Indian Journal of Biochemistry & Biophysics
Vol. 52, February 2015, pp. 34-44

Computational prediction of MHC class I epitopes for most common viral diseases in cattle (*Bos taurus*)

Tanmaya Kumar Sahu[1], A R Rao[1]*, Prabina Kumar Meher[2], Bishnu Charan Sahoo[1], Satakshi Gupta[1] and Anil Rai[1]

[1]Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, New Delhi, 110012, India
[2]Division of Statistical Genetics, ICAR-Indian Agricultural Statistics Research Institute, New Delhi, 110012, India

Received 31 March 2014; revised 01 December 2014

Viral diseases like foot-and-mouth disease (FMD), calf scour (CS), bovine viral diarrhea (BVD), infectious bovine rhinotracheitis (IBR) etc. affect the growth and milk production of cattle (*Bos taurus*) causing severe economic loss. Epitope-based vaccine designing have been evolved to provide a new strategy for therapeutic application of pathogen-specific immunity in animals. Therefore, identification of major histocompatibility complex (MHC) binding peptides as potential T-cell epitopes is widely applied in peptide vaccine designing and immunotherapy. In this study, MetaMHCI tool was used with seven different algorithms to predict the potential T-cell epitopes for FMD, BVD, IBR and CS in cattle. A total of 54 protein sequences were filtered out from a total set of 6351 sequences of the pathogens causing the said diseases using bioinformatics approaches. These selected protein sequences were used as the key inputs for MetaMHCI tool to predict the epitopes for the BoLA-A11 MHC class I allele of *B. taurus*. Further, the epitopes were ranked based on a proposed principal component analysis based epitope score (PbES). The best epitope for each disease based on its predictability through maximum number of predictors and low PbES was modeled in PEP-FOLD server and docked with the BoLA-A11 protein for understanding the MHC-epitope interaction. Finally, a total of 78 epitopes were predicted, out of which 27 were for FMD, 25 for BVD, 12 for CS and 14 for IBR. These epitopes could be artificially synthesized and recommended to vaccinate the cattle for the considered diseases. Besides, the methodology adapted here could also be used to predict and analyze the epitopes for other microbial diseases of important animal species.

Keywords: Foot-and-mouth disease**,** Calf scour, Bovine viral diarrhea**,** Infectious bovine rhinotracheitis, Major histocompatibility complex**,** Principal component analysis.

The common viral diseases in cattle like foot-and-mouth disease (FMD), calf scour (CS), bovine viral diarrhea (BVD), infectious bovine rhinotracheitis (IBR) etc. adversely affect the growth and milk production of cattle, in turn affecting the economy of the countries dependent on agriculture. FMD is a highly contagious viral infection in wild and domestic cloven-hoofed animals. The foot-and-mouth disease virus (FMDV) is classified as a member of the *Aphthovirus* genus and causes a drastic decrease in performance of production traits due to the formation of painful blisters in epithelial sites, primarily mouth and feet[1]. There are seven serotypes of FMDV (A, O, C, Asia1, SAT1, SAT2 and SAT3) exist in nature[2] and out of them, the serotype O is the most prevalent and found in many parts of the world[3].

Some of the FMD symptoms like lesions in nose or mouth resemble BVD, which often creates confusion between these two diseases. However, BVD is characterized by severe erosive lesions in oral and intestinal mucosa, diarrhea and death[4] and is caused by a single-stranded RNA virus called bovine viral diarrhea virus (BVDV) which belongs to the *Pestivirus* genus of Flaviviridae family. BVDV is divided into two different genotypes (genotypes I and II), which are distinct from one another, but responsible for same disease[5]. Like BVD, CS is also caused by an RNA virus, which is double stranded and referred as bovine rotavirus (BRV) that belongs to the family Reoviridae. It causes infantile

*Corresponding author:
Email: arrao@iasri.res.in; rao.cshl.work@gmail.com
Phone: +91-9999422935
Fax: +91-11-25841564

Abbreviations: BHV-1, bovine herpesvirus 1; BoLA, bovine lymphocyte antigen; BRV, bovine rotavirus; BVD, bovine viral diarrhea; BVDV, bovine viral diarrhea virus; CMI, cell mediated immunity; CS, calf scour; CTLs, cytotoxic T lymphocytes; DS, discovery studio; FMD, foot and mouth disease; FMDV, foot and mouth disease virus; H-bond, hydrogen bond; IBR, infectious bovine rhinotracheitis; MHC, major histocompatibility complex; MSA, multiple sequence alignment; NCBI, national center for biotechnology information; PbES, principal component analysis based epitope score; PCA, principal component analysis; RMSD, root mean square deviation; TCRs, T-cell receptors; UTR, untranslated region.

gastroenteritis and results several deaths worldwide each year[6]. Similarly, infectious bovine rhinotracheitis (IBR) is another harmful cattle disease caused by herpesvirus infection and has diverse clinical manifestations. It is mainly known as respiratory tract disease, characterized by tracheitis, rhinitis and fever. The causal virus, bovine herpesvirus 1 (BHV-1) is easily transmitted and is distributed worldwide[7]. BHV-1 infection is also an important component of an upper respiratory tract infection referred to as shipping fever or bovine respiratory complex[8].

The vaccines are developed to configure the immune system to protect the body from a specific viral attack in future. Biologically, vaccine contains an agent resembling to a disease-causing pathogen, which is generally the weakened or heat-killed form of the pathogen. Generally, these agents bind with two types of major histocompatibility complex (MHC) molecules that are involved in the antigen presentation to T-Cell receptors. MHC class I molecules present endogenous antigens to CD8+ cytotoxic T-Cells, whereas MHC class II molecules present the exogenously-derived proteins to CD4+ helper T-Cells[9]. The cytotoxic T lymphocytes (CTLs) play a major role in cell-mediated immunity (CMI) that is vital for the defense against viral diseases, as these CTLs recognize the endogenous antigenic peptides presented by the MHC class I molecules[10,11] and responsible for the immune elimination of intracellular pathogens like viruses[12].

Biologically experimental determination of these antigenic peptides is time-consuming and expensive[13].To address such problems, computational vaccine designing through bioinformatics approaches is being adapted by using various strategies to design novel antigen-specific, epitope-based vaccines[14]. Thus, significant efforts have been made on the development of computational tools for the identification of MHC-binding peptides[15]. Due to diverse advantages of epitope based vaccines, nowadays epitope-based vaccine designing has become a prime interest in the field of modern computational biology. These conserved epitopes are used as vaccines with the expectation that it can protect against multiple pathogenic strains or species. On the other hand, specific epitopes are used for given infectious pathogen strain. However, the degree of conservancy of epitopes is vital in both the cases[16].

In cattle, the MHC class I molecules are potentially available for antigen presentation to CTLs. Data from several bovine lymphocyte antigen (BoLA) workshops have demonstrated the existence of more than 50 serologically defined MHC class I specificities[17]. However, many researchers have conducted computational experiments on the identification of antigenic peptides for binding to the BoLA-A11 molecules. Hegde *et al*.[18] reported the amino acid sequence of nonapeptides with a motif for binding to BoLA class I A11 molecules. Similarly, Hegde and Srikumaran[19] used computer simulation methods to search for potential CTL epitopes from BHV proteins for major BoLA class I alleles (including BoLA A11). Further, Hegde and Srikumaran[20] also identified several antigenic peptides from polyprotein of BVD virus for BoLA-A11, BoLA-A20, BoLA-HD1 and BoLA-HD6. Becker[21] presented a computer analysis that predicts the availability of putative nonapeptides with amino acid motifs from FMDV for binding to BoLA class I A11 and A20 alleles.

In the present study, the antigenic CD8+ cytotoxic T-Cell epitopes from the viral protein sequences for BoLA A11 MHC class I allele of *B. taurus* have been predicted using seven existing epitope predictors (four individual predictors and three meta-predictors). In addition, an index-based approach using principal component analysis (PCA) is proposed by taking into account the epitope scores obtained from all the existing predictors to screen the potential epitopes. Further, the proposed approach has been compared with the existing meta-predictors. Besides, the 3D structures of the best epitope for each disease have been modeled and docked with the BoLA A11 allele to analyze the epitope-MHC interaction. The predicted T-Cell epitopes are expected to contribute in the development of vaccine candidates for the major cattle diseases like FMD, BVD, IBR and CS.

Materials and Methods

Pathogen, diseases and antigenic proteins

For the present study, FMD, CS, BVD and IBR diseases were considered on the basis of mode of infection, virulency and economic importance. All the protein sequences of different serotypes of the causative viruses were downloaded from the protein database of National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov). The total number of downloaded sequences was 6351 from which 4623 were from FMDV, 1414 were from BVDV, 31 sequences from BHV and 283 were from BRV. Short sequences of length less than 50 amino acids and the redundant sequences with 90%

Table 1—Details of pathogenic sequences considered for epitope prediction for different cattle viruses

Disease	Type	No. of downloaded sequences	No. of sequences after redundancy removal	No. of sequences after removal of short sequences
FMDV	Asia 1	371	10	3
FMDV	SAT 1	280	46	2
FMDV	SAT 2	382	39	4
FMDV	SAT 3	72	22	3
FMDV	A	1258	87	6
FMDV	C	156	14	2
FMDV	O	2104	80	5
BVD	1	1216	163	10
BVD	2	198	36	4
BHV	1	31	12	9
BRV	-	283	6	6
Total		6351	515	54

identity were removed. Upon filtration, a total of 54 sequences from all the viruses were selected for epitope prediction (Table 1). The selected virulent proteins along with the information on pathogens, corresponding serotypes, NCBI accessions and related references are listed in Supplementary Table 1 and the corresponding fasta sequences are given in Supplementary Dataset 1 (http://cabgrid.res.in/ijbb/Suppl/).

MHC Class I allele of *Bos taurus* and homology analysis

MHC class I molecules bind to the short antigenic peptides (8-11 aa) that are derived from the pathogen proteins at their antigen binding groove formed by α1 and α2 domains of the MHC molecule. The antigenic peptides are presented by the cell surface of the MHC to the cytotoxic T-Cells (CTLs). Further, these peptides are recognized by T-Cell receptors (TCRs)[22]. Most of the reviewed online epitope prediction tools, including MetaMHCI predict epitopes for the MHC alleles of *Homo sapiens, Mus musculus, Pan troglodytes* and *Macaca mulatta* (*Rhesus macaque*). As such, there exists no provision in these tools to predict potential epitopes against the MHC alleles of *B. taurus.*

Therefore, the closest MHC homologue of *B. taurus* was identified from the above-mentioned species. Hence, a proteome-specific BLAST search was performed for BoLA-A11 (accession 3PWU_A) with the proteomes of *H. sapiens*, *M. musculus*, *P. troglodytes* and *M. mulatta*, respectively. Then, the closest protein sequences were downloaded and subjected to multiple sequence alignment (MSA) along with BoLA-A11 of *B. taurus*. The MSA result was used as prior information for dendrogram generation by neighbor joining algorithm with 1000 bootstrap replications in the interface of ClustalX2. The nearest neighbor of *B. taurus* was identified based on MHC alleles and selected for further analysis. Besides, the crystal structure of BoLA A11 protein with PDB ID: 3PWU and the MHC class I allele of the closest species identified from MSA were downloaded from protein data bank (http://rcsb.org). These two structures were superimposed in discovery studio (DS) visualizer 3.5 and root mean square deviation (RMSD) was calculated and the superimposition at the antigen binding grooves of both the proteins was analyzed.

Epitope prediction using MetaMHCI

MetaMHCI is a server based web application that uses ensemble approaches to predict epitopes having binding affinity towards MHC class I proteins. The length of these epitopes vary from 8 to 11 amino acid residues[23,24], however, the antigen peptides that bind to MHC class I molecules are approximately nine amino acid residues long[25]. MetaMHCI tool computes scores based on efficient prediction algorithms *viz.* ANN[26], SMM[27], NetMHC[28], NetMHCpan[29], Consensus[30], PM[15] and AvgTanh[31] for MHC class I peptide binding prediction. ANN, SMM, NetMHC and NetMHCpan are the individual predictors, whereas Consensus, PM and AvgTanh are meta-predictors. Generally, the individual predictors are standalone models and meta-predictors need the results of individual predictors to give an ensemble prediction and also perform analysis of long protein sequences (max. 1000 sequence length).

The sequences of all the 54 viral proteins listed in Supplementary Table 1 were provided as the sequence input to the MetaMHCI for 9 mer epitope prediction considering MHC of the species close to *B. taurus*. All the predictors (both individual and meta-predictors) were selected for computation of scores. Besides, the species found closer to *B. taurus* in the previous section was set as 'MHC source species' and the corresponding allele as 'MHC allele'. The MetaMHCI predicted all the scores for each of the 9-mer fragments from each sequence.

Index based selection of potential epitopes

Different scores used in MetaMHCI designate different number of nonapeptides as epitopes and the ranks of the epitopes differ from method to method. Therefore, an index score was calculated based on principal component analysis (PCA) using all the

predictors. This index score is expected to rank the epitopes in such a way that the high ranked epitopes are those, which are predicted through maximum number of predictors. Initially, the scores of individual predictors and meta-predictors obtained from MetaMHCI were used to compute a correlation matrix. This correlation matrix was further used to generate the heat map for graphically visualizing the extent of correlation between the considered predictors. Upon confirmation of higher degree of correlation between the individual predictors and meta-predictors, these correlated predictors were transformed to uncorrelated variables called principal components (PC). Then, the PC scores were computed from the principal components using SPSS v16.0. A weighted PC score called PCA based epitope score (PbES) was computed by considering the eigen values of the PCs as weights. The PbES is computed as shown below:

$$PbES = \frac{\left(\sum_{i=1}^{p} \lambda_i f_i\right)}{\left(\sum_{i=1}^{p} \lambda_i\right)}$$

where λ_i is the i^{th} eigen value, f_i is the i^{th} PC score and p is the total number of PCs. Further, the correlation between PbES and the score of existing predictors were computed and based on which the tendency of PbES (increasing or decreasing) for the good binders was decided. Since PbES is also a meta-predictor, it was compared with the other meta-predictors by considering the highest ranked epitopes predicted through the individual predictors. The meta-predictor scores of top 10, 20, 30 and 40 epitopes were analyzed and compared with those of individual predictors separately for all the considered diseases. One potential epitope for each disease was selected based on its predictability through maximum number of predictors and the PbES score. Further, the selected epitopes were considered for the epitope-MHC interaction analysis.

Prediction of epitope structures

The tertiary structures of the best epitopes selected for each disease were predicted using PEP-FOLD server, which adopts a *de novo* approach to predict 3D peptide structures from sequence information based on *ab initio* protein structure modeling. This server is more efficient than other prediction servers for *ab initio* structure prediction, because it can predict the structure of peptide size ranging from 9 to 25 residues[32]. The top five results for each submitted epitope obtained by PEP-FOLD server were validated by PROCHECK of SAVES server[33] and the structures were further refined by Modloop Server[34] and then revalidated by SAVES. This process was repeated till the final structures with an acceptable Ramachandran plot[35] were obtained. The final structures were then energy minimized through YASARA energy minimization server (http://www.yasara.org/minimizationserver.htm).

MHC-Epitope interaction analysis

In the downloaded 3D structure of BoLA-A11 of *B. taurus* (PDB ID 3PWU), one CTL epitope from rinderpest virus was found inside its antigen binding groove. From the analysis of such interaction, active residues of the antigen binding groove were identified as *a priori* information (Supplementary Fig. 1; http://cabgrid.res.in/ijbb/Suppl/) to perform docking of potential epitopes with BoLA-A11. The CTL epitope from rinderpest virus, water molecules and other hetero atoms were removed from the structure and then the structure was minimized using YASARA energy minimization server. Further, the energy minimized selected epitopes for all the four diseases were docked separately with BoLA-A11 protein using the pre-identified binding site information.

The docking was performed using the PatchDock server (http://bioinfo3d.cs.tau.ac.il/PatchDock/). Patch Dock is an efficient algorithm used for bio-molecular docking *i.e.* protein-protein, protein-peptide, protein-DNA and protein-drug molecules. It results a list of potential complexes sorted by shape complementarity criteria[36, 37]. Further, the top 100 results of PatchDock were refined using FireDock server (http://bioinfo3d.cs.tau.ac.il/FireDock/) that addresses the refinement problem of protein-protein docking solutions by targeting the problem of flexibility and scoring of solutions produced by fast rigid-body docking algorithms. Given a maximum of 1000 potential docking candidates, FireDock refines and scores them according to an energy function[38].

From the result of FireDock, top ten solutions based on the global energy were downloaded for all the four diseases. These solutions were visualized in DS visualizer 3.5 and the interaction between the epitope and MHC was analyzed by identifying the intermolecular hydrogen bond (H-Bond) and a Van-der-Waals interaction (Bumps)[39]. The amino acid residues participating in the MHC-Epitope interaction

from the top ten solutions were also identified and the best solutions were considered based on global energy of the solution, total number of interaction and number of receptor residues participated in the reference model 3PWU for each of the four diseases.

Results

A total of 54 antigenic proteins considered in this study for epitope prediction mostly belong to the categories of capsid proteins, glycoproteins, polyproteins, polymerases, proteases, DNA packaging and tegument proteins etc. These proteins are evidenced to be related with FMD, CS, BVD and IBR from the citations given in Supplementary Table 1.

From the result of proteome-specific BLAST search for BoLA-A11 (3PWU_A) with the proteomes of considered organisms, 2 proteins of *H. sapiens i.e.*, MHC class I antigen and HLA B7 (CAX3389.1 and 3VCL_A), 2 proteins of *M. musculus i.e.,* MHC class I antigen precursor and H-2Kd (AAT06794.1 and 2FWO_A), one MHC class I antigen protein of *P. troglodytes* (ABG02214.1) and one Hb2m protein of *M. mulatta* (3JTS_A) were obtained. The result of MSA based phylogenetic analysis revealed H-2Kd of *M. musculus* as the closest homologue of BoLA-A11 of *B. taurus* in terms of MHC alleles (Fig. 1) with sequence identity of 74.9% and similarity of 84.4%. The result of structural super-imposition of BoLA-A11 and H-2Kd also revealed that both the proteins were closely related with each other. An RMSD of 1.888 Å was found between the structures of BoLA-A11 and H-2Kd proteins with a negligible variation in the antigen binding-groove (Fig. 2). Hence, the epitope prediction was done against H-2Kd of *M. musculus* as a proxy for the BoLA-A11 of *B. taurus* in MetaMHCI.

For each of the nonapeptides, fragmented from the antigenic protein sequences through a sliding window of length 9, different prediction scores were obtained using the predictors of MetaMHCI tool. The thresholds are reported for four individual predictors and one meta-predictor (Consensus). The threshold for ANN, SMM, NetMHC and NetMHCpan is less than 500, whereas for Consensus meta-predictor, it is greater than 0.9 (http://www.biokdd.fudan.edu.cn/Service/MetaMHCI/help.html). The total number of nonameric fragments scored were 43888 (Supplementary Dataset 2; http://cabgrid.res.in/ijbb/Suppl/), of which a total number of 78 nonapeptides (Supplementary Table 2; http://cabgrid.res.in/ijbb/Suppl/) were found to satisfy the thresholds of all five methods and a total of 4350 nonapeptides (including 78 nonapeptides of Supplementary Dataset 3; http://cabgrid.res.in/ijbb/Suppl/) satisfied the threshold score of at least one method.

The correlations obtained among the scores of 4 individual predictors and 3 meta-predictors are presented in the form of heat maps for four diseases in Fig. 3. It was observed from the Fig. 3 that there existed high degree of correlation between the predictors (the dark gray color represents highest correlation and white color represents lowest correlation). The correlation between PbES and the scores of individual predictor is presented in Table 2. From both Fig. 3 and Table 2, it was inferred that PbES was correlated with four individual predictors, as well as with the meta-predictors. However, the nature of linear relationship was positive in case of individual predictors and negative in case of meta-predictors.

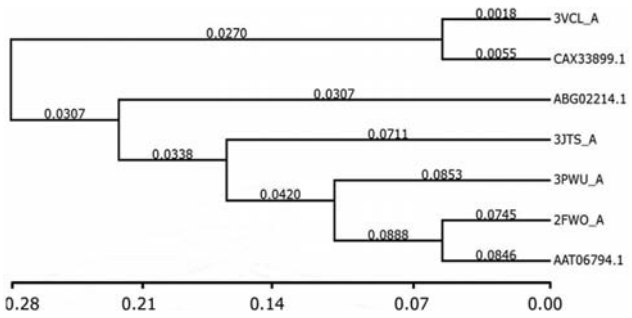


Fig. 1—Phylogenetic tree showing *M. muscullus* (2FWO_A) as the closest homologue of *B. taurus* (3PWU_A)

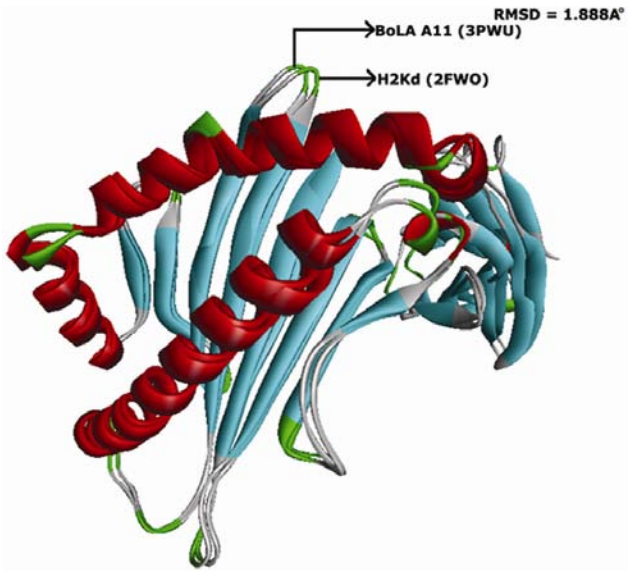


Fig. 2—Structural superimposition of BoLA A11 with H-2Kd [It shows very less RMSD (1.888A^0) implying that these two proteins are structurally very similar]

Table 2—Correlation of PbES with individual and meta- predictors under different disease categories

	Individual predictors				Meta predictors		
Disease	ANN	SMM	NetMHC	NetMHCpan	Consensus	PM	AvgTanh
FMD	0.861438	0.449872	0.857638	0.8301731	-0.80712	-0.88794	-0.9504
BVD	0.896585	0.429393	0.868535	0.8337876	-0.8053	-0.89561	-0.9546
IBR	0.880181	0.419838	0.863575	0.8382829	-0.79673	-0.87968	-0.95493
CS	0.901708	0.422924	0.875518	0.8372867	-0.79504	-0.8733	-0.95608

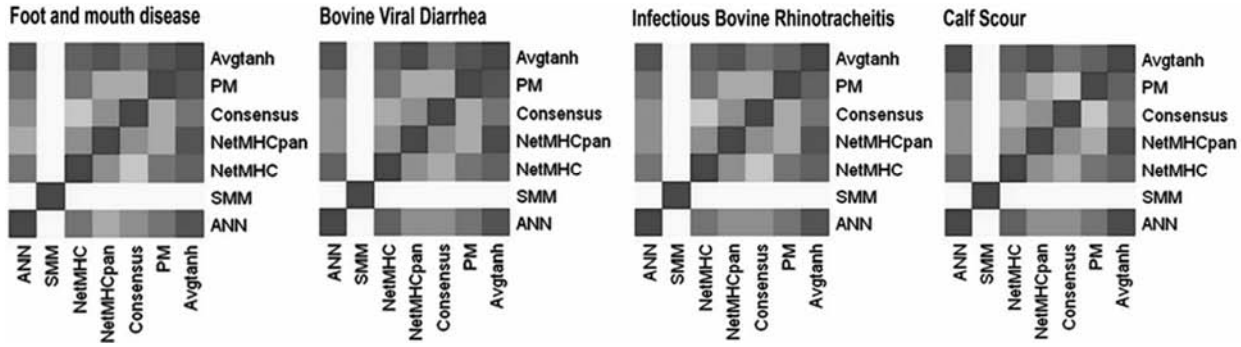
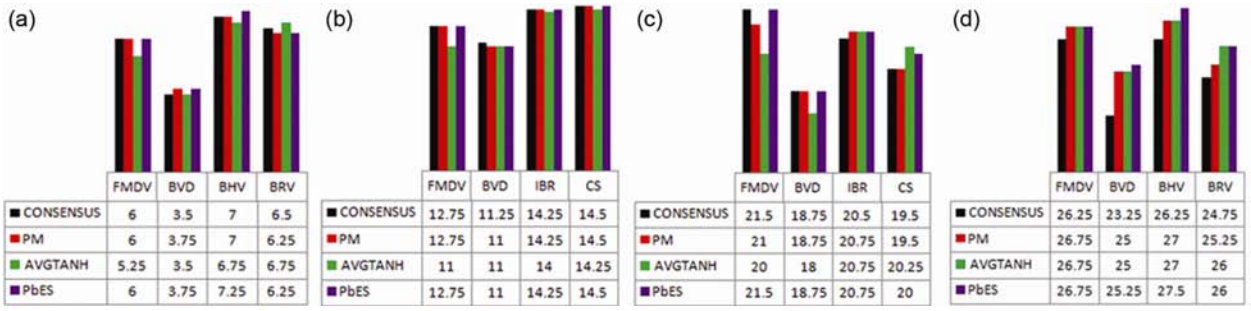


Fig. 3—Heat maps showing the correlations among the scores of 4 individual predictors and 3 meta-predictors [The dark color shows high correlation, whereas the light color shows low correlation]



(a)	FMDV	BVD	BHV	BRV
CONSENSUS	6	3.5	7	6.5
PM	6	3.75	7	6.25
AVGTANH	5.25	3.5	6.75	6.75
PbES	6	3.75	7.25	6.25

(b)	FMDV	BVD	IBR	CS
CONSENSUS	12.75	11.25	14.25	14.5
PM	12.75	11	14.25	14.5
AVGTANH	11	11	14	14.25
PbES	12.75	11	14.25	14.5

(c)	FMDV	BVD	IBR	CS
CONSENSUS	21.5	18.75	20.5	19.5
PM	21	18.75	20.75	19.5
AVGTANH	20	18	20.75	20.25
PbES	21.5	18.75	20.75	20

(d)	FMDV	BVD	BHV	BRV
CONSENSUS	26.25	23.25	26.25	24.75
PM	26.75	25	27	25.25
AVGTANH	26.75	25	27	26
PbES	26.75	25.25	27.5	26

Fig. 4—Ggraphical representation of the average number of epitopes predicted by each meta-predictor (including PbES) that are appearing in the top 10(5a), 20(5b), 30(5c) and 40(5d) epitopes predicted by individual predictors

The average number of epitopes predicted by each meta-predictor (including PbES) appearing in the top 10, 20, 30 and 40 results of individual predictors (ANN, SMM, NetMHC and NetMHCpan) for each disease are graphically shown in Fig. 4a-d. Each cell value in Fig. 4a indicates the average number of epitopes computed from intersections of top ten meta-predictor results with that of four individual predictors separately. Similarly, the averages are presented in Fig. 4b-d for top 20, 30 and 40 results of each meta-predictor, respectively. From Fig. 4a, it was observed that out of top 10 epitopes predicted by the individual predictors, the average number of common epitopes predicted through (i) PbES and PM were equal in case of three diseases, except in IBR, (ii) PbES was more than Consensus in case of BVD and IBR, and (iii) PbES was more than AvgTanh in case of three diseases, but slightly less in case of CS.

Figure 4b shows that out of top 20 epitopes predicted by the individual predictors, the average number of common epitopes predicted through (i) PbES and PM were equal in case of all the diseases, ii) PbES and Consensus were equal in three diseases, but Consensus was slightly more in BVD, and (iii) PbES was greater than or equal to AvgTanh in all diseases. From Fig. 4c, it was observed that out of top 30 epitopes predicted by the individual predictors, the average number of common epitopes predicted through PbES were greater than or equal to PM, Consensus and AvgTanh in all the diseases. Out of the top 40 epitopes (Fig. 4d) predicted by the individual predictors, the average number of common epitopes predicted through PbES, PM, Consensus and AvgTanh were equal in case FMDV and in all other diseases number of epitopes predicted through PbES were more than that of all other meta-predictors.

The potential epitopes for each disease selected for the epitope-MHC interaction analysis based on their predictability through maximum number of predictors and the PbES score were *KYSSAKHSL* from polyprotein

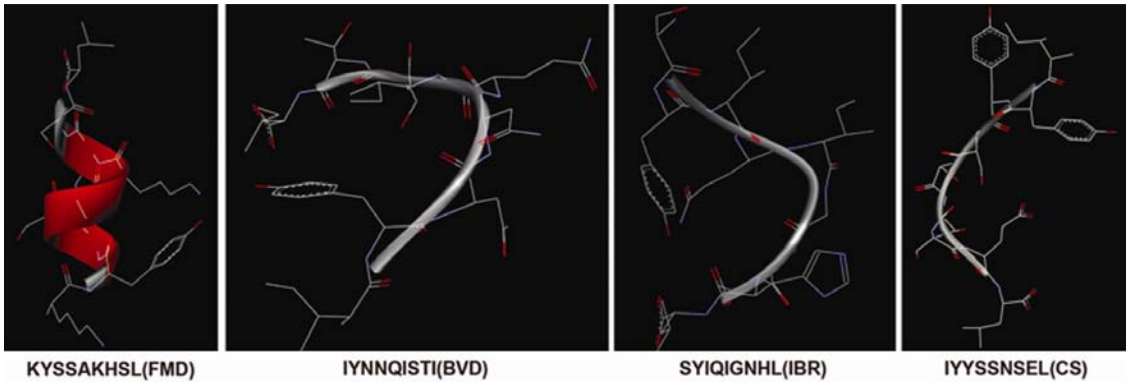


Fig. 5—Eenergy minimized structures of best epitopes for FMD, BVD, IBR and CS diseases of cattle

Table 3—Best epitope identified for each disease with sequence, number of methods predicted the epitope, PbES, results of Ramachandran plot for the loop refined structures, initial and final global energies and type of secondary structure

Disease	Epitope sequence	No. of methods where epitope appeared in top 10 results	PbES	Ramachandran plot		Global energy		Secondary structure
				Core (%)	Allowed (%)/ Generously allowed (%)/ Disallowed (%)	Initial (kJ/mol)	Final (kJ/mol)	
FMD	KYSSAKHSL	6/8	-4.69165	100	Nil	-3339.6	-5390.7	Helix
BVD	IYNNQISTI	6/8	-4.66919	100	Nil	-2476.2	-4146.6	Coil
IBR	SYIQIGNHL	8/8	-4.86339	100	Nil	-2407.6	-3734.7	Coil
CS	IYYSSNSEL	7/8	-4.85508	100	Nil	-2927.5	-3806.4	Coil

(CAB62902.1: 808~816) for FMD, *IYNNQISTI* from polyprotein (AFI81915.1: 3255~3263) for BVD, *IYYSSNSEL* from Dsrna-dependent polymerase, Vp1 (4AU6_E: 32~40) for CS and *SYIQIGNHL* from capsid triplex subunit 1 (NP_045320.1:9~17) for IBR. Top five structures for these submitted epitope sequences were obtained from the results of PEP-FOLD server. Table 3 presents the final Ramachandran plot scores of the loop refined structures with initial and minimum global energies. The energy minimized structures of all the four epitopes are presented in Fig. 5. The epitope structure for FMD was found helical, whereas others were found coiled.

Though few residues were found in the outlier region of the Ramachandran plot before loop refinement, all residues were observed in the core region for all the epitopes after loop refinement. Analysis of tertiary structure of BoLA A11 MHC allele of *B. taurus* (3PWU) revealed its interaction with one CTL epitope from rinderpest virus (Supplementary Fig. 1). The interacting residues observed in Supplementary Fig. 1 were used as *a priori* information for docking. The initial energy for BoLA-A11 (3PWU) was -185880.6 kJ/mol, but the final energy became -233893.7 KJ/mol (Supplementary Fig. 2; http://cabgrid.res.in:8080/ ijbb/Suppl/) after minimization. The 100 best results obtained from patch dock server, refined through Firedock server were found arranged based on the global energy in the result page. The interactions of 10 best Firedock results for each docked epitope are given in the Supplementary Table 3 (http://cabgrid.res.in/ijbb/Suppl/). Further, the analysis of docking solutions for 4 diseases revealed ten different amino acid residues of BoLA-A11 those participated in at least one solution from each of the four diseases (Table 4). The interacting residues of one selected Firedock solution for each epitope are presented in Table 5 and three- dimensional view of the binding of epitopes in MHC binding groove is shown in Fig. 6. All the epitopes were found half-buried in the MHC binding groove of BoLA-A11.

Table 4—Number of residues of BoLA A11 taking part in epitope-MHC interaction in the four diseases of cattle

Residues	FMD	BVD	IBR	CS	Total
ARG162	2	3	4	5	14
TYR98	3	2	3	6	14
ARG154	2	1	7	2	12
ARG154	2	1	7	2	12
GLN155	2	3	4	3	12
GLU96	2	1	2	6	11
TYR8	3	2	2	2	9
THR69	2	4	2	1	9
GLU151	1	3	2	1	7
ASP68	1	1	2	2	6

Table 5—Interacting residues and global energies of the selected Firedock solutions for the four diseases

Disease	Interacting MHC residue	Interacting MHC atom*	Interacting epitope residue	Interacting epitope atom*	No. of FireDock solutions	No. of total interactions	No. of reference residues involved	Global energy
IBR	TYR8	OH	LEU9	O	3	9	2	-46.48
IBR	TYR8	OH	LEU9	OXT	3			
IBR	ASN76	ND2	SER1	O	3			
IBR	TYR98	OH	LEU9	O	3			
IBR	ARG154	NE	GLN4	OE1	3			
IBR	ASN76	OD1	SER1	N	3			
IBR	GLU151	OE1	GLN4	N	3			
IBR	GLN155	OE1	GLN4	NE2	3			
IBR	ASP68	OD2	HIS8	ND1	3			
CS	ASN62	ND2	TYR2	OH	3	16	2	-56.50
CS	GLN155	NE2	TYR3	OH	3			
CS	ARG162	NE	TYR2	O	3			
CS	ARG162	NE	TYR3	O	3			
CS	ARG162	NH1	TYR3	O	3			
CS	ARG162	NH1	GLU8	OE2	3			
CS	ARG162	NH2	TYR3	O	3			
CS	ARG169	NH1	LEU9	O	3			
CS	ARG169	NH1	LEU9	OXT	3			
CS	TYR8	OH	ILE1	N	3			
CS	TYR98	OH	ILE1	N	3			
CS	ASN62	OD1	TYR2	OH	3			
CS	GLN155	OE1	TYR3	OH	3			
CS	GLU165	OE2	SER7	N	3			
CS	GLU165	OE1	GLU8	N	3			
CS	TYR158	CD1	TYR2	O	3			
BVD	ARG74	NH1	THR8	O	7	12	4	-29.00
BVD	ARG74	NH2	THR8	O	7			
BVD	ARG74	NH2	ILE9	O	7			
BVD	ARG74	NH2	ILE9	OXT	7			
BVD	THR79	OG1	GLN5	O	7			
BVD	TYR122	OH	ASN3	OD1	7			
BVD	THR142	OG1	ASN3	OD1	7			
BVD	LYS145	NZ	TYR2	O	7			
BVD	LYS145	NZ	ASN3	O	7			
BVD	TRP146	NE1	TYR2	O	7			
BVD	TYR83	OH	ASN4	N	7			
BVD	ALA75	O	SER7	OG	7			
FMD	TYR6	OH	LEU9	O	8	11	2	-20.81
FMD	TYR6	OH	LEU9	OXT	8			
FMD	ARG61	NE	HIS7	O	8			
FMD	THR69	OG1	SER4	OG	8			
FMD	TYR158	OH	LEU9	OXT	8			
FMD	ARG162	NE	LYS6	O	8			
FMD	ARG162	NE	SER8	O	8			
FMD	ARG162	NH2	LYS6	O	8			
FMD	ARG162	NH2	HIS7	O	8			
FMD	ILE65	O	SER4	OG	8			
FMD	THR69	OG1	SER4	OG	8			

FMD, foot and mouth disease; CS, calf scour; BVD, bovine viral diarrhea; IBR, infectious bovine rhinotracheitis

*The first character is the chemical symbol of the atom (usually C, N or O), the second character is the remoteness indicator and the third character is the branch designator.

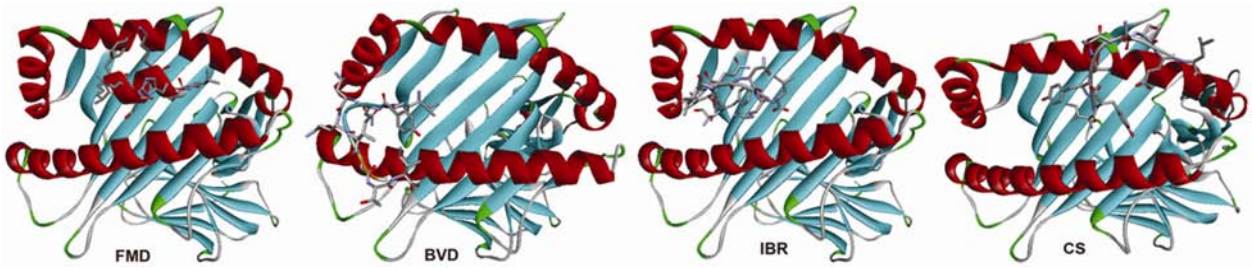


Fig. 6—Three-dimensional view of epitope-MHC interaction for FMD, BVD, IBR and CS diseases of cattle. The epitopes are bound to binding groove of the BoLA A11 MHC allele of cattle

Discussion

FMD, CS, BVD and IBR are the most important viral diseases, which cause low milk production and reduced growth of cattle resulting high economic loss. Thus, an attempt was made to identify potential epitopes and study their interaction with MHC alleles in *B. taurus*. For this, all the pathogenic sequences were downloaded and the sequences with more than 90% identity were removed due to the fact that almost all epitopes from identical sequences will be identical. The total number of sequences selected for epitope prediction belongs to different categories that are found to be related with the diseases considered under the present study (Supplementary Table 1).

The MetaMHCI is a widely used web-server that uses ensemble approaches to predict epitopes having binding affinity towards MHC class I proteins of *H. sapiens, M. musculus, P. troglodytes* and *R. macaque*. Thus, to predict epitopes for *B. taurus,* identification of most closely related species of *B. taurus* among the species listed in MetaMHCI was necessary. Accordingly, the closest homologue of *B. taurus* was identified as *M. musculus* through phylogenetic analysis (Fig. 1). Besides, the structure comparison between MHC class I proteins of *B. taurus* and *M. musculus* revealed an RMSD of less than 2A^{o}, confirming the high structural similarity between the MHC proteins of these two species. Also, a number of studies in relation to FMD[40], BVD[41] and IBR[42] were conducted on mouse species in the past. The rotavirus that causes CS was first isolated and characterized from mice[43]. This indicates the proximity of the mouse species with the cattle species. In the present study, MHC of cattle showed similarity with many MHC class I proteins of mouse, as evident from the results of BLAST. The first hit of BLAST was related to the "H-2Dr MHC allele", which is not available in MetaMHCI tool. However, the next known MHC allele available in MetaMHCI was found to be H-2Kd. Though the query coverage of H-2Kd was 3% lower than that of H-2Dr, the percentage of identity of former over latter was 1% higher. Hence, *M. musculus* was considered here as the source species and H-2Kd as the MHC allele in MetaMHCI tool for the prediction of probable epitopes.

MetaMHCI database includes seven types of predictors (both individual and meta-predictors) based on different algorithms. These predictors were found to be correlated with each other (Fig. 3), when analyzed using the dataset considered under the present study. Since different predictors ranked different epitopes based on their threshold scores and are correlated with each other, the proposed PbES that is a principal components analysis[45] based index score was used to rank the epitopes. The PbES was found to be positively correlated with all the individual predictors, whereas negatively correlated with all the Meta-predictors. The epitopes with low individual predictor scores were referred as good binders in MetaMHCI[44]. Hence a low PbES (proposed in this study) can be preferred for selecting the good binders.

The comparative analysis between the meta-predictors including PbES using the top 10, 20, 30 and 40 results showed that the good binders predicted through the individual predictors appeared in the top results of PbES. Probably, the PbES performed well due to its capability to capture the correlation structure existing among individual and meta-predictors and its nature as a weighted index, where eigen values are used as weights. As PbES is able to capture the good binders on a consensus basis, it can be used as a complementary procedure to the existing scoring methods for the selection of good binders.

The MetaMHCI server considers the epitopes as good binders when they score less than 500 for ANN, SMM, NetMHC and NetMHCPan. Besides, an epitope with score less than 0.9 obtained from Consensus meta-predictor is considered as a good

binder[44] in MetaMHCI server. However, the thresholds are not available for the rest of the meta-predictors. As 78 nonapeptides were found to satisfy the available thresholds, these were considered as strong binders. However, the nonapeptides satisfying at least one of these thresholds should not be rejected, but can be considered as weak binders. Further, the nomenclature of the epitopes as strong or weak binders can be decided by looking at their PbES with low or high values, respectively.

The structures of best epitopes for FMD, BVD, CS and IBR were predicted using *ab initio* method. The three-dimensional (3D) confirmation of the epitope considered for FMD was found to be a helix, while the epitopes considered for rest of the diseases were the coils. The loop refinement and validation results (Table 3) confirmed the 3D arrangement of the amino acids in the structures with correct Φ and Ψ angles. In Ramachandran plot, above 90% residues were found in the favored region with no residue in disallowed region for all the structures, signifying the reliability of the predicted structures. As the epitope structures were found closer to accuracy, the interaction analysis was carried out with these refined structures.

From the PatchDock result, 100 best solutions (docked poses of epitopes and MHC) were taken and refined with the Firedock server. The analysis of top ten solutions obtained from Firedock revealed that maximum of solutions showed good number of interactions with minimum global energies (negative), implying the stability of the solutions. However, out of the ten selected solutions, the best solution was chosen, based on global energy, number of interactions and number of reference residues involved in the interactions. As the reference residues were identified from 3PWU protein model (a crystallographic structure), their presence in the theoretical models was taken into consideration in the selection of a best solution. From the best solutions, it was found that the docked epitopes were well-fitted in the MHC binding groove of the BoLA A11 protein and hence were expected to be good binders. Thus, their recognition by the CTLs would help enable the immune elimination of FMDV, BVDV, BRV and BHV form the cattle. Similarly, for rest of the predicted epitopes, the docking studies could be followed to analyze and validate their interaction with the class I MHC molecule of *B. taurus.* In addition, ten amino acids of BoLA A11 protein were found to participate in at least one docking solution from each of the four diseases (Table 4). Thus, these amino acids might be the crucial residues of BoLA A11 that are expected to play an important role in antigen presentation.

Conclusion

Epitope-based vaccine designing has recently been attracting growing interest in the field of modern computational biology. However, the experimental discovery of candidate epitopes is expensive in terms of time and money. Thus, in this study, computational prediction of CD8+ T-Cell epitopes was done for vaccine development against the four major viral diseases *i.e.* FMD, CS, BVD and IBR. A total of 78 epitopes were identified and out of which 27 were for FMD, 25 for BVD, 12 for CS and 14 were for IBR. Besides, a set of 4272 epitopes (excluding the above 78) that satisfied the thresholds of at least one predictor score was filtered out. A PCA-based epitope score (PbES) was proposed using the scores of the existing predictors to simplify the selection of potential epitopes. The epitopes having low PbES could be recommended to vaccinate the cattle for FMD, CS, BVD and IBR viral diseases. Besides, the epitopes predicted here could be an addition to the existing epitope databases like immuno epitope database (IEDB) and the methodology adapted here could also be used to predict and analyze the epitopes for the viral diseases in other important animal species.

Acknowledgements

This study was supported by World Bank Funded National Agricultural Innovation Project (NAIP), ICAR Grants 30(68)/2009/Bio Informatics/NAIP/ O&M and NAIP/Comp-4/C4/C-30033/2008-09.

References

1 Brown C C, Meyer R F, Olander H J, House C & Mebus C A (1992) *Can J Vet Res* 56, 189-193
2 Fry E E, Newman J W, Curry S, Najjam S, Jackson T, Blakemore W, Lea S M, Miller L, Burman A, King A M & Stuart D I (2005) *J Gen Virol* 86, 1909-1920
3 Samuel A R & Knowles N J (2001) *J Gen Virol* 82, 609-621
4 Handel I G, Willoughby K, Land F, Koterwas B, Morgan K L, Tanya V N & Bronsvoort B M (2011) *PLoS ONE* 6(7) e21620. [doi:10.1371/journal.pone.0021620]
5 Weiskircher E, Aligo J, Ning G & Konan K V (2009) *Virol J* 6. [doi: 10.1186/1743-422X-6-185]
6 Lorrot M & Vasseur M (2007) *Virol J* 4, 31. [doi:10.1186/1743-422X-4-31]
7 Kahrs R F (1977) *J Am Vet Med Assoc* 171, 1055-1064.
8 Lovato T, Inman M, Henderson G, Doster A & Jones C (2003) *J Virol* 77, 4848-4857

9 Parham P (2009) *The Immune System*, Garland Science, New York
10 Yewdell J W & Bennink J R (1992) *Adv Immunol* 52, 1-123
11 Barber L D & Parham P P (1993) *Annu Rev Cell Biol* 9, 163-206
12 McMichael A (1992) *Cancer Surv* 13, 5-21
13 Toussaint N C & Kohlbacher O (2009) *Expert Opin Drug Discov* 4, 1047-1060
14 Khan A M, Miotto O, Heiny A T, Salmon J, Srinivasan K N, Nascimento E J, Marques E T, Jr Brusic V, Tan T W & August J T (2006) *Cellular Immunol* 244, 141-147
15 Karpenko O, Huang L & Dai Y (2008) *Immunogenetics* 60, 25-36
16 Bui H H, Sidney J, Li W, Fusseder N & Sette A (2007) *BMC Bioinformatics* 8, [doi:10.1186/1471-2105-8-361]
17 Davies C J, Joosten I, Bernoco D, Arriens M A, Bester J, Ceriotti G, Ellis S, Hensen E J., Hines H C, Horin P, Kristensen B, Lewin H A, Meggiolaro D, Morgan A L G, Morita M, Nilsson P R, Oliver R A, Orlova A, Ostergard H, Park C A, Schuberth H J, Simon M, Spooner R L & Stewart J A (1994) *Eur J Immunogenet* 21,239-258
18 Hegde N R, Ellis S A, Gaddum R M, Tregaskes C A, Sarath G & Srikumaran S (1995) *Immunogenetics* 42,302-303
19 Hegde N R & Srikumaran S (1996) *Virus Genes* 13, 121-133
20 Hegde N R & Srikumaran S (1997) *Virus Genes* 14, 111-121
21 Becker Y (1997) *Virus Genes* 14, 123-129.
22 Altschul S F, Gish W, Miller E W, Myers D J & Lipman (1990) *J Mol Bio* 215, 403-410
23 Lin H H, Ray S, Tongchusak S, Reinherz E L & Brusic V (2008) *BMC Immunol* 9. [doi: 10.1186/1471-2172-9-8]
24 Peters B, Bui H H, Frankild S, Nielson M, Lundegaard C, Kostem E, Basch D, Lamberth K, Harndahl M, Fleri W, Wilson S S, Sidney J, Lund O, Buus S & Sette A (2006) *PLoS Comput Biol* 2. [doi: 10.1371/journal.pcbi.0020065]
25 Bleek G M V & Nathenson S G (1991) *PNAS* 88, 11032-11036
26 Nielsen M, Lundegaard C, Worning P, Lauemaller S L, Lamberth K, Buus S, Brunak S & Lund O (2003) *Protein Sci* 12, 1007-1017
27 Peters B & Sette A (2005) *BMC Bioinformatics* 6. [doi: 10.1186/1471-2105-6-132]
28 Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O & Nielsen M (2008) *Nucleic Acids Res* 36, W509–W512
29 Hoof I, Peters B, Sidney J, Pedersen le, Sette A, Lund O, Buus S & Nielsen M (2009) *Immunogenetics* 61, 1-13
30 Moutaftsi M, Peters B, Pasquetto V, Tscharke D C, Sidney J, Bui H H, Grey H & Sette A (2006) *Nat Biotechnol* 24, 817-819
31 Jain A, Nandakumar K & Ross A (2005) *Pattern Recognition* 38, 2270-2285
32 Maupetit J, Derreumaux P & Tufféry P (2009) *Nucleic Acids Res* 37, W498-W503
33 Laskowski R A, Macarthur M W, Moss D S & Thornton M J (1993) *J Appl Cryst* 26, 283-291
34 Fiser A & Sali A (2003) *Bioinformatics* 19, 2500-2501
35 Ramachandran G N, Ramakrishnan C & Sasisekharan V (1963) *J Mol Biol* **7**, 95–99
36 Duhovny D, Nussinov R & Wolfson H J (2002) Efficient Unbound Docking of Rigid Molecules, In: *Algorithms in Bioinformatics, Lecture Notes in Computer Science* 2452 (Guigo R & Gusfield D, eds), pp. 185-200, Springer Berlin, Heidelberg
37 Schneidman-Duhovny D, Inbar Y, Nussinov R & Wolfson H J (2005) *Nucleic Acids Res* 33, W363-367
38 Mashiach E, Schneidman-Duhovny D, Andrusier N, Nussinov R & Wolfson H J (2008) *Nucleic Acids Res* 36,W229-W32
39 Dzyaloshinskii I E, Lifshitz E M & Pitaevskii L P (1961) *Sov Phys Uspekhi* 4, 153-176
40 Richmond J Y (1971) *Infect Immun* 3, 249-253
41 Weiland E, Thiel H J, Hess G & Weiland G (1998) *J Virol Meth* 24(1-2), 237-243
42 Forman A J, Babiuk L A, Misra V & Baldwin F (1982) *Infect Immun* 35, 1048-1057
43 Schroeder B A, Sproule R & Saywell D (1985) *Surveillance* 12, 2-3
44 http://www.biokdd.fudan.edu.cn/Service/MetaMHCI/help.html
45 Johnson R A & Wichern D W (2002) *Applied Multivariate Statistical Analysis*, Prentice Hall, USA