ORIGINAL RESEARCH

Antigenic heterogeneity of capsid protein VP1 in foot-and-mouth disease virus (FMDV) serotype Asia1

SM Sabbir Alam[1]
Ruhul Amin[1]
Mohammed Ziaur Rahman[2]
M Anwar Hossain[1]
Munawar Sultana[1]

[1]Department of Microbiology, University of Dhaka, Dhaka, Bangladesh; [2]International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh

Abstract: Foot and mouth disease virus (FMDV), with its seven serotypes, is a highly contagious virus infecting mainly cloven-hoofed animals. The serotype Asia1 occurs mainly in Asian regions. An in-silico approach was taken to reveal the antigenic heterogeneities within the capsid protein VP1 of Asia1. A total of 47 VP1 sequences of Asia1 isolates from different countries of South Asian regions were selected, retrieved from database, and were aligned. The structure of VP1 protein was modeled using a homology modeling approach. Several antigenic sites were identified and mapped onto the three-dimensional protein structure. Variations at these antigenic sites were analyzed by calculating the protein variability index and finding mutation combinations. The data suggested that vaccine escape mutants have derived from only few mutations at several antigenic sites. Five antigenic peptides have been identified as the least variable epitopes, with just fewer amino acid substitutions. Only a limited number of serotype Asia1 antigenic variants were found to be circulated within the South Asian region. This emphasizes a possibility of formulating synthetic vaccines for controlling foot-and-mouth disease by Asia1 serotypes.

Keywords: protein modeling, antigenic sites, sequence variation

Introduction

Foot and mouth disease (FMD) is a highly contagious, economically devastating epidemic disease of cloven-hoofed animals, affecting up to 70 domesticated and wild species.[1] The etiological agent, foot-and-mouth disease virus (FMDV), is a nonenveloped ribonucleic acid (RNA) virus belonging to the *Aphthovirus* genus within the family *Picornaviridae*.[2] The virus is composed of a positive-sense, single-stranded RNA genome of about 8.3 kb, enclosed within a protein capsid. This capsid is composed of 60 copies of four different structural polypeptides; the VP1, VP2, and VP3 proteins are surface exposed, while VP4 is entirely internal. The coding sequence for the VP1 protein has been extensively used for molecular epidemiological studies.[3–6]

VP1 is the major antigen of FMDV capsid protein that contains the major B-cell epitope, which is the major immunodominant epitope eliciting protective humoral immunity, and antigenic variants of FMDV can be screened by the presence of specific antibodies against VP1.[7,8]

The virus exists in seven distinct serotypes: O, A, C, Asia-1, and Southern African Territories (SAT) 1–3.[9,10] Among the seven serotypes, serotypes O, A, and C mainly occur in Europe, South America, Africa, and Asia; SAT 1, SAT 2, and SAT 3 are

Correspondence: Munawar Sultana
Department of Microbiology, University of Dhaka, Dhaka-1000, Bangladesh
Tel +880 155 363 6825
Email munawar@du.ac.bd; munamicro@yahoo.com

generally restricted to sub-Saharan Africa; and Asia1 occurs only in Asia.[11] In Bangladesh, the emerging FMDV serotypes are mainly O and A.[12]

The high rate of mutation, the quasispecies dynamics, and recombination are responsible for the wide antigenic diversity of the virus.[13,14] Genetic heterogeneity within FMDV may arise as a result of normal genetic drift[15] due to selection pressure[16,17] or as a result of recombination between different FMDV genomes.[13] Therefore, in addition to its serotypic diversity, the antigenic shift and drift, wide host range, and its virulence potential to infect animals with a small infectious dose make it difficult to prevent the disease by vaccination.[18–20] The study of antigenic variability is thus important for proper vaccine design and to prevent obstacles in control of the disease.[21]

Currently, the rapid expansion of biologic databases and computational programs facilitate more elaborate research in postgenomic sequence analysis, characterization of sequence variability, and especially, the mapping of epitopes and computer-aided rational vaccine design.[22,23] Nucleotide and protein sequences of VP1 from a number of FMDV strains/isolates are now available in public domain repositories. A wide variety of computational programs and algorithms have been recently developed in that can be effectively used for postgenomic sequence analysis and computer-aided conformational epitope prediction.[23–25] The time and effort required for screening potential epitopes can be reduced by up to 95% by using computational immunology methods.[26] The development of more effective and accurate procedures and algorithms in the last few years allows more accurate prediction of major histocompatibility complex (MHC) binding affinity. Many web-based tools are publicly used for predicting T-cell and B-cell epitopes from protein sequences.[27] For the induction of humoral- or cell-mediated immunity using synthetic peptides, understanding the nature of T and B cell epitopes may play an important role.[28–30]

In the present study, bioinformatic analyses of the FMDV major capsid protein VP1 of the serotype Asia1 were done to map conformational epitopes and to characterize antigenic variability. It is known that the patterns, which are apparently hidden at the sequence level, become evident when mapped into a structure. Experimentally derived three-dimensional (3D) structure data of the VP1 protein of the serotype Asia1 virus were not already available. Therefore, an attempt was made to predict its 3D structure, using a knowledge-based homology modeling approach. The structure of the VP1 protein of FMDV serotype Asia1 was modeled and used to

determine the conformational epitopes, and each epitopes were analyzed in the context of observed mutations.

Materials and methods
Compilation of datasets
This study was particularly focused on FMDV isolates of the South Asian regions. A total of 47 VP1 nucleotide and complementary protein sequence data were retrieved from the National Center for Biotechnology Information (NCBI) GenBank sequence database.[31] Sequence choices were based on independently originated isolates from different South Asian countries over the past 12 years.

Sequence alignment and calculation of variability index
Retrieved sequences were aligned using the EBI ClustalW program[32] and the Gonnet matrix. The Protein Variability Server (PVS) was used to calculate the protein variability index, using the Wu–Kabat variability coefficient.[33] The variability coefficient was computed using the following formula:

$$\text{Variability} = N*k/n, \tag{1}$$

where, N is the number of sequences in the alignment, k is the number of different amino acids at a given position, and n is the number of times that the most common amino acid at that position is present.[34] The protein variability index was used to determine whether the predicted epitopes were positioned in the least variable, moderately variable, or hypervariable regions. A consensus sequence was also derived from this alignment, using the PVS server. The consensus sequence utilized a VP1 protein sequence that was derived from the most common amino acid residues occurring at each position. This consensus sequence was later used for modeling the VP1 protein 3D structure.

Prediction and evaluation of the protein 3D model
As there was no 3D structure of FMDV Asia1 in the Protein Data Bank (PDB), the 3D structure of Asia1 was determined using a homology modeling method. A consensus sequence, derived from the PVS server from ClustalW alignment, was used to generate the protein 3D structure, using the SWISS-MODEL 3D server.[35] The VP1 protein structure (PDB id: 1fod) from the O serotype was selected as the template for maximum sequence identity and E value. The derived model was evaluated using a Ramachandran plot version

2.0 server.[36] A Ramachandran plot was used to check the stereochemical properties of the structure. For model comparison, the SuperPose server was used.[37] Using SuperPose, two VP1 structures were superimposed onto one another to find the divergent region, as the RMSD (root mean square deviation) value. For each loop region, a RMSD value was calculated.

Antigenic site detection

Different bioinformatic algorithm and computational tools were used to predict antigenic sequences. The Disco Tope −1.2[38] server and ElliPro[39] server with default parameter were used to predict antigenic fragments, using protein 3D structure. Disco Tope predicts discontinuous epitope, using protein 3D structural data. The method is based on amino acid statistics, spatial information, and surface accessibility from a compiled data set of discontinuous epitopes determined by X-ray crystallography of antibody/antigen protein complexes. Disco Tope detects 15.5% of residues located in discontinuous epitopes, with a specificity of 95%.[38] ElliPro, is a web tool that applies Thornton's method and, together with a residue clustering algorithm, the MODELLER program and the Jmol viewer, allows the prediction and visualization of antibody epitopes in a given protein sequence or structure. ElliPro has been tested on a benchmark dataset of discontinuous epitopes inferred from the 3D structures of antibody-protein complexes.[39]

The BCPREDS server 1.0 and BepiPred 1.0b server with default threshold were used to predict epitopes of at least 12 mer lengths from input amino acid sequences.[48,61] BCPREDS utilizes a novel method – string kernels – for predicting linear B-cell epitopes. String kernels[40–44] are a class of kernel methods that have been successfully used in many sequence classification tasks.[42–47] In these applications, a protein sequence is viewed as a string defined on a finite alphabet of 20 amino acids. Four string kernels were explored: spectrum,[42] mismatch,[43] local alignment,[44] and subsequence,[41] in predicting the linear B-cell epitopes. BepiPred predicts linear B-cell epitopes using a hidden Markov model. It uses three data sets of annotated linear B-cell epitopes. A data set was collected from the literature, another data set was extracted from the AntiJen database,[62] and a data set of the epitopes in the proteins of HIV was collected from the Los Alamos HIV database.[48] Results from all the servers (Disco Tope, ElliPro, BCPREDS, BepiPred) were compared and combined to predict the antigenic regions.

Antigenic heterogenicity analysis

The predicted epitopes were further checked for location in the 3D structure, presence of mutations among other isolates, and for possible combination of mutations. The average Wu–Kabat protein variability index[34] was calculated for each of the predicted sites.

Results
## Alignment of sequences and calculation of protein variability index

Retrieved sequences were aligned using the EBI ClustalW program.[32] Multiple sequence alignment of the VP1 sequences of Asia1 (shown in Figure 1) was done using a Gonnet matrix. A protein variability plot was derived from the PVS server, using Wu–Kabat as the variability coefficient.[33] (Figure 2). The consensus sequence (TTTTGESADPVTTTVENYG GETQTARRLHTDVAFVLDRFVKLTAPKNTQTLDL MQIPSHTLVGALLRSATYYFSDLEVALVHTG PVTWVPNGSPKDALDNQTNPTAYQKQPITRLA LPYTAPHRVLATVYNGKTTYGETTSRRGD MAALAQRLSGRLPTSFNYGAVKAETITELLIRMK RAETYCPRPLLALDTTQDRRKQEIIAPEKQMM) of the VP1 protein was derived from the multiple sequence alignment, using the PVS server.

Protein modeling and model evaluation

The modeling process involved several steps, such as target-template selection and alignment, model building, and model evaluation. The accuracy of the models increased along with the increase in the numbers of known protein structures and the improvement in protein model software.[49] Since the FMDV VP1 3D structure for serotype Asia1 was not found on a database and since there exists a substantial difference in the VP1 sequence and conformational epitopes among the different serotypes of FMDV,[13,14] it was necessary to model the FMDV Asia1 VP1 sequence prior to epitope design. The model was built using SWISS-MODEL (Figure 3A and B). The PDB id: 1fod (2.60 Å) was used as a template for the model build up. This was selected due to its relatively better sequence identity (72.857%) and highly significant E value (6.56e-78).[50] A Ramachandran plot was found for the 3D model (Figure 4), which showed that the quality of the designed model was acceptable. The Ramachandran plot showed that model had most residues in the most favorable region and had overall good quality.

Using ClustalW multiple sequence alignment and the PyMOL visualization tool, each loop of the model was
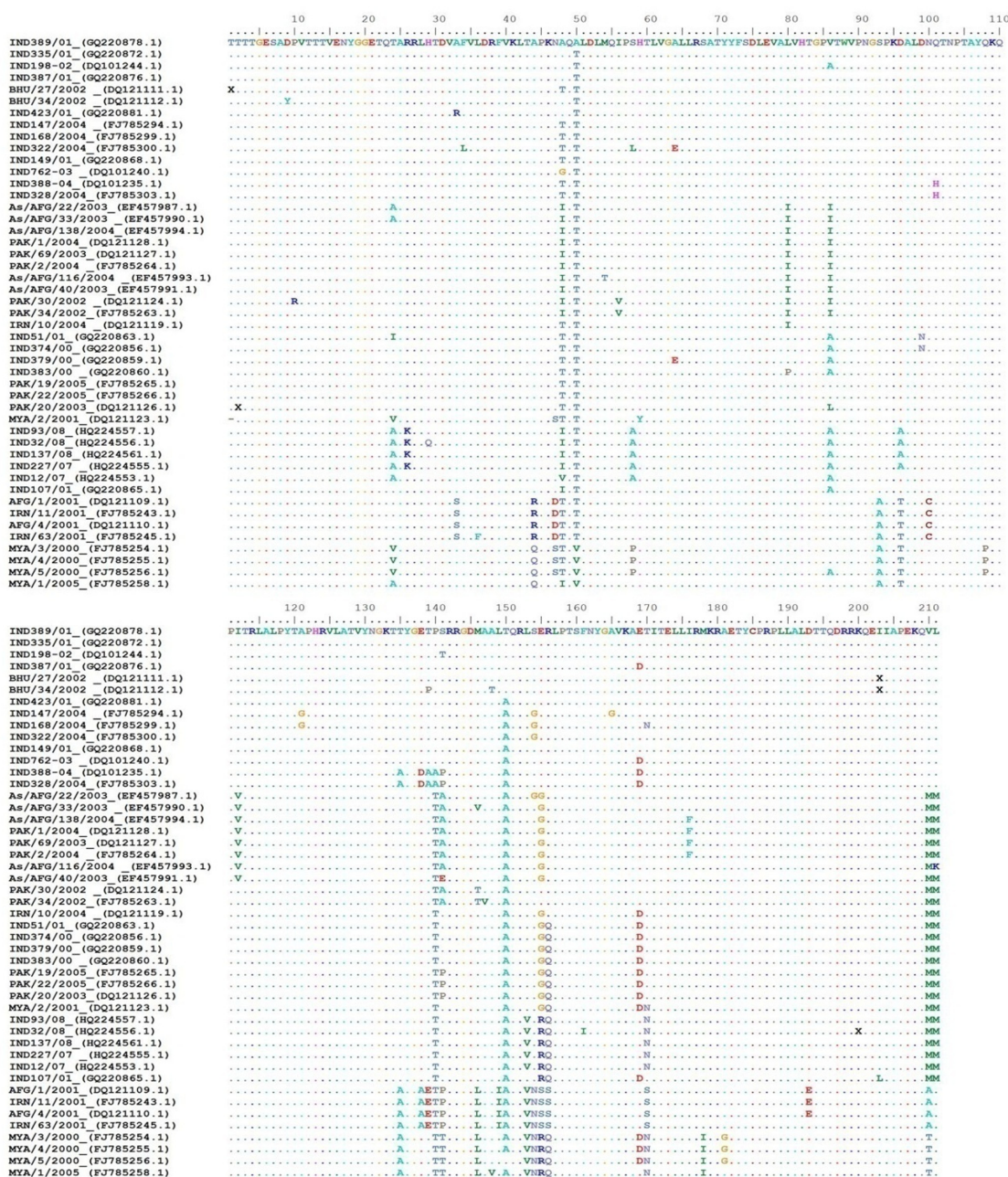
Figure 1 Multiple alignment of 47 VP1 sequences of the deduced 211 amino acids of Asia1 FMDV.
**Notes:** Dots represent the sequence identity with the consensus; amino acid changes relative to the consensus are indicated using single letter codes; asterisks represent missing amino acids; gaps and "x" in the sequence indicates region not sequenced.
Abbreviation: FMDV, foot-and-mouth disease virus.
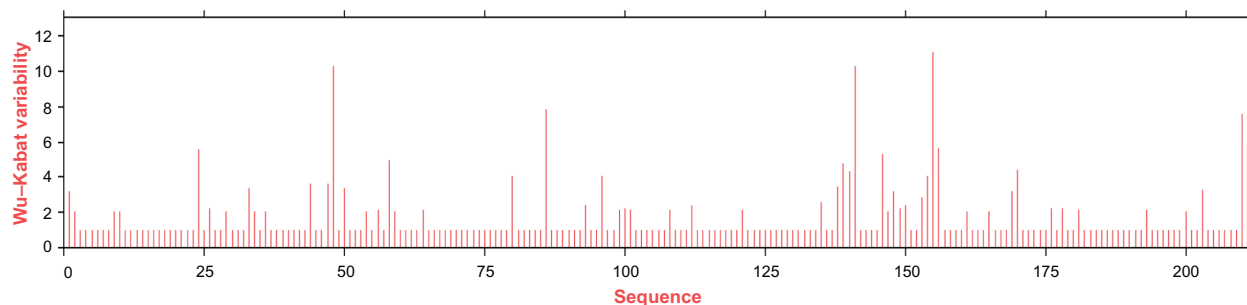
Figure 2 Wu–Kabat protein variability index for each of the 211 amino acid residues of VP1 protein of the FMDV serotype Asia1.
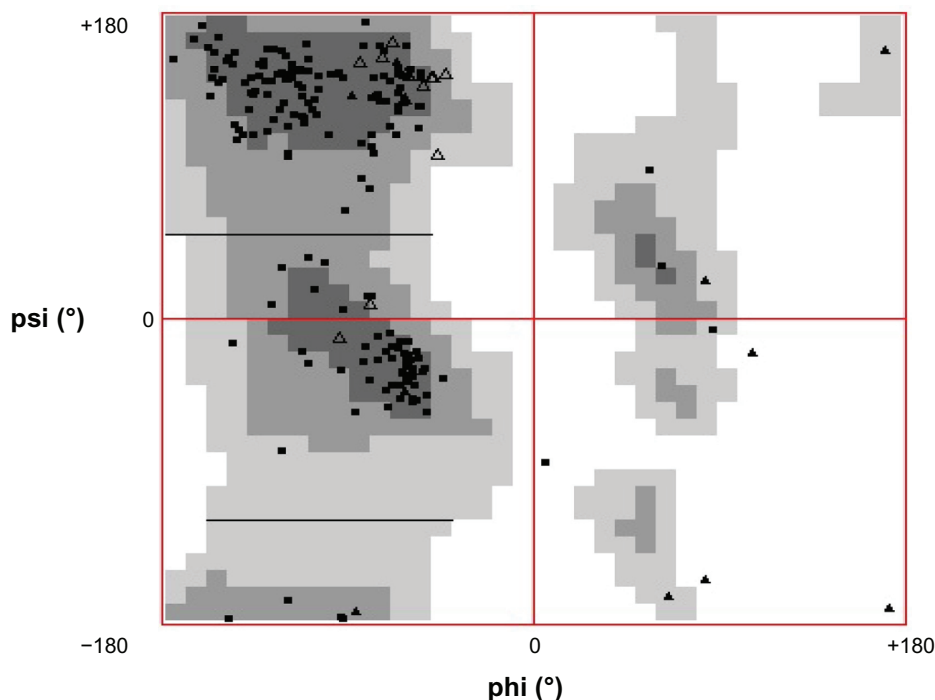**Abbreviation:** FMDV, foot-and-mouth disease virus.
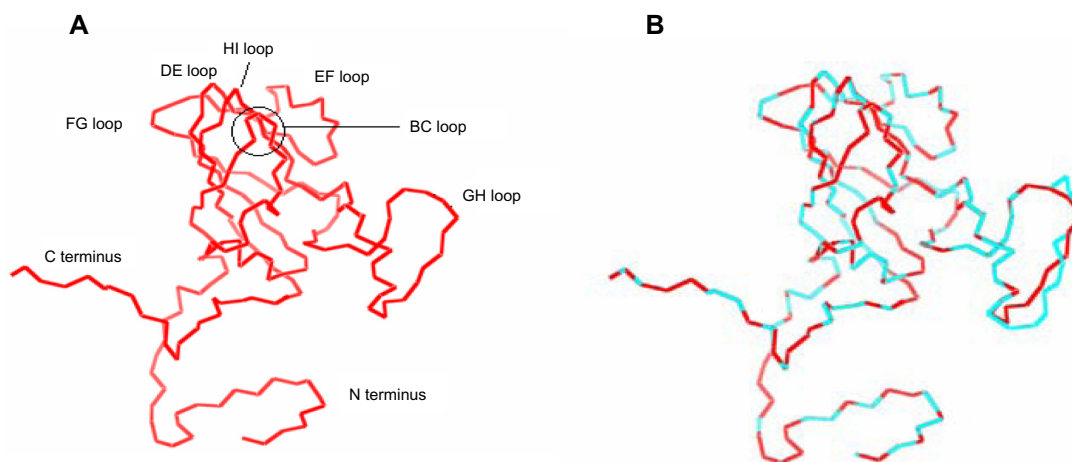
Figure 3 (**A**) Protein three-dimensional (3D) model of the VP1 protein of serotype Asia1, shown as a ribbon. Each of the important loops was detected using ClustalW alignment with the template VP1 structure and using the PyMOL visualization tool. Each loop is indicated in this figure. (**B**) Superimposed model and template (PDB id: 1fod). Here, the protein model is shown as red-colored, and the template shown as cyan-colored.

identified (Figure 3A). They were the C terminus, FG loop, DE loop, HI loop, EF loop, BC loop, GH loop, and the N terminus. To find structural distinction with the template 1fod, a RMSD value for the whole model and for all loops of the VP1 region was calculated, using the SuperPose server.[37] It was found that the RMSD value was the highest for the GH loop and the C terminus. The data are shown in Table 1.

Mapping of antigenic sites

B cell epitopes were predicted using both sequence data and 3D modeling of the FMDV Asia1 serotype. Several

epitopes were first predicted by BCPREDS and BepiPred, but only those epitopes that were fully or mostly overlapped with the Disco Tope and ElliPro prediction were chosen as the final epitope candidates. Thus, six predicted B-cell epitopes were common to all four programs (Table 2). It was found that among the six predicted epitopes, two epitopes were located in the N terminus, one in the EF loop, one between the EF and FG loops, one in the GH loop, and one in the C terminus. The predicted epitopes were mapped onto a protein 3D structure using PyMOL, as shown in Figure 5.

Figure 4 Ramachandran plot for the VP1 model of FMDV serotype Asia1, showing that most residues were located in favorable regions.
Abbreviation: FMDV, foot-and-mouth disease virus.

Table 1 Comparison of RMSD values of type O and type Asia1, for different positions

Region/ loop	Residues	RMSD value (A°)		
		Alpha carbons	**Back bone**	**All**
VP1 model	All	0.77 (208 atoms)	0.77 (832 atoms)	0.95 (1,501 atoms)
GH loop	136–150	3.53 (15 atoms)	3.38 (60 atoms)	4.76 (86 atoms)
BC loop	45–65	0.10 (21 atoms)	0.12 (84 atoms)	0.35 (146 atoms)
FG loop	103–113	0.07 (11 atoms)	0.11 (44 atoms)	0.33 (80 atoms)
EF loop	87–102	0.10 (16 atoms)	0.12 (64 atoms)	0.22 (113 atoms)
C terminus	184–210	5.27 (25 atoms)	5.07 (100 atoms)	5.72 (147 atoms)
N terminus	1–32	0.08 (32 atoms)	0.11 (128 atoms)	0.34 (231 atoms)

Abbreviation: RMSD, root mean square deviation.

Variability study

For each of the predicted antigenic peptides, the average variability index was calculated, mutation types were identified, and all possible combinations of mutations were identified from multiple sequence alignment (Table 3). From this data, it was found that, epitope $_{132}$(GKTTYGETTSRR)$_{143}$ located in GH loop region showed a higher variability index than did the other regions. A very minimal number of mutation combinations were found in the terminal regions of N and C. Three antigenic peptides predicted at these two regions, ($_1$(TTTTGESADPVT)$_{12}$, $_{10}$(PVTTTVENYGGE)$_{21}$, and $_{194}$(TTQDRRKQEIIA)$_{205}$), showed only two mutation combinations. Here, for position 135, the T→A substitution was found to occur at higher rates (ten times or 21.27% cases), and for position 140, the T→P substitution was also found to occur at elevated rates (12 times or 25.53% cases). This indicates a mutation proneness of specific bases at these two sites. Significant mutation proneness was also found for position 93 (occurred at S93A eight times), 96 (occurred at D96T eight times), 141 (occurred at S141P nine times) and 141 (occurred at S141A nine times). This mutation proneness of specific amino acid bases lowered the antigenic variability towards a minimum mutation combination, which could otherwise be more variable.

Discussion

In this work, VP1 protein sequences of the FMDV serotype Asia1 of South Asian origin were targeted and analyzed for an antigenic variability study. Several prediction methods were used to predict the antigenic sites. Both sequence-based and structure-based algorithm methods of epitope prediction were applied to identify the antigenic sites of VP1 protein. As this study combines the results from four different epitope prediction servers, the results were homologous and were expected to be correct. In a previous study of FMDV serotype Asia1,[51] experiments using Western blot and enzyme-linked immunosorbent assay (ELISA) demonstrated three potential B-cell epitopes located in VP1 region. It was found that two of our predicted B-cell epitopes, (VP1 $_{16}$ENYGGETQSARR$_{28}$) and (VP1$_{193}$TTHDRRKQEIIA$_{205}$), were located in regions that were significantly antigenic, as determined by Zhang et al,[51] which validated this prediction. The present study also revealed two additional antigenic sites that were predictably antigenic located on the EF loop and between the EF and FG loop regions. Previous studies have showed that major antigenic sites are located on the GH loop.[52,53] This study also showed epitope $_{132}$(GKTTYGETTSRR)$_{143}$ to be located in the GH loop, but here variability was comparatively higher than for the others. Five other epitopes located in the N terminus, C terminus, EF loop, and between the EF and FG loops had lower antigenic variability and are therefore suggested for consideration in vaccine design. These had only a limited number of mutations, and these mutations were biased towards specific amino acid bases. This bias limits their antigenic variability, which could otherwise be more divergent.

Most studies illustrate that FMDV is a highly variable single-stranded RNA virus[13,14,18] that is genetically very diverse, but its variability can be lower than expected and vac-

Table 2 Predicted antigenic sites with prediction score and location in the three-dimensional (3D) structure

Sequence	Position	Location in loop region	BCPREDS score	ElliPro score	Disco Tope score (average)	BepiPred score (average)
TTTTGESADPVT	1–12	N terminus	0.866	0.830	−5.83	1.67
PVTTTVENYGGE	10–21	N terminus	0.515	0.830	−5.23	1.25
SPKDALDNQTNP	93–104	EF loop	0.542	0.642	−4.74	1.61
LDNQTNPTAYQK	98–109	EF, and FG loop	0.83	0.642	−5.753	1.44
GKTTYGETTSRR	132–143	GH loop	0.441	0.726	−4.14	1.23
TTQDRRKQEIIA	194–205	C terminus	0.95	0.810	−1.85	0.70
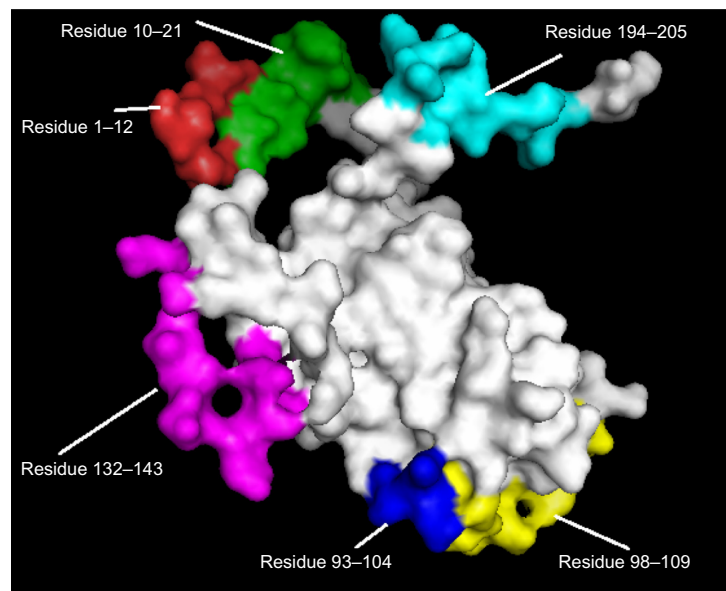
Figure 5 Predicted antigenic sites mapped onto the protein three-dimensional (3D) structure, using the PyMOL visualization tool.

Table 3 Average Wu–Kabat protein variability index, mutation type, and mutation combinations found for each predicted epitope

Peptide	Variability index average	Mutation type	Mutation combinations
$_1$(TTTTGESADPVT)$_{12}$	1.27	D9Y	Two combinations: $_1$(TTTTGESADPVT)$_{12}$ $_1$(TTTTGESA<u>Y</u>PVT)$_{12}$
$_{10}$(PVTTTVENYGGE)$_{21}$	1.08	D9Y, P10R	Two combinations: $_{10}$(PVTTTVENYGGE)$_{21}$ $_{10}$(<u>R</u>VTTTVENYGGE)$_{21}$
$_{93}$(SPKDALDNQTNP)$_{104}$	1.65	S93A*8, D96A*4, D96T*8, D99N*2, N100C*4, Q101H*2	Six combinations: $_{93}$(SPKDALDNQTNP)$_{104}$ $_{93}$(SPK<u>A</u>ALDNQTNP)$_{104}$ $_{93}$(SPKDALDN<u>H</u>TNP)$_{104}$ $_{93}$(SPKDAL<u>N</u>NQTNP)$_{104}$ $_{93}$(<u>A</u>PK<u>T</u>ALD<u>C</u>QTNP)$_{104}$ $_{93}$(<u>A</u>PK<u>T</u>ALDNQTNP)$_{104}$
$_{98}$(LDNQTNPTAYQK)$_{109}$	1.37	D99N*2, N100C*4, Q101H*2, Q108P*3	Five combinations: $_{98}$(LDNQTNPTAYQK)$_{109}$ $_{98}$(LDN<u>H</u>TNPTAYQK)$_{109}$ $_{98}$(L<u>N</u>NQTNPTAYQK)$_{109}$ $_{98}$(LD<u>C</u>QTNPTAYQK)$_{109}$ $_{98}$(LDNQTNPTAY<u>P</u>K)$_{109}$
$_{132}$(GKTTYGETTSRR)$_{143}$	2.68 (hypervariable)	T135A*10,E138D*2, E138A*4, T139P, T139A*2, T139E*4 T140P*12, T140A*2 S141T*5, S141P*9, S141A*9, S141E	Ten combinations: $_{132}$(GKTTYGETTSRR)$_{143}$ $_{132}$(GKTTYGET<u>P</u>SRR)$_{143}$ $_{132}$(GKTTYGE<u>PP</u>SRR)$_{143}$ $_{132}$(GKTTYGET<u>P</u>TRR)$_{143}$ $_{132}$(GKT<u>A</u>YG<u>DAA</u>PRR)$_{143}$ $_{132}$(GKTTYGETT<u>A</u>RR)$_{143}$ $_{132}$(GKTTYGETT<u>E</u>RR)$_{143}$ $_{132}$(GKTTYGETT<u>P</u>RR)$_{143}$ $_{132}$(GKT<u>A</u>YG<u>A</u>ET<u>P</u>RR)$_{143}$ $_{132}$(GKT<u>A</u>YGETT<u>T</u>RR)$_{143}$
$_{194}$(TTQDRRKQEIIA)$_{205}$	1.27	I203L	Two combinations: $_{194}$(TTQDRRKQEIIA)$_{205}$ $_{194}$(TTQDRRKQE<u>L</u>IA)$_{205}$

Note: *Number indicates mutation occurring by times for that position.

cine escape mutants can arise through only limited sequence variations at several antigenic sites.[54,55] Our study supports this hypothesis.

Previous studies have shown that synthetic peptides can induce antibodies reactive to the cognate sequences within the associated proteins.[56,63,64] Synthetic peptides that can be used as antigens are generally available through chemical synthesis. These peptides can produce potential immunogenic responses.[56–58] Conventionally, these peptides are conjugated with known protein or synthetic polymer carriers and are administered to laboratory animals to produce antibodies against these synthetic peptides.[56,57] The use of a carrier is sometimes avoided through use of polymers of synthetic peptide antigens.[58] A similar system, known as a multiple antigen peptide system (MAP), has been developed and used for the preparation of antipeptide antibodies and synthetic vaccines. The MAP system uses a small peptidyl core matrix bearing radially branching synthetic peptides as dendritic arms.[59,60] For the FMDV VP1 protein, a convenient polypeptide could be designed, utilizing a core matrix of a heptalysine containing eight dendritic arms of 12-residue peptides in length.[59]

The predicted antigenic sites and mutation combinations that were identified through this study may be considered in the design of MAPs, in which each mutation combination is present at least once.

Conclusion

This study showed that in South Asian regions, most antigenic variations of the FMDV serotype Asia1 come from very few mutations at several antigenic sites. Using computational predictive methods, six epitopes were predicted to be present within the VP1 region. Among these, five epitopes were found to be less variable, with only few mutations. This underlines the possibility of devising better synthetic vaccines for controlling FMD in that region, by considering these mutation combination types and the need of further synthetic vaccine study.

Acknowledgment

Disclosure

The authors report no conflicts of interest in this work.

References

1. Hedger RS. Foot-and-mouth disease. In: Davis JW, Karstad LH, Trainer DO, editors. *Infectious Diseases of Wild Mammals.* 2nd ed. Ames, USA: Iowa State University Press. 1981:87–96.
2. Belsham GJ. Translation and replication of FMDV RNA. *Curr Top Microbiol Immunol*. 2005;288:43–70.
3. Beck E, Strohmaier K. Subtyping of European foot-and-mouth disease virus strains by nucleotide sequence determination. *J Virol*. 1987;61(5):1621–1629.
4. Tosh C, Sanyal A, Hemadri D, Venkataramanan R. Phylogenetic analysis of serotype A foot-and-mouth disease virus isolated in India between 1977 and 2000. *Arch Virol*. 2002;147(3):493–513.
5. Knowles NJ, Samuel AR. Molecular epidemiology of foot-and-mouth disease virus. *Virus Res*. 2003;91(1):65–80.
6. Knowles NJ, Samuel AR, Davies PR, Midgley RJ, Valarcher JF. Pandemic strain of foot-and-mouth disease virus serotype O. *Emerging Infect Dis*. 2005;11(12):1887–1893.
7. Gurumurthy CB, Sanyal A, Venkataramanan R, Tosh C, George M, Hemadri D. Genetic diversity in the VP1 gene of foot-and-mouth disease virus serotype Asia 1. *Arch Virol*. 2002;147(1):85–102.
8. Brown F. Antibody recognition and neutralization of foot-and-mouth disease virus. *Semin Virol*. 1995;6(4):243–248.
9. Bachrach HL. Foot-and-mouth disease. *Annu Rev Microbiol*. 1968;22:201–244.
10. Domingo E, Escarmis C, Baranowski E, et al. Evolution of foot-and-mouth disease virus. *Virus Res*. Jan 2003;91(1):47–63.
11. Ansell DM, Samuel AR, Carpenter WC, Knowles NJ. Genetic relationships between foot-and-mouth disease type Asia 1 viruses. *Epidemiol Infect*. 1994;112(1):213–224.
12. Nandi SP, Rahman MZ, Momtaz S, Sultana M, Hossain MA. Emergence and distribution of Foot-and-Mouth Disease virus serotype A and O in Bangladesh. *Transbound Emerg Dis*. Epub June 4, 2013.
13. Carrillo C, Tulman ER, Delhon G, et al. Comparative genomics of foot-and-mouth disease virus. *J Virol*. 2005;79(10):6487–6504.
14. Domingo E, Mateu MG, Escarmis C, et al. Molecular evolution of aphthoviruses. *Virus Genes*. 1995;11(2–3):197–207.
15. Dopazo J, Sobrino F, Palma EL, Domingo E, Moya A. Gene encoding capsid protein VP1 of foot-and-mouth disease virus: a quasispecies model of molecular evolution. *Proc Natl Acad Sci U S A*. 1988;85(18):6811–6815.
16. Haydon DT, Bastos AD, Knowles NJ, Samuel AR. Evidence for positive selection in foot-and-mouth disease virus capsid genes from field isolates. *Genetics*. 2001;157(1):7–15.
17. Tully DC, Fares MA. Shifts in the selection-drift balance drive the evolution and epidemiology of foot-and-mouth disease virus. *J Virol*. 2009;83(2):781–790.
18. Kitching RP. Global epidemiology and prospects for control of foot-and-mouth disease. *Curr Top Microbiol Immunol*. 2005;288:133–148.
19. Domingo E, Escarmis C, Martinez MA, Martinez-Salas E, Mateu MG. Foot-and-mouth disease virus populations are quasispecies. *Curr Top Microbiol Immunol*. 1992;176:33–47.
20. Domingo E, Mateu MG, Martinez MA, Dopazo J, Moya A, Sobrino F. Genetic variability and antigenic diversity of foot-and-mouth disease virus. *J Applied Virology Research*. 1990;2:233–266.
21. Longjam N, Tayo T. Antigenic variation of Foot and Mouth Disease Virus – An Overview. *Vet World*. 2011;4(10):475–479.
22. Sobolev BN, Poroĭkov VV, Olenina LV, Kolesanova EF, Archakov AI. Computer-assisted vaccine design. *Biomed Khim*. 2003;49(4):309–332. Russian.
23. De Groot AS, Sbai H, Aubin CS, McMurry J, Martin W. Immunoinformatics: Mining genomes for vaccine components. *Immunol Cell Biol*. 2002;80(3):255–269.
24. De Groot AS, Martin W. From immunome to vaccine: epitope mapping and vaccine design tools. *Novartis Found Symp*. 2003;254:57–72; discussion 72–76, 98.

25. De Groot AS, Rappuoli R. Genome-derived vaccines. *Expert Rev Vaccines*. 2004;3(1):59–76.

26. De Groot AS, Bosma A, Chinai N, et al. From genome to vaccine: in silico predictions, ex vivo verification. *Vaccine*. 2001;19(31):4385–4395.

27. Khan AM, Heiny AT, Lee KX, et al. Large-scale analysis of antigenic diversity of T-cell epitopes in dengue virus. *BMC Bioinformatics*. 2006;7 Suppl 5:S4.

28. Singh H, Raghava GP. ProPred: prediction of HLA-DR binding sites. *Bioinformatics*. 2001;17(12):1236–1237.

29. Singh H, Raghava GP. ProPred1: prediction of promiscuous MHC Class-I binding sites. *Bioinformatics*. 2003;19(8):1009–1014.

30. Lata S, Bhasin M, Raghava GP. MHCBN 4.0: A database of MHC/TAP binding peptides and T-cell epitopes. *BMC Res Notes*. 2009;2:61.

31. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. *Nucleic Acids Res*. 2008;36(Database issue):D25–D30.

32. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994;22(22):4673–4680.

33. Garcia-Boronat M, Diez-Rivero CM, Reinherz EL, Reche PA. PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic Acids Res*. 2008;36(Web Server issue):W35–W41.

34. Kabat EA, Wu TT, Bilofsky H. Unusual distributions of amino acids in complementarity-determining (hypervariable) segments of heavy and light chains of immunoglobulins and their possible roles in specificity of antibody-combining sites. *J Biol Chem*. 1977;252(19):6609–6616.

35. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res*. 2003;31(13):3381–3385.

36. Gopalakrishnan K, Sowmiya G, Sheik SS, Sekar K. Ramachandran plot on the web (2.0). *Protein Pept Lett*. 2007;14(7):669–671.

37. Maiti R, Van Domselaar GH, Zhang H, Wishart DS. SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res*. 2004;32(Web Server issue):W590–W594.

38. Haste Andersen P, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci*. 2006;15(11):2558–2567.

39. Ponomarenko J, Bui HH, Li W, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*. 2008;9:514.

40. Haussler D. *Convolution Kernels on Discrete Structures*. UCSC-CRL-99-10 [Technical Report]. Santa Cruz, CA: University of California at Santa Cruz; 1999.

41. Lodhi H, Saunders C, Shawe-Taylor J, Cristianini N, Watkins C. Text classification using string kernels. *J Mach Learn Res*. 2002;2:419–444.

42. Leslie C, Eskin E, Noble WS. The spectrum kernel: a string kernel for SVM protein classification. *Pac Symp Biocomput*. 2002:564–575.

43. Leslie CS, Eskin E, Cohen A, Weston J, Noble WS. Mismatch string kernels for discriminative protein classification. *Bioinformatics*. 2004;20(4):467–476.

44. Saigo H, Vert JP, Ueda N, Akutsu T. Protein homology detection using string alignment kernels. *Bioinformatics*. 2004;20(11):1682–1689.

45. Zaki NM, Deris S, Illias R. Application of string kernels in protein sequence classification. *Appl Bioinformatics*. 2005;4(1):45–52.

46. Rangwala H, DeRonne K, Karypis G. Protein structure prediction using string kernels. In: Hu X, Pan Y, editors. *Knowledge Discovery in Bioinformatics: Techniques, Methods, and Applications*. Hoboken, NJ: John Wiley & Sons, Inc; 2007:145–168.

47. Wu F, Olson B, Dobbs D, Honavar V. Comparing kernels for predicting protein binding sites from amino acid sequence. *Proceedings of the International Joint Conference on Neural Networks (IJCNN06)*, Vancouver, BC, Canada; July 16–21, 2006. New York, NY: IEEE; 2006:1612–1616.

48. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Res*. 2006;2:2.

49. Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*. 2000;29:291–325.

50. Logan D, Abu-Ghazaleh R, Blakemore W, et al. Structure of a major immunogenic site on foot-and-mouth disease virus. *Nature*. 1993;362(6420):566–568.

51. Zhang ZW, Zhang YG, Wang YL, et al. Screening and identification of B cell epitopes of structural proteins of foot-and-mouth disease virus serotype Asia1. *Vet Microbiol*. 2010;140(1–2):25–33.

52. Höhlich BJ, Wiesmüller KH, Haas B, et al. Induction of an antigen-specific immune response and partial protection of cattle against challenge infection with foot-and-mouth disease virus (FMDV) after lipopeptide vaccination with FMDV-specific B-cell epitopes. *J Gen Virol*. 2003;84(Pt 12):3315–3324.

53. Höhlich BJ, Wiesmüller KH, Schlapp T, Haas B, Pfaff E, Saalmüller A. Identification of foot-and-mouth disease virus-specific linear B-cell epitopes to differentiate between infected and vaccinated cattle. *J Virol*. 2003;77(16):8633–8639.

54. Martínez MA, Dopazo J, Hernández J, et al. Evolution of the capsid protein genes of foot-and-mouth disease virus: antigenic variation without accumulation of amino acid substitutions over six decades. *J Virol*. 1992;66(6):3557–3565.

55. Mateu MG, Hernández J, Martínez MA, et al. Antigenic heterogeneity of a foot-and-mouth disease virus serotype in the field is mediated by very limited sequence variation at several antigenic sites. *J Virol*. 1994;68(3):1407–1417.

56. Lerner RA. Tapping the immunological repertoire to produce antibodies of predetermined specificity. *Nature*. 1982;299(5884):593–596.

57. Bittle JL, Houghten RA, Alexander H, et al. Protection against foot-and-mouth disease by immunization with a chemically synthesized peptide predicted from the viral nucleotide sequence. *Nature*. 1982;298(5869):30–33.

58. DiMarchi R, Brooke G, Gale C, Cracknell V, Doel T, Mowat N. Protection of cattle against foot-and-mouth disease by a synthetic peptide. *Science*. 1986;232(4750):639–641.

59. Merrifield RB. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J Am Chem Soc*. 1963;85(14):2149–2154.

60. Tam JP. Synthetic peptide vaccine design: synthesis and properties of a high-density multiple antigenic peptide system. *Proc Natl Acad Sci U S A*. 1988;85(15):5409–5413.

61. El-Manzalawy Y, Dobbs D, Honavar V. Predicting linear B-cell epitopes using string kernels. J Mol Recognit. 2008;21(4):243-255.

62. Toseland CP, Clayton DJ, McSparron H, et al. AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. Immunome research. 2005;1(1):4.

63. Tam JP. Synthetic peptide vaccine design: synthesis and properties of a high-density multiple antigenic peptide system. Proceedings of the National Academy of Sciences of the United States of America. 1988;85(15):5409-5413.

64. Grant GA. Synthetic peptides for production of antibodies that recognize intact proteins. Current protocols in immunology. Edited by John E. Coligan at al. 2003;Chapter 9:Unit 9 2.