Available online at www.scholarsresearchlibrary.com

Scholars Research Library

Annals of Biological Research, 2010, 1 (2) : 16-22
(http://scholarsresearchlibrary.com/archive.html)

Annals of Biological Research
Scholars Research Library
ISSN 0976-1233

In silico epitope structure prediction for matrix protein of H1N1

AnupTripathi, Deepesh Kumar, Sachidanand Singh, Atul Kumar, Manitha T. P., Babu Ram, K. Panneerselvam and J. Jannet Vennila

Department of Bioinformatics, School of Biotechnology and Health Sciences, Karunya University, Coimbatore, Tamil Nadu, India

---

Abstract

H1N1is a current endemic in both human and pig populations and is the cause of Swine flu in humans. Bioinformatics tools enable us to move rapidly from protein sequence to vaccine design. ProPred-I, Rankpep and PeptGen are servers used for identification of epitope with the help of artificial neural network approach. For H1N1 we identified 10 matrix proteins (M1) which are mainly responsible for propagation of H1N1. From ProPred-I,Rankpep and PeptGen conserved 40 epitopes were identified by their selective algorithms and scoring matrices. A virtual library was designed for the 40 epitopes and further it was used for epitope conservancy analysis tool (IEDB) to narrow down the list of putative epitopes to 20 only. A structural library of all conserved putative epitopes was then minimized with Prime Schrodinger module and then ten putative epitopes were designedwith the motto of identifying best virtual vaccine. The pace of vaccine design will accelerate when these in silico results combined with in vitro methods for screening and confirming epitope.

Key words: Capsid Protein, peptide, epitope, matrix protein, H1N1.

INTRODUCTION

A serotype of Influenza virus A, H1N1 is a current endemic in both human and pig populations and is the cause of Swine flu in humans [1]. Influenza 'A' viruses are enveloped RNA viruses with an eight-segmented, single-stranded, negative-sense genome belonging to the family *Orthomyxoviridae*. The segment of influenza A virus having eight gene, encoded 10 proteins: hemagglutinin (HA), neuraminidase (NA), matrix proteins M2 and M1, nonstructural (NS) proteins NS1 and NS2, the nucleocapsid, and the three polymerases, the PB1 (polymerase basic 1), PB2, and PA (polymerase acidic) proteins [2].Influenza type A viruses are sub-typed based upon the HA and NA antigens, which are surface proteins found on the viral envelope [3]. The capsid is the protein shell of a virus encloses the genetic material of the virus. Matrix protein of H1N1 virus is the outer covering which contains the epitope detection site [4]. So by designing

the peptides complementary to the epitopes, an insight of preventing infection can be done. Epitopes were predicted for H1N1 virus with the help of Propred1 [5], PeptGen and Rankpep software. After that IEDB [6]conservancy analysis has been performed for the identified epitopes. The nonamer epitopes obtained are then designed in ISIS Draw and 3D optimized in ChemSketch.A virtual library is prepared by finally minimizing the structures through Prime (Minimization) module of schrödinger which gave the value of potential energy for each epitope and enable to select the best ten epitope on the basis of minimum potential energy.

MATERIALS AND METHODS

The complete genome information was collected from NCBI (M1_I33A0) and the protein sequence (Indian Strain) was retrieved from SWISS-PROT (Q76V10). Then the protein sequence was submitted to three different online tools for epitope prediction Propred I, Rank Pep and Peptgen. Common eptitopes predicted from the three tools were taken and submitted to "Epitope Conservancy Analysis Tool" from Immunological epitope databaseto find the degree of conservancy of an epitope within a given protein sequence.Structures of the resulting epitopes from "Epitope Conservancy Analysis Tool" were drawn using ISIS Draw and optimized in ACD Chemsketch software. The energy of designed peptide structure was minimized using Schrödinger software.

RESULTS AND DISCUSSION

T cell immune responses are driven by antigenic epitopes, and hence their identification is important for understanding disease pathogenesis and etiology, and for vaccine design. There are two types of T cell epitopes, named CD8 and CD4, which are only recognized in the context of the MHCI and MHCII molecules, respectively, by the correspondent T-cell types. Engaging both sets of T-cells is desirable for mounting a strong defensive immune response against cancer cells and pathogens. Appropriate processing of antigen peptides must occur prior to their binding to the relevant MHC molecules. Incidentally, the C-terminus of most MHCI-restricted epitopes (CD8-T cell epitopes) results from cleavage by the proteasome, and thus, proteasome specificity is important for determining T-cell epitopes. MHC-I ligands are of short length (8-11), as they are constrained into the MHCI peptide binding groove, with their N- and C-terminal ends connected by a network of hydrogen bonds to conserved residues of the MHCI molecule.

Epitope Prediction by Propred1

The ProPred-I is an on-line service for identifying the MHC Class-I binding regions in antigens. It implements matrices for 47 MHC Class-I alleles, proteasomal and immunoproteasomal models. It is a matrix based method that allows prediction of MHC binding site in an antigenic sequence for 47 MHC class I alleles. The matrix is from BIMAS server that helps in prediction of proteosome&immunoproteosome cleavage sites in an antigenic sequence [7]. It also helps in finding MHC binders, having cleavage site at C-terminus, because of becoming promiscuous potential T-cell Epitopes [8], these epitope can serve as suitable vaccine components. Overlapping 9-mer peptides were calculated by using quantitative matrix for all MHC alleles. The highly predictor binders were given as a result which are greater than the threshold of 4%. Threshold value for binders depends upon its selectivity and sensitivity as shown in figure 1.The 9-mer peptides were selected on the basis of addition matrix, where score of peptide is calculated on the basis of summing the score at each position.The centre position that is 4 rights and 4 left are considered as predicted proteasome cleavage site.

Epitope Selection

Epitopes are selected on the basis of their presence in different alleles. The number of alleles **(n)** is decided by subtracting the number of overlaps **(O)** from the total number of alleles **(N)** of their presence (Eq. 1).

$$\mathbf{n = N - O} \qquad \mathbf{(1)}$$

If the value of 'n' >= 3, then the epitope is taken into consideration.

Figure 1: Overlapping nonamer peptides obtained by propred 1

PeptGen

PeptGen generates peptides in the stairstep patterns having sequence length in (9). The selected peptides were C-terminal cleavage site. The peptides are generated on the basis of proline rule. Hydropathy of each amino acid is calculated as Kyte-Dolittlehydropathy index and also represented in [9]. In figure 2 most hydrophobic amino acids are represented by dark blue and light blue whereas most hydrophilic with red and pink.

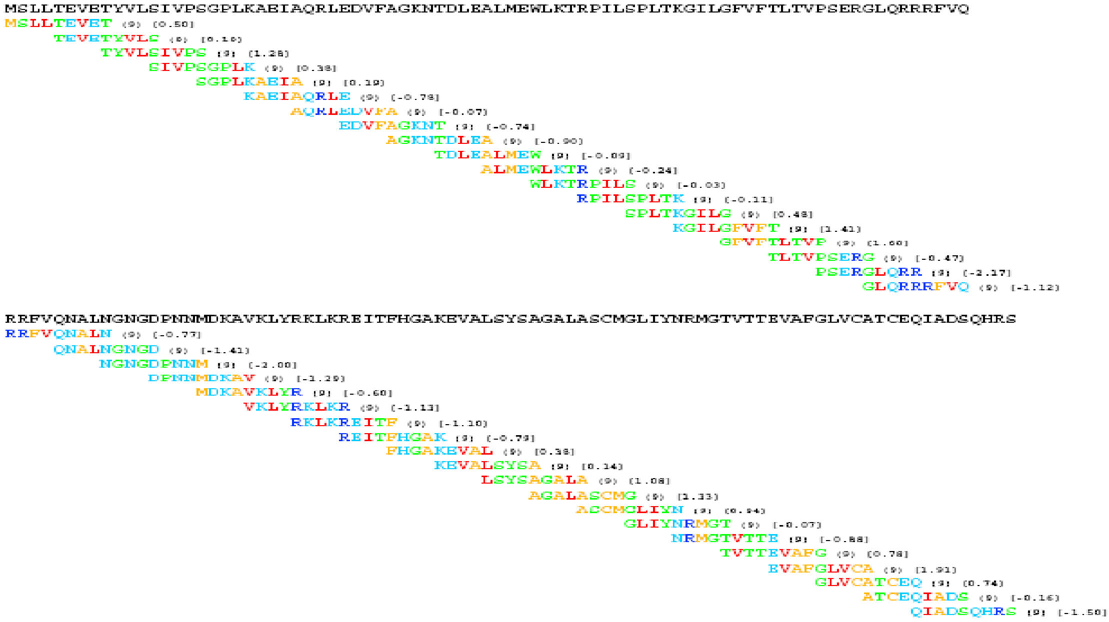

Figure 2: Nonamer epitopes obtained by PeptGen shown in stair step arrangement

Rankpep

This server predicts peptide binders to MHCI molecules from protein sequence/s or sequence alignments using Position Specific Scoring Matrices (PSSMs). In addition, it predicts those MHC-I ligands whose C-terminal end is likely to be the result of proteasomal cleavage[9]. Peptides that bind to a given MHC molecule share sequence similarity. PSSM was used from these alignments of MHC ligands using profile weight. Rankpep selects effective epitope binders on the basis of SVM based classifier trained on both residue properties and amino acid sequence.

RANK	POS.	N	SEQUENCE	C	MW (Da)	SCORE	% OPT.
1	134	IYN	RMGTVTTEV	AFG	975.11	14.781	37.71 %
2	3	MS	LLTEVETYV	LSI	1048.2	12.174	31.06 %
3	58	LTK	GILGFVFTL	TVP	948.17	11.375	29.02 %
4	164	SHR	QMVTTTNPL	IRH	986.13	11.093	28.30 %
5	51	TRP	ILSPLTKGI	LGF	923.16	10.826	27.62 %
6	41	DLE	ALMEWLKTR	PIL	1106.38	10.387	26.50 %
7	116	KEV	ALSYSAGAL	ASC	833.95	9.669	24.67 %
8	211	QAR	QMVQAMRTI	GTH	1059.3	7.299	18.62 %
9	130	CMG	LIYNRMGTV	TTE	1048.26	7.151	18.24 %
10	47	EWL	KTRPILSPL	TKG	1006.26	6.839	17.45 %

Figure 3: Consensus Epitope obtained by Rankpep on the basis of score and optimal score

The output of RANKPEP as shown in figure 3 consists of a list of peptides ordered by their binding potential (score) to the selected MHC molecule. On the basis of score and optimal score (39.198) of the predicted peptide relative to that of the consensus and keeping a binding thresh hold of 8.49, a specific consensus Epitope was obtained as FLWKWHWCV. Also all rows highlighted in red represent predicted binders and a peptide highlighted in violet has a C-terminus predicted by the cleavage model used[10].

Table 1: Library of conserved Epitopes selected by three on line tools i.e. Propred1, PeptGen, Rankpep

S.No.	P.S*	S.No.	P.S*	S.No.	P.S*	S.No.	P.S*
M 1	RMGTVTTEV	M 11	VFAGKNTDL	M 21	RRRFVQNAL	M 31	VETYVLSIV
M 2	LLTEVETYV	M 12	AMEVASQAR	M 22	GAKEVALSY	M 32	TEVETYVLS
M 3	GILGFVFTL	M 13	LYRKLKREI	M 23	QARQMVQAM	M 33	KEVALSYSA
M 4	QMVTTTNPL	M 14	KAVKLYRKL	M 24	QAYQKRMGV	M 34	GAKEVALSY
M 5	ILSPLTKGI	M 15	SLLTEVETY	M 25	MEWLKTRPI	M 35	TEVAFGLVC
M 6	ALMEWLKTR	M 16	VTTTNPLIR	M 26	SAGLKDDLL	M 36	VTTEVAFGL
M 7	QMVQAMRTI	M 17	LKDDLLENL	M 27	SSAGLKDDL	M 37	HENRMVLAS
M 8	LIYNRMGTV	M 18	ALASCMGLI	M 28	THPSSSAGL	M 38	NNMDKAVKL
M 9	KTRPILSPL	M 19	IRHENRMVL	M 29	FHGAKEVAL	M 39	DPNNMDKAV
M 10	RLEDVFAGK	M 20	RGLQRRRFV	M 30	EVETYVLSI	M 40	LGFVFTLTV

** Nonamer peptide obtained by comparative approach of three on line tools Propred1,Rankpep,PeptGen*

IEDB (Epitope conservancy analysis)

Finally 40 epitopes were selected on the basis of the result of the three online tools used for the prediction of epitopes. The common epitopes were selected from the large data set obtained and their virtual library was made as shown in table 1.Conservancy analysis for the obtained epitope was done with the help of IEDB (Immune Epitope Database and Analysis Resource) which is a project hosted by scientists at the La Jolla Institute for Allergy and Immunology (LIAI), with support from the National Institute of Health (NIH), and Department of Health and Human Services (HHS).This tool calculates the degree of conservancy of an epitope within a given protein sequence set at different degrees of sequence identity. Conservancy is defined as the fraction of protein sequences that contain the epitope, and Identity is the degree of correspondence (similarity) between two sequences.

Epitope No.	Epitope name	Epitope sequence	Epitope length	Percent of protein sequence matches at identity ≥100%	Minimum identity	Maximum identity	View details
1	M1	RMGTVTTEV	9	**82.35%** (14/17)	88.89%	100.00%	Go
2	M2	LLTEVETYV	9	**94.12%** (16/17)	88.89%	100.00%	Go
3	M3	GILGFVFTL	9	**88.24%** (15/17)	88.89%	100.00%	Go
4	M4	QMVTTTNPL	9	**70.59%** (12/17)	88.89%	100.00%	Go
5	M5	ILSPLTKGI	9	**88.24%** (15/17)	88.89%	100.00%	Go
6	M6	ALMEWLKTR	9	**76.47%** (13/17)	88.89%	100.00%	Go
7	M7	QMVQAMRTI	9	**64.71%** (11/17)	88.89%	100.00%	Go
8	M8	LIYNRMGTV	9	**82.35%** (14/17)	88.89%	100.00%	Go
9	M9	KTRPILSPL	9	**100.00%** (17/17)	100.00%	100.00%	Go
10	M10	RLEDVFAGK	9	**100.00%** (17/17)	100.00%	100.00%	Go
11	M11	VFAGKNTDL	9	**100.00%** (17/17)	100.00%	100.00%	Go
12	M12	AMEVASQAR	9	**76.47%** (13/17)	88.89%	100.00%	Go
13	M13	LYRKLKREI	9	**88.24%** (15/17)	88.89%	100.00%	Go
14	M14	KAVKLYRKL	9	**23.53%** (4/17)	77.78%	100.00%	Go
15	M15	SLLTEVETY	9	**94.12%** (16/17)	88.89%	100.00%	Go
16	M16	VTTTNPLIR	9	**76.47%** (13/17)	88.89%	100.00%	Go
17	M17	LKDDLLENL	9	**70.59%** (12/17)	77.78%	100.00%	Go
18	M18	ALASCMGLI	9	**100.00%** (17/17)	100.00%	100.00%	Go
19	M19	IRHENRMVL	9	**94.12%** (16/17)	88.89%	100.00%	Go
20	M20	RGLQRRRFV	9	**100.00%** (17/17)	100.00%	100.00%	Go
21	M21	RRRFVQNAL	9	**100.00%** (17/17)	100.00%	100.00%	Go
22	M22	GAKEVALSY	9	**82.35%** (14/17)	77.78%	100.00%	Go
23	M23	QARQMVQAM	9	**82.35%** (14/17)	88.89%	100.00%	Go
24	M24	QAYQKRMGV	9	**100.00%** (17/17)	100.00%	100.00%	Go
25	M25	MEWLKTRPI	9	**100.00%** (17/17)	100.00%	100.00%	Go
26	M26	SAGLKDDLL	9	**70.59%** (12/17)	77.78%	100.00%	Go
27	M27	SSAGLKDDL	9	**70.59%** (12/17)	77.78%	100.00%	Go
28	M28	THPSSSAGL	9	**94.12%** (16/17)	88.89%	100.00%	Go
29	M29	FHGAKEVAL	9	**70.59%** (12/17)	77.78%	100.00%	Go
30	M30	EVETYVLSI	9	**100.00%** (17/17)	100.00%	100.00%	Go
31	M31	VETYVLSIV	9	**70.59%** (12/17)	88.89%	100.00%	Go
32	M32	TEVETYVLS	9	**100.00%** (17/17)	100.00%	100.00%	Go
33	M33	KEVALSYSA	9	**11.76%** (2/17)	77.78%	100.00%	Go
34	M34	GAKEVALSY	9	**82.35%** (14/17)	77.78%	100.00%	Go
35	M35	TEVAFGLVC	9	**94.12%** (16/17)	88.89%	100.00%	Go
36	M36	VTTEVAFGL	9	**94.12%** (16/17)	88.89%	100.00%	Go
37	M37	HENRMVLAS	9	**94.12%** (16/17)	88.89%	100.00%	Go
38	M38	NNMDKAVKL	9	**23.53%** (4/17)	88.89%	100.00%	Go
39	M39	DPNNMDKAV	9	**23.53%** (4/17)	88.89%	100.00%	Go
40	M40	LGFVFTLTV	9	**100.00%** (17/17)	100.00%	100.00%	Go

Figure 3: IEDB Epitope conservancy tabular view

The Figure 3 view shown for each epitope, the calculated degree of conservancy (percent of protein sequence matches a specified identity level) and the matching minimum/maximum identity levels within the protein sequence set. On the basis of identity and similarity the conservancy of 20 epitopes was selected as the best epitope. The nonamer peptides were designed by ISIS Draw and then were 3D optimized by ChemSketch. The epitopes designed were then minimized using Schrodinger Prime module and related potential energy was obtained. The Minimization refinement task performsa truncated-Newton energy minimization, using the OPLS_2005 all-atom force field (protein-optimized) for proteins and OPLS_2001 for cofactors, and treating solvation energies and effects via the Surface Generalized Born (SGB) continuum solvation model. For the ten epitopes minimum energy was obtained as mentioned in table 2 And optimized structure shown in figure 4.

Table 2: Nonamer peptide sequences with their Potential Energy

S.No.	Peptide	Potential Energy (Kcal/mol)	S.No.	Peptide	Potential Energy (Kcal/mol)
1	RRRFVQNAL	-883.373	6	LYRKLKREI	-411.967
2	RGLQRRRFV	-829.907	7	QMVQAMRTI	-390.327
3	QARQMVQAM	-670.982	8	HENRMVLAS	-382.239
4	IRHENRMVL	-545.981	9	LKDDLLENL	-380.821
5	QAYQKRMGV	-461.361	10	DPNNMDKAV	-329.507

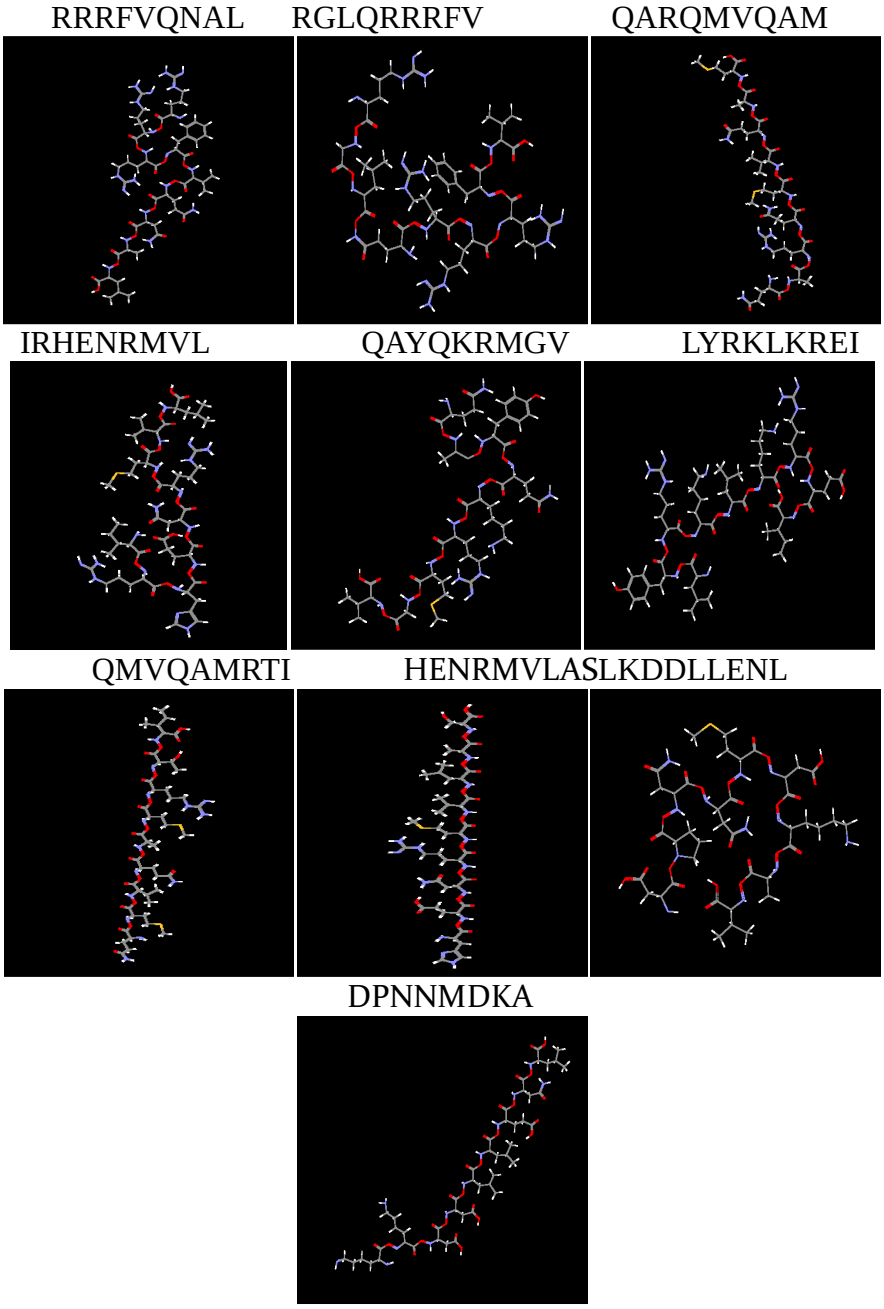

Figure 4: Minimized and optimezed strucure of ten nonamer epitope

The optimized and minimized structure of epitope can be further used for the ADME prediction and for the in-vitro testing on cell lines so that the efficacy of the epitopes can be identified as a effective peptide vaccine for H1N1.

CONCLUSION

T cell immune responses are driven by antigenic epitopes, and hence their identification is important for understanding disease pathogenesis and etiology, and for vaccine design.Epitope designing for the serotype of Influenza virus A i.e H1N1 was performed by retrieving the sequence from Swiss prot, Indian strain Q76V10. Identical 40 nonamers epitopes were selected as conserved epitopes obtained from Propred I, Rankpep and Peptgen by different scoring algorithms. The 40 epitopes selected are basically C-terminal cleavage which are proteosomal site. For forty epitopes selected, IEDB analysis was performed for getting the percentage of identity and similarty , the minimum identity was upto 77.78% and max identity was about 100%. Conservancy calculated for all nonamer peptide was in average 92%. Finally structure was designed for the nonamer's and energy was minimized using Schrodinger Prime module. Ten best selected peptides on the basis of minimum potential energy were selected, further these epitope can be checked for the invitro activities on specific cell lines . Epitope prediction done gives an insight to predict more peptide vaccine relavant for the influenza A virus (H1N1).

REFERENCES

[1] Markushin,S., Ghiasi,H., Sokolov,N., Shilov,A., Sinitsin,B., Brown,D., Klimov,A. and Nayak,D. (**1988**). *Virus Res.* 10: 263-271.
[2] Bouvier, N.Mand Palese, P. (**2008**). *Vaccine* 4: 49– 53.
[3] Ito,T., Gorman,O.T., Kawaoka,Y., Bean,W.J. and Webster.R.G. **1991,** *J. Virol*. 65:5491-5498.
[4] Kobasa, D., Jones, S.M and Shinya, K. (**2007**) *Nature* 445: 319–323.
[5] Bhasin,M., Singh,H. and Raghava, G.P.S. (**2002**) *Nucleic Acid Res*. 43:1276-1287.
[6] Bui,H.H., Sidney,J., Li,W., Fusseder,N. and Sette,A. (**2007**) *BMC Bioinformatics*.8:361.
[7] Toes,R.E., Nussbaum,A.K., Degermann,S., Schirle,M., Emmerich,N.P.N., Kraft, M. Laplace, C., Zwinderman, A., Dick,T.P., Muller,J., Schonfisch,B., Schmid,C., Fehling,H.J., Stevanovic,S., Rammensee,H.G., Schild,H. (**2001**) *J. Exp. Med*., 194, 1-12.
[8] Kessler, J.H., Beekman, N.J., Bres-Vloemans, S.A., Verdijk, P., Veelen, P.A., Kloosterman Joosten,A.M., Vissers,D.C.J., Bosch, G.J.A., Kester, M.G.D., Sijts, A., Drijfhout, J.W., Ossendrop,F., Offringa,R. and Melief, C.J.M. (**2001**) *J. Exp. Med*. 193, 73-88.
[9] Kyte,J. and Doolitle,R.F. (**1982**). *J. Mol. Biol*. 157:105-132.
[10] Ruppert, J., Sidney, J., Celis, E., Kubo,R.T., Grey,H.M andSette. *Cell* 74:929, **1993**.