*6801 allele.

MHC molecules showed that the residue at peptide position 2 bound in pocket B.[4,6] There are different residues lining pocket B in the different MHC-A3 molecules: Tyr9 in A*1101 and A*6801, Phe9 in A*0301 and Thr9 A*3101.[27] This means more space in pocket B for A*0301 and A*3101, allowing them to accommodate larger side chains. Electrostatic potential, hydrophobicity and hydrogen bond acceptance maps were very varied at this position. This was in good agreement with the broad spectrum of amino acids observed at this position, from the bulky hydrophobic Leu to the small polar Thr.

The most important property for the amino acid at position 9 was hydrogen-bond donor ability. It was favoured by A*6801 and A*3101, and was disfavoured by A*1101. For A*0301 were found areas of favoured and disfavoured hydrogen bond donor groups at this position. In some cases, the change of Lys to the larger residue Arg could affect the expression of the molecule.[26] Results from the present study suggested the interaction between the residue at peptide position 9 and the MHC molecule may play an important role. The side chain of larger basic residue Arg could extend to the bottom of the pocket F of A*6801 and A*3101, forming complex stabilising hydrogen bonds with residues at the bottom of the pocket.

Among the secondary anchors, positions 1, 3, 5, 6 and 7 were of great importance. The common favoured property for position 1 was hydrogen-bond donor/acceptor ability. Hydrogen-bond donor groups with negative electrostatic potential were preferred at position 3 for three of the alleles. Sidney and co-workers[28] found that peptides with an aromatic residue, like Tyr, Phe and Trp, had a 31-fold increase in binding affinity to A*0301. Bulky side chains with negative electrostatic potential were preferred at position 5. Hydrogen-bond donors and acceptors were disfavoured here. Hydrophilic amino acids capable of forming hydrogen bonds were well accommodated at position 6. The only common favoured property for position 7 was hydrophobicity.

Positions 4 and 8 face the T-cell receptor,[29] but can still contribute to the affinity. Hydrogen-bond donor ability was important for position 4. Steric bulk and negative electrostatic potential were favoured at position 8.

Conclusion

Our 3D-QSAR study has characterised the contribution to A3-supertype binding made at each peptide position with respect to favoured or disfavoured physiochemical properties. Besides the detailed explanatory ability, CoMSIA can also be used for quantitative binding affinity prediction. CoMSIA has proved to be an extremely effective method in describing ligand–receptor interactions in small molecule drug design. The present study demonstrated that it could also be used to characterise binding motifs for MHC molecules, and that it was a reliable method for epitope prediction.

Experimental

Molecular modelling and CoMSIA

No X-ray data was available for a peptide–HLA-A3 molecule complex. As the A2 supertype is the closest to the A3 supertype, a crystallographic structure for the peptide TLTSCNTSV binding to HLA-A*0201 allele was chosen as the starting conformation.[23] All molecular modelling calculations were performed on a Silicon Graphics workstation using SYBYL6.7[24] as previously described.[19] Briefly, using the X-ray peptide as a template all the studied peptides were built, energy optimised, their charges calculated using the AM1 Hamiltonian as implemented in MOPAC[30] and finally the structures were aligned based on the backbone atoms. The molecules were considered in their neutral forms.

The peptides were evaluated using the five CoMSIA physicochemical properties included in the QSAR module of Sybyl6.7: steric bulk, electrostatic potential, local hydrophobicity, hydrogen-bond donor and acceptor abilities. The properties were evaluated using a probe atom placed at regular intervals within the grid. The atom had radius 1 Å, charge, hydrophobicity, hydrogen bond donor and acceptor properties all equal to +1. Similarity indices were calculated using a Gaussian-type distance dependence between the probe and the atoms of the peptides tested.

The CoMSIA analyses were performed using the partial least square (PLS) method. The initial models were calibrated with respect to grid spacing, attenuation factor and column filtering. The grid was extended 2.0 and 4.0 Å beyond the aligned molecules. As no significant differences in the q^2 values were found, the smaller grid box was used further in the study. Different values were tested for grid spacing: 1.0–2.5 Å in steps of 0.5 Å. Values for the attenuation factor α were varied from 0.3

to 0.7 Å in steps of 0.1 Å, column filtering σ_{min}- from 0.5 to 3.5 kcal/mol in steps of 0.5 kcal/mol. The cross-validated models were assessed by the leave-one-out cross-validated (LOO-CV) coefficient q^2_{LOO}, standard error of prediction SEP and the mean absolute value of the residuals between experimental affinities and those predicted by LOO-CV, presented as negative logarithms of IC_{50}. The optimal number of components was defined as the number of components leading to the highest q^2_{LOO} and the lowest SEP. The difference in q^2 values for the optimal (n_{opt}) and preceding number ($n_{opt}-1$) of components was more than 5%. The non-cross-validated models were assessed by the non-cross-validated r^2, standard error of estimate SEE and F-ratio. The ratio of the standard errors to the affinity range was used as a more effective measure of model predictivity and goodness of fit. Cross-validation in five groups was used as a more robust method for predictivity assessment of the models and the mean of 20 runs is given as q^2_{CV5}. A bootstrap analysis was performed to assess the stability of the non-cross-validated models; $r^2_{boostrap}$ is a mean of 20 runs.

The contour maps

The results of the non-cross-validated CoMSIA models were displayed as contour maps. The contour map highlighted whether different changes in the peptide structure favour or disfavour binding to MHC molecule. The maps from the present study were generated using the stdev*coeff option, using the actual values. Five maps for each of the four alleles were generated for each of the five physicochemical properties. The contours of the CoMSIA steric map are shown in green (more bulk is favoured) and yellow (less bulk is favoured). The electrostatic map has blue (negative potential is favoured) and red (negative potential is disfavoured) coloured contours. CoMSIA hydrophobic fields are coloured yellow (hydrophobic amino acids enhance affinity) and white (hydrophilic amino acids reduce affinity). In the hydrogen-bond donor map cyan and purple areas indicate where hydrogen-bond donor groups on the ligand enhance or reduce the binding, respectively. Analogously, in the hydrogen-bond acceptor map magenta and red areas indicate where hydrogen-bond acceptor groups on the ligand enhance or reduce the binding, respectively.

References and Notes

1. Rau, L.; Cohan, N.; Robert *J. Transplant*. **2001**, *72*, 1830.
2. Ljunggren, H. G.; Thorpe, C. J. *Histol. Histopath*. **1996**, *11*, 267.
3. Madnaka, K.; Yvonne Jones, E. *Curr. Opin. Immunol*. **1999**, *9*, 745.
4. Saper, M. A.; Bjorkman, P. J.; Wiley, D. C. *J. Mol. Biol*. **1991**, *219*, 277.
5. Garrett, T. P. J.; Saper, M. A.; Bjorkman, P. J.; Strominger, J. L.; Wiley, D. C. *Nature* **1989**, *342*, 692.
6. Madden, D. R.; Gorga, J. C.; Strominger, J. L.; Wiley, D. C. *Nature* **1991**, *353*, 321.
7. Krensky, A. M.; Clayberger, C. *Int. Rev. Immunol*. **1996**, *13*, 173.
8. Takiguchi, M. *Nippon Rinsho. Jpn. J. Clin. Med*. **1994**, *52*, 2817.
9. Yvonne Jones, E. *Curr. Opin. Immunol*. **1997**, *9*, 75.
10. Shields, M. J.; Hodgson, W.; Ribaudo, R. K. *Mol. Immunity* **1999**, *36*, 561.
11. Lawlor, D. A.; Zemmour, J.; Ennis, P. D.; Parham, P. *Annu. Rev. Immunol*. **1990**, *8*, 23.
12. Parham, P.; Lomen, C. E.; Lawlor, D. A.; Ways, J. P.; Holmes, N.; Coppin, H. L.; Salter, R. D.; Wan, A. M.; Ennis, P. D. *Proc. Natl. Acad. Sci. U.S.A*. **1988**, *85*, 4005.
13. Sette, A.; Sidney, J. *Curr. Opin. Immunol*. **1998**, *10*, 478.
14. Sidney, J.; Grey, H. M.; Kubo, R. T.; Sette, A. *Immunol. Today* **1996**, *17*, 261.
15. Klebe, G. *Perspect. Drug. Discov. Des*. **1998**, *12-14*, 87.
16. Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem*. **1994**, *37*, 4130.
17. Kim, K. H. *Biol. Med. Chem*. **2002**, *10*, 1367.
18. Huang, X.; Liu, T.; Gu, J.; Luo, X.; Ji, R.; Cao, Y.; Xue, H.; Wong, J. T.; Wong, B. L.; Pei, G.; Jiang, H.; Chen, K. *J. Med. Chem*. **2001**, *44*, 1883.
19. Doytchinova, I. A.; Flower, D. R. *J. Med. Chem*. **2001**, *44*, 3572.
20. Doytchinova, I. A.; Flower, D. R. *J. Comp. Aided Mol. Des*. **2002**, *16*, 535.
21. Blythe, M. J.; Doytchinova, I. A.; Flower, D. R. *Bioinformatics* **2002**, *18*, 434.
22. Ruppert, J.; Sidney, J.; Celis, E.; Kubo, R. T.; Grey, H. M.; Sette, A. *Cell* **1993**, *74*, 929.
23. Madden, D. R.; Garboczi, D. N.; Wiley, D. C. *Cell* **1993**, *75*, 693.
24. *Sybyl 6.7*; Tripos Inc.: 1699 Hanley Road, St. Louis, MO 63144, USA.
25. *Apex-3-D 95.0 User Guide*; Biosym/MSI: 9685 Scranton Road, San Diego, CA 92121, USA.
26. Zhang, Q. J.; Gavioli, R.; Klein, G.; Masucci, M. G. *Proc. Natl. Acad. Sci. U.S.A*. **1993**, *90*, 2217.
27. Schönbach, C.; Koh, J. L. Y.; Sheng, X.; Wong, L.; Brusic, V. *Nucleic Acids Res*. **2000**, *28*, 222.
28. Sidney, J.; Grey, H. M.; Southwood, S.; Celis, E.; Wentworth, P. A.; del Guercol, M. F.; Kubo, R. T.; Chestnut, R. W.; Sette, A. *Human Immunol*. **1996**, *45*, 79.
29. Silver, M. L.; Guo, H. C.; Strominger, J. L.; Wiley, D. C. *Nature* **1992**, *360*, 367.
30. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc*. **1985**, *107*, 3902.