Archives of Razi Institute, Vol. 73, No. 3 (2018) 183-198

Copyright © 2018 by
Razi Vaccine & Serum Research Institute

Original Article

Sequencing and In Silico Multi-aspect Analysis of S1 Glycoprotein in 793/B Serotype of Infectious Bronchitis Virus Isolated from Iran in 2003 and 2011

Vasfi Marandi [1], M., Malekan [1,*], M., Ranjbar [2], M.M., Dadashpour Davachi [3], N., Alamian [4], S.

1. Department of Poultry Diseases, Faculty of Veterinary Medicine, University of Tehran, Tehran, Iran
2. Department of Animal Virology, Research and Diagnosis, Razi Vaccine and Serum Research Institute, Agricultural Research, Education, and Extension Organization(AREEO), Karaj, Iran
3. Department of Research, Breeding and Production of Laboratory Animals, Razi Vaccine and Serum Research Institute, Agricultural Research, Education, and Extension Organization(AREEO), Karaj, Iran
4. Department of Brucellosis, Razi Vaccine and Serum Research Institute, Agricultural Research, Education, and Extension Organization(AREEO), Karaj, Iran

Received 19 January 2018; Accepted 26 February 2018
Corresponding Author: dr.m.malekan@gmail.com

ABSTRACT

Infectious bronchitis (IB) is an acute, highly contagious, and economically important viral disease of chickens. The S1 subunit from Spike (S) protein plays the major role in protective immunity and is involved in the host-virus interactions, as well as infectious bronchitis virus (IBV) serotyping. Aim of the present study was multi-aspect analysis of the molecular and immunological features of 5' part belonging to the S1 glycoprotein sequence of Iranian 793/B IBV strain isolates. This might ideally help in characterization, prevention, and vaccine development. The tissue samples were prepared, followed by virus isolation, reverse transcription polymerase chain reaction and restriction fragment length polymorphism analysis. In addition, sequencing and registration of the sequences in the National Center for Biotechnology Information were performed. Moreover, 12 sequences were retrieved from Fars province, Iran. The next steps included evaluation of conservation/variability along the sequences, phylogenetic analysis, estimation of the average evolutionary divergence over all the sequence pairs, predicting the phosphorylation/N-glycosylation/palmitoylation sites, and the final analysis of antigenicity. The findings of alignment, entropy plot, and pairwise similarity analysis revealed 17 hypervariable regions. The isolates belonging to Tehran were clustered in phylogenetic tree, and the most similar isolates to them were ADW11182 and ADW11183. Location of some of the N-glycosylation/phosphorylation/palmitoylation points indicated that these sites were conserved among the isolates. Furthermore, the frequency of epitopes and their scores reflect the high immunogenicity of S1 protein in 793/B serotype. Analysis of the primary and secondary structures demonstrated that their parameters had variable values and were different regarding the number and location of α-helix, β-strand, and coils. According to our findings, the Iranian isolates of 793/B serotype change their molecular characteristics during time and in different geographical regions. These alterations might account for failure in prevention programs and differences in virulence and pathogenicity.

Keywords: 793/B Serotype, Bioinformatics, Infectious Bronchitis Virus, Molecular Features, S1 Glycoprotein

Séquençage et Analyse In silico d'Aspects multiples de la Glycoprotéine S1 dans le Sérotype 793/B de la bronchite Infectieuse Isolé en l'Iran dans les Années 2003 et 2011

Résumé: La bronchite infectieuse (BI) est une maladie virale aiguë, hautement contagieuse et économiquement importante chez les poulets. La sous-unité S1 de la protéine Spike (S) joue un rôle majeur dans l'immunité protectrice et est impliquée dans les interactions hôte-virus, ainsi que dans le sérotypage du virus de la bronchite infectieuse (IBV). Le but de cette étude était l'analyse de plusieurs aspects des

caractéristiques moléculaires et immunologiques de la partie 5 'appartenant à la séquence glycoprotéique S1 d'isolats de souche IBV 793 / B iraniens. Cela pourrait idéalement aider à la caractérisation, à la prévention et au développement de vaccins. Les échantillons de tissu ont été préparés, le virus a été ensuite isol, et une réaction en chaîne par polymérase en transcription inverse a a été menée pour l'analyse du polymorphisme de la longueur des fragments de restriction. De plus, le séquençage et l'enregistrement des séquences ont été effectués dans le Centre national d'information sur la biotechnologie. Douze séquences ont été récupérées de la province de Fars, en Iran. Les étapes suivantes comprenaient l'évaluation de la conservation / variabilité des séquences, l'analyse phylogénétique, l'estimation de la divergence évolutive moyenne sur toutes les paires de séquences, la prédiction des sites de phosphorylation / N-glycosylation / palmitoylation et l'analyse finale de l'antigénicité. Les résultats de l'alignement, du tracé d'entropie et de l'analyse de similarité par paires ont révélé 17 régions hypervariables. Les isolats àprovenant de Téhéran regroupés dans un arbre phylogénétique et les isolats les plus similaires étaient ADW11182 et ADW11183. L'emplacement de certains des points de N-glycosylation / phosphorylation / palmitoylation indique que ces sites sont conservés parmi les isolats. De plus, la fréquence des épitopes et leurs scores reflètent l'immunogénicité élevée de la protéine S1 dans le sérotype 793 / B. L'analyse des structures primaires et secondaires a montré que leurs paramètres avaient des valeurs variables et étaient différents pour le nombre et la localisation de l'hélice α, du brin β et des spires. Selon nos résultats, les isolats iraniens de sérotype 793/ B modifient leurs caractéristiques moléculaires au cours du temps et selon les différentes régions géographiques. Ces modifications pourraient expliquer l'échec des programmes de prévention et les différences de virulence et de pathogénicité.

Mots-clés: Sérotype 793/B, Bioinformatique, Virus de la Bronchite Infectieuse, Caractéristiques Moléculaires, Glycoprotéine S1

INTRODUCTION

Infectious bronchitis (IB) is a common, highly contagious, acute, and economically important viral disease of chickens caused by a coronavirus of Coronaviridae family. It has adverse effects on egg quality and production factors in layers. In broiler chickens, IB leads in reduced weight gain and feed efficiency[1]. Serous, catarrhal, or caseous exudates are observed in the trachea, nasal passages, and sinuses of the infected chickens. IB can also be associated with swollen pale kidneys and visceral urate depositions. The disease has been identified in chicken flocks of Iran by serological and virus isolation methods (Saif et al., 2011). This virus has a single stranded linear RNA as genome with the molecular weight of 27.6 kb (Bochkov et al., 2006). The four structural proteins which form the viral virion include spike glycoprotein (S), an integral membrane glycoprotein (M), a nucleocapsid protein (N), and a small membrane envelope protein (E) (Saif et al., 2011). The main structural protein that determines the antigenicity of infectious bronchitis virus (IBV) is S protein. It is composed of S1 and S2 subunits with the same molecular weights. The S1 subunit is essential for induction of protective immunity and has most of the virus neutralizing epitopes, including the serotype-specific epitopes (Cavanagh et al., 1992). The S1 protein is expressed in the membrane of host cell. Therefore, the combination of S1 protein and S1 receptor is crucial for the fusion of virus and host cell. Furthermore, S1 protein can act as a site of tissue tropism and accounts for virulence of IBV. The gene part that encodes the S1 subunit of spike glycoprotein is usually used for genotyping of IBV. As a result, even some small changes in S1 amino acid composition may cause changes in IBV serotype (Cavanagh et al., 1992). These alterations result in epitopes variations responsible for virus neutralization, whereas some larger mutations at other parts of S1 might not be effective enough to change the virus antigenicity. It means that IBVs of different serotypes/genotypes not

only have different epitopes, but also share common epitopes, which have important role in cross-immunity and cell-mediated immune responses (Saif et al., 2011; Cavanagh et al., 1992). Adzhar et al. (1997) identified the differences between 793/B (also known as 4/91 and CR88 serotypes) and other known strains using molecular techniques. They also stated that the new isolates could be divided into three subgroups (Adzhar et al., 1997). Regular vaccination with live attenuated and inactivated vaccines of mass strains is performed in Iran. However, an epidemiological survey showed that IBV is still present in Iran resulting in serious financial losses in poultry industry. In recent years, several researchers have reported the prevalence of a new serotype, known as 793/B in Iran by reverse transcription polymerase chain reaction (RT-PCR) technique (Seyfi Abad Shapouri et al., 2004; Yan, 2008). This finding may partially explain the failure of Mass-type vaccines in protecting the chicken flocks and importance of revising the prevention programs against IB in Iran. Virus-related databases and bioinformatics analysis are crucial for understanding the relationships within complex datasets regarding viruses, virus-host interactions, and immune responses (Ranjbar et al., 2015, Dadmanesh, et al., 2015, Ranjbar et al., 2014). Prediction of different features, such as N-glycosylation sites, epitope recognition, similarity, in addition to the primary and secondary structures of proteins are essential for analyzing the structure-function relationship of the proteins encoded by viral genomes (Yan, 2008, Keyvani et al., 2016). Antigenic variations of the IBV strains are usually investigated by nucleotide sequencing of the gene coding S protein, or more specifically, by amino acid sequencing of the S1 subunit in S protein, to which most of the neutralizing antibodies bind (Kusters et al., 1989; Kant et al., 1992). Various serotypes generally have large differences in the deduced amino acid sequences of the S1 subunit (Kusters et al., 1989). Glycoprotein N-glycosylation products have become an area of intense interest in recent years, and are shown to be associated with changes in virulence, cellular tropism in enzymes, and survival of viruses (Vigerust and Shepherd, 2007). Palmitoylation is one of the post-translational modification processes involving lipids (Linder and Deschenes, 2007; Wan et al., 2007). It enhances protein surface hydrophobicity, membrane affinity, and aggregation. Furthermore, plays an important role in modulating membrane trafficking of the proteins, stability, cell signaling, and apoptosis (Draper et al., 2007; Linder and Deschenes, 2007). Computational methods that identify the potential of palmitoylation provide useful information. Protein phosphorylation is of importance in modulation and regulation of physiological functions of virus proteins during replication and assembly (Ivanov et al., 2001; Ingrell et al., 2007). With this background in mind, the present study aimed to sequence and analysis the molecular features of the 5' part belonging to the S1 glycoprotein in IBV serotype 793/B isolates from Tehran and Fars provinces in two different years. Consequently, the differences in their molecular and biological properties are revealed.

MATERIALS AND METHODS

Sampling. Tissue samples, including lung, trachea, and kidney taken from broiler and layer flocks were submitted to the avian virology laboratory of the department of poultry diseases in order to isolate the respiratory viruses. Twenty IBV isolates obtained from the different parts of Iran were isolated in embryonated chicken eggs. Infected embryos displayed stunting, urate deposition in the mesonephros, or death after three passages. The whole S1 gene of the three local IBV isolates showed a band of above 1600 base pair (bp) in gel electrophoresis following RT-PCR. Restriction fragment length polymorphism (RFLP) analysis using three restriction enzymes (HaeIII, HindIII, and EcorI) revealed the 793/B serotype pattern. The S1 gene of these strains was sequenced according to the method of Vasfi Marandi et al. (Vasfi Marandi et al., 2007).

S1 Gene Sequencing and Registration. The sequence data were edited and then the S1 spike glycoprotein nucleotide sequences were registered in National Center for Biotechnology Information (NCBI). GenBank approved this sequence with the accession numbers (AC) of AAS48624, AAS48625, and AAS48626 (http://ncbi.nlm.nih.gov). In the next steps, the amino acid sequences related to the S1 gene were used for comparison purposes.

Retrieving Iranian S1 Serotype, Alignment, and Entropy Plot for Conserved and Hypervariable Regions. Spike (S1) glycoprotein sequences of all Iranian IBV isolates already registered by other Iranian researchers were retrieved from UniprotKB (http://www.expasy.org/uniprot) database. The accession numbers included ADW11164, ADW11165, ADW11167, ADW11168, ADW11169, ADW11173, ADW11180, ADW11181, ADW11182, ADW11183, ADW11184, and ADW 11185. All of the existing S1 glycoproteins were reported by Boroomand et al. (2011) and Poorbaghi et al. (2011) in Fars province, Iran and belonged to the 793/B serotype of IBV.

Tehran and Fars S1 protein sequences were integrated into a FASTA file and aligned using ClustalW, followed by analysis and trim in CLC sequence viewer software version 6.6.2. At the final step, the areas with ambiguous alignment were excluded from the analyses. Entropy values (Hx) were measured as the variations at each amino acid position in the set of aligned sequences. The values could vary from 0 (i.e., no variation seen) to 3.04 (i.e., all the possible 20 amino acids or a gap occur with equal frequency). Evaluation of the most conserved and variable regions is demonstrated in figures 1 and 2. The sequence AAS48624 was used for further analysis due to the complete similarity between AAS48624 and AAS48625. In all the isolates, size of the target part in S1 protein was 346 amino acids in length after alignment and edition.

Phylogenetic Analysis. The phylogenetic analysis was conducted by MEGA 5.3 software package based on different sets of aligned sequences of the S1 glycoprotein in the IBV isolated from two different provinces of Iran. Trees were constructed using the neighbor-joining (NJ) algorithm under the global gap removal option and Kimura's two-parameter substitution model. Robustness of phylogenetic analysis was measured using bootstrap analysis with 10,000 replications. The percentage of replicate trees, in which the associated taxa clustered together in bootstrap test, is shown next to the branches.

Estimates of Average Evolutionary Divergence over All Sequence Pairs. The number of amino acid substitutions per site were calculated based on the average sequence, and standard error estimate(s). Fifteen sequences from two provinces were analyzed by the Poisson correction model. All the positions containing gaps and missing data were eliminated and there were a total of 346 amino acids. Evolutionary divergence analyses were conducted using the MEGA5.

Prediction of N-glycosylation. The NetNGlyc 1.0 Server (http://www.cbs.dtu.dk/services/NetNGlyc) was used for predicting the N-glycosylation sites in S1 glycoprotein using artificial neural networks. These networks examine the sequence context of Asn-Xaa-Ser/Thr sequons. By default, predictions are performed only on the Asn-Xaa-Ser/Thr sequons (including Asn-Pro-Ser/Thr).

Analysis of Potential Phosphorylation Sites. Potential phosphorylation sites were determined using the website http://www.cbs.dtu.dk/services/NetPhos. The NetPhos 2.0 server produces neural network predictions for serine, threonine, and tyrosine phosphorylation sites in eukaryotic proteins.

Analysis of the S Glycoprotein Primary Structure and Amino Acid Distributions. The primary structure (sequence statistics) of S1 glycoprotein belonging to the IBV was investigated in terms of length, molecular weight (MW), isoelectric point (pI), and amino acid distribution. The outcomes were arranged in tables by the Expasy tools (http://us.expasy.org/tools/pi_tool.html). In order to evaluate the amino acid distribution, we applied http://www.lrrfinder.com/lrrfinder.php.

Prediction of S Glycoprotein Secondary Structures The secondary structures (i.e., alpha helix, beta strand, and random coil) of the protein were predicted using the bioinformatics tools available on website http://npsa-pbil.ibcp.fr. The method of GOR4 (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page= npsa_gor4.html) was used to identify the alpha helices, beta strands, and coil residues. A minimum of three or more turns were taken into account for one helix, strand, or coil in the structure of spike glycoprotein.

Prediction of Palmitoylation Sites. Palmitoylation sites were assessed with the medium threshold frequency through services at http://csspalm.biocuckoo. org/prediction.php. In addition, we used CSS-Palm 2.0 software to predict the palmitoylation sites.

Antigenicity Prediction Based on Kolaskar and Tongaonkar Method. The antigenic epitopes of S1 glycoprotein in the IBV were determined using the method of Kolaskar and Tongaonkar. This prediction is a semi-empirical approach developed based on the physico-chemical properties of amino acid residues (i.e. hydrophilicity, accessibility, and flexibility) and their frequencies among the 156 experimentally determined epitopes from 34 different proteins. This approach has the accuracy of 75% in detecting the antigenic peptides.

RESULTS

Alignment. Alignment results revealed an almost complete similarity between the isolates of Tehran. However, there were significant differences between the S1 glycoproteins of IBVs from Tehran and Fars (Figure 1). Pair-wise comparison between the isolated viruses revealed that among Fars isolates the sequence closest to AAS48624 (after Tehran sequences) was ADW11181 with 92% similarity. Moreover, the sequences closest to AAS48626 were ADW11181, ADW11182, and ADW11184, all with 92% similarity.

Based on alignment and entropy plot the hypervariable regions (HVR) were observed in the amino acids 22-24, 37-39, 53-56, 62-64, 86-89, 116-120, 129-130, 142-145, 156-157, 155-156, 166-167, 168-169, 185-187, 266-267, 285-286, 288-289, and 332-333. The amino acids 53-56 and 116-120 showed the highest variability among all.

Phylogenetic Study. For the phylogenetic analysis of S1 glycoprotein in the IBV isolated from the two different provinces of Iran, amino acid sequences of Tehran and Fars isolates were aligned, compared, and edited using CLC sequence viewer 6.6.2. The MEGA 5.3 software package was utilized to construct the phylogenetic tree (Figure 3). The results demonstrate that the Tehran isolates are clustered in a branch and are more similar to each other, compared to the Fars isolates. In addition, close relationships were observed between the Tehran isolates with AC ADW11182, ADW11184, ADW11181, and ADW11180 from Fars isolates. Furthermore, AC ADW11169 and ADW11183 have the least similarity to the other sequences from both Tehran and Fars.

Estimate of Average Evolutionary Divergence over All the Sequence Pairs. The mean pair-wise distance and standard error for the set of sequences under study were computed. The findings revealed that the mean distances for all the sequences, Tehran isolates, and Fars isolates were 0.298, 0.027, and 298, respectively. Moreover, the standard errors for all the sequences, Tehran isolates, and Fars isolates were 0.021, 0.007, and 0.021, respectively.

Prediction of N-glycosylation. A range of 10-14 and 12 N-glycosylation sites were found in the isolates from Fars and Tehran, respectively. Conservation of some N-glycosylation sites and total number of the N-glycosylation sites are indicated in details in Table 1. The results demonstrated that most of the N-glycosylation sites were conserved within Tehran and Fars isolates.

Analysis of Potential Phosphorylation Sites. According to the analysis of potential phosphorylation sites, there were differences in number and site of potential phosphorylated peptides in the S1 glycoprotein amino acid sequences (Figure 4 and Table 2). Details of the phosphorylation analysis are depicted

in Figure 4. In the isolates from Tehran, serine phosphorylation sites at positions Ser80, Ser95, Ser135, Ser151, Ser153, Ser157, Ser227, Ser264, and Ser265 were found to be conserved. In addition, the conserved threonine phosphorylation sites included Thr69, Thr179, and Thr268. The conserved tyrosine phosphorylation sites entailed Tyr21, Try26, Try56, Try71, Try155, Try 185, Try212, Try246, Try305, Try318, and Try325. In Fars isolates, the serine phosphorylation sites at positions Ser37, Ser62, Ser227, Ser265, and Ser303 were found as conserved. ADW11169 and ADW11183 had similar pattern of serine phosphorylation but were different from the others. Furthermore, the conserved threonine phosphorylation sites were 179 and to some extent 264 and 268. Again ADW11169 and ADW11183 had a similar pattern of threonine phosphorylation but were different from the others. It was observed that the conserved Tyrosine phosphorylation sites were 246, as well as 305 and to some extent 318.

Antigenic Determinants Analysis. The approach introduced by Kolaskar and Tongaonkar is known as the simplest and most applicable for prediction of antigenic determinants (Kolaskar and Tongaonkar, 1990). The mentioned method relies on occurrence of the amino acid residues in experimentally determined epitopes. Antigenicity prediction plots for a part of S1 glycoprotein in the IBV isolated from two different provinces of Iran using Kolaskar-Tongaonkar algorithm is represented in Figure 5. The method of Kolaskar-Tongaonkar evaluates hydrophilicity, accessibility, and flexibility together to predict the epitopes. The findings show that about 235 epitopes in all sequences were predicted. It could be concluded that these epitope frequencies and their scores show a high immunogenicity for S1 protein (Figure 5).

Analysis of Primary Structure and Amino acid Distribution in a Part of S1 Glycoprotein. The data obtained from Expasy's ProtParam regarding the primary structures are demonstrated in Table 3. Analysis of the primary structure showed that except instability index I, all parameters had variable values. However, the ranges of different parameters are close to each other. The average length of protein sequences and molecular weight of constructs are mentioned in the table. Isoelectric Point (pI) is the pH point in which the surface of protein is covered with charges, while the net charge of protein is zero. The isoelectric point (pI) is essential for estimation of solubility and mobility in an electric field. The calculated isoelectric points (pI) in Tehran isolates were 7.52 and 6.69. It reflects the neutral nature of protein and a little tendency from neutral point to basic and acidic pH. All the values for Fars province isolates are above seven, indicative of the basic nature for the S1 glycoprotein in these isolates. The instability index provides an estimation of the protein in-vitro stability. Regarding the instability index results, all the S1 glycoproteins of isolates can be classified as stable protein. The high aliphatic index (range of 69.02-73.79 in both isolates) indicates that the S1 protein is stable in a range of temperatures. The grand average of hydropathicity values were negative in most of the isolates, except the four isolates of Fars province that are a little positive. The latter result indicates a high hydrophilicity for the proteins and a high interaction between the protein and surrounding water molecules. Finally, the data obtained concerning the primary structure and amino acid distribution (Figure 6) showed differences in various physical and chemical parameters, in addition to the hydrophobic and hydrophilic nature of all the isolates. This reveals a relative variability in the strains regarding the S1 glycoprotein.

Secondary Structure of S1 Glycoprotein. Results of the secondary structures evaluation are presented in Figure 7. Both the Fars and Tehran isolates had three or four α-helix structures in the same or different locations of their S1 glycoprotein belonging to the IBV. The numbers of β-strands in the Tehran isolates were 25 positions and ranged from 23 to 30 in the Fars isolates. Furthermore, calculation of the coils (beta turns) revealed 27 and 28 coils in the Tehran isolates and a range of 27 to 33 coils in the Fars isolates.

Table 1. Predicted N-glycosylation sites. For picking up the N-glycosylation sites with high specificity (asparagines very likely to be glycosylated), only the (+) predictions were used (and better) for asparagines that occured within the Asn-Xaa-Ser/Thr triplet (no proline at the Xaa position). Note that identifying the sites by this method would include sensitivity

AC.	ADW11164	ADW11165	ADW11167	ADW11168	ADW11169	ADW11173	ADW11180	ADW11181	ADW11182	ADW11183	ADW11184	ADW11185
N-g*	24, 51, 75, 103, 147, 166, 181, 215, 240, 250, 267, 274, 279	24, 51, 75, 103, 147, 166, 181, 215, 240, 250, 267, 274, 279	24, 51, 75, 103, 141, 147, 166, 181, 215, 240, 250, 267, 274, 279	24, 51, 75, 103, 147, 166, 181, 215, 240, 250, 267, 274, 279	51, 75, 103, 147, 166, 181, 215, 250, 265, 274	24, 51, 75, 103, 147, 166, 181, 215, 240, 250, 267, 274, 279	23, 54, 75, 103, 147, 166, 181, 215, 240, 250, 267, 274, 279	23, 54, 54, 103, 147, 166, 181, 215, 240, 250, 267, 274, 279	23, 54, 75, 103, 147, 166, 181, 215, 240, 250, 267, 274, 279	51, 75, 103, 147, 166, 181, 215, 250, 265, 274	23, 54, 75, 103, 147, 166, 181, 215, 240, 250, 267, 274, 279	51, 77, 103, 147, 166, 181, 215, 240, 250, 274
T**	13	13	14	13	10	13	13	13	13	10	13	10

*N-glycosylation sites **Total number

AC.	AAS48624	AAS48626
N-glycosylation sites	23, 75, 103, 147, 166, 181, 215, 240, 250, 267, 274, 279	23, 75, 103, 147, 166, 181, 215, 240, 250, 267, 274, 279
Total number count	12	12

Table 2. Phosphorylation site predictions

AC	Serine phosphorylation sites	Threonine phosphorylation sites	Tyrosine phosphorylation sites
AAS48624	9	3	12
AAS48626	10	4	11
ADW11164	7	2	7
ADW11165	7	2	7
ADW11167	5	3	7
ADW11168	7	2	7
ADW11169	4	5	6
ADW11173	7	2	7
ADW11180	13	2	11
ADW11181	7	4	11
ADW11182	7	4	11
ADW11183	4	5	6
ADW11184	7	4	11
ADW11185	9	2	9

Table 3. Parameters computed using Expasy's ProtParam

AC	AAS 48624	AAS 48626	ADW 11164	ADW 11165	ADW 11167	ADW 11168	ADW 11169	ADW 11173	ADW 11180	ADW 11181	ADW 11182	ADW 11183	ADW 11184	ADW 11185
M W	38527.4	38367.0	38248	38248	38322.1	38248.0	38568.5	38248.0	38465.3	38530.4	38560.5	38496.5	38560.5	38056.6
pI	7.52	6.69	7.10	7.10	7.97	7.10	7.50	7.10	8.65	8.21	8.40	7.93	8.40	8.22
Total-R&+R	20, 21	22, 21	19,19	19,19	18, 20	19,19	22, 23	19, 19	17, 23	19, 22	19, 23	22, 24	19, 23	15, 18
II	stable	stable	stable	stable	stable	stable	stable	stable	stable	stable	stable	stable	stable	stable
AI, GRAVY	69.02, -0.060	67.05, -0.144	70.69, 0.009	70.69, 0.009	69.57, -0.040	70.69, 0.009	72.66, -0.135	70.69, 0.009	67.63, -0.086	70.72, -0.067	71.85, -0.053	73.79, -0.114	71.85, -0.053	68.44, -0.084

MW: molecular weight; pI: theoretical isoelectric point; -R: number of negatively-charged residues (Arg+Lys); +R: number of positively-charged residues (Asp+Glu); II: instability index; AI: aliphatic index; GRAVY: grand average of hydropathicity

Table 4. Details of palmitoylation sites predictions

AC	Position	Peptide	Score	Cut-off	Cluster
AAS48624	13, 16, 65	LLLVTLWCALCSALL,VTLWCALCSALLYDN, AGSAPECTAGTFYE	0.648, 1.314, 1.347	0.308, 1.225, 1.225	A, C, C
AAS48626	13, 16, 65	LLLVTLRCALCSTLL, VTLRCALCSTLLYDN, NAGSASECTAGTFYE	0.686, 1.298, 1.24	0.308, 1.225, 1.225	A, C, C
ADW11164	16, 65	VTLLFALCSAALFDN, NAGSSSECTAGAIYW	1.426, 1.289	0.497, 1.225	B, C
ADW11165	16, 65	VTLLFALCSAALFDN, NAGSSSECTAGAIYW	1.426, 1.289	0.497, 1.225	B, C
ADW11167	16, 65	VTLLFALCSAALFDN, NAGSSTQCTAGAIYW	1.417, 1.314	0.497, 1.225	B, C
ADW11168	16, 65	VTLLFALCSAALFDN, NAGSSSECTAGAIYW	1.426, 1.289	0.497, 1.225	B, C
ADW11169	13, 163	LLLVTLWCALCSALL, PVFKSLHCLNNDTSV	0.552, 0.319	0.308, 0.308	A, A
ADW11173	16, 65	VTLLFALCSAALFDN, NAGSSSECTAGAIYW	1.426, 1.289	0.497, 1.225	B, C
ADW11180	13, 65	LLLVTLWCALCSALL, NAGSSSDCTAGTFYE	0.605, 1.421	0.308, 1.225	A, C
ADW11181	13, 65	LLLVTLWCALCSALL, NAGSSSDCTAGTFYE	0.624, 1.421	0.308, 1.225	A, C
ADW11182	13, 65	LLLVTLWCALCSALL, NAGSSSDCTAGTFYE	0.624, 1.421	0.308, 1.225	A, C
ADW11183	13, 163	LLLVTLWCALCSALL, PVFKSLHCLNNDTSV	0.552, 0.319	0.308, 0.308	A, A
ADW11184	13, 65	LLLVTLWCALCSALL, NAGSSSDCTAGTFYE	0.624, 1.421	0.308, 1.225	A, C
ADW11185	13, 16, 98	LLLVTLLCALCSAAL, VTLLCALCSAALYDS, AWSSSQFCTAYCNFS	0.967, 1.667, 1.479	0.308, 0.497, 1.225	A, B, C

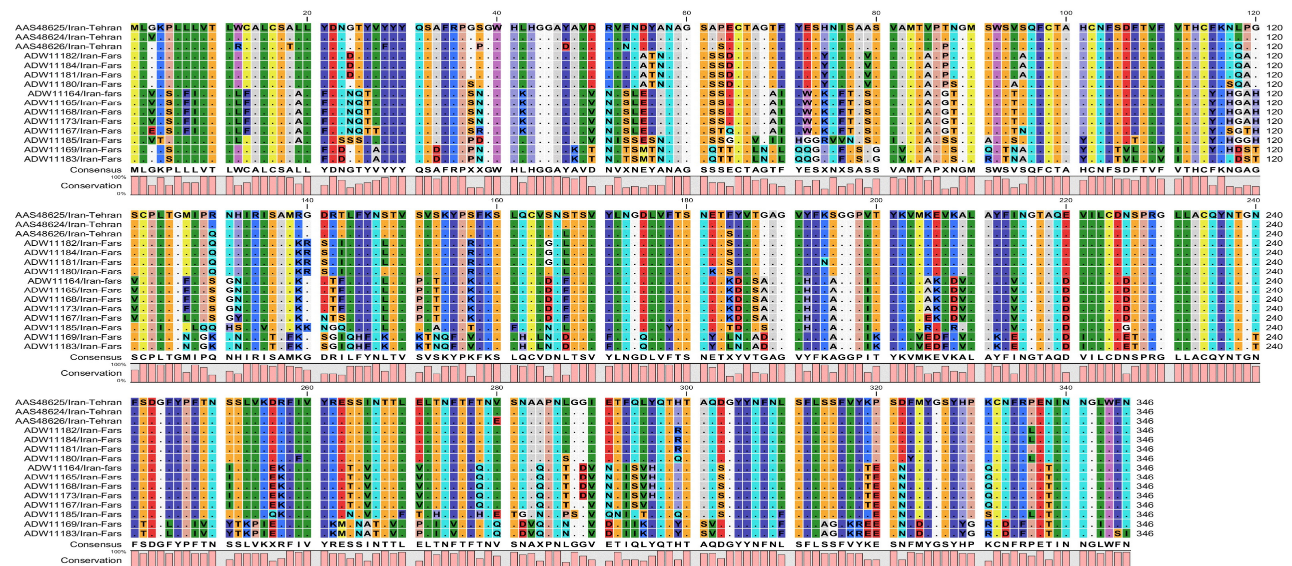

Figure 1. Comparison of amino acid sequences in the S1 glycoprotein of Iranian IBV isolates. Gene bank accession numbers are shown. All the sequenced proteins have been used for alignment. Identical, conserved, and semi-conserved regions along the spike glycoprotein sequences in IBV can be seen. All the residues that are identical to the top sequence in an alignment are marked by a dot ('.'). The consensus and rate of conservation are indicated in the bottom panel.

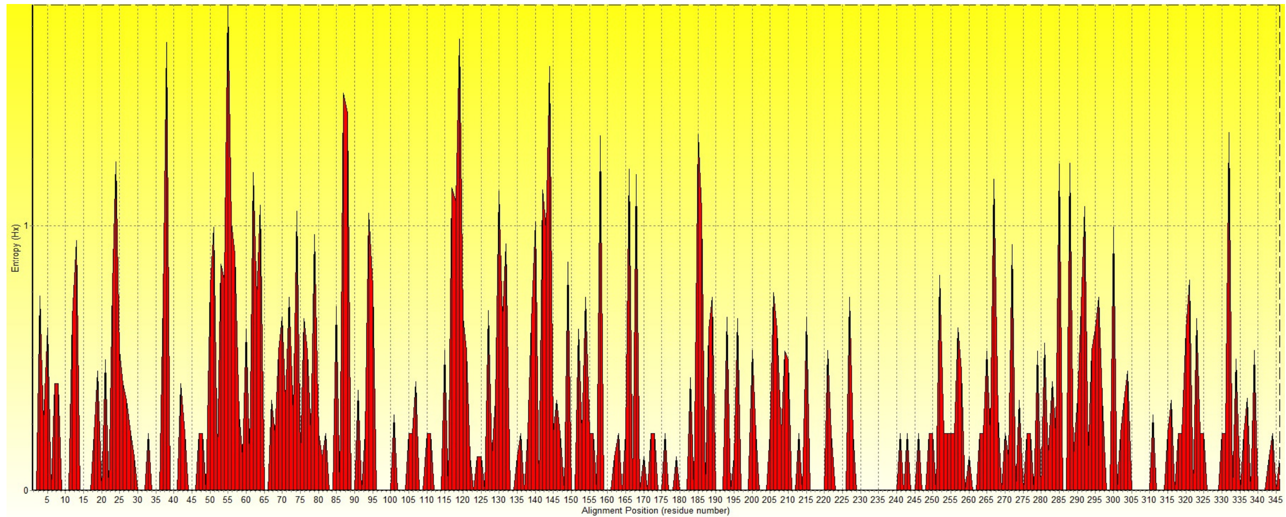

Figure 2. Variations along the 5' part of the S1 glycoproteins in the IBV shown by entropy plots. Circles show the borderline variations not included in the variation analysis.

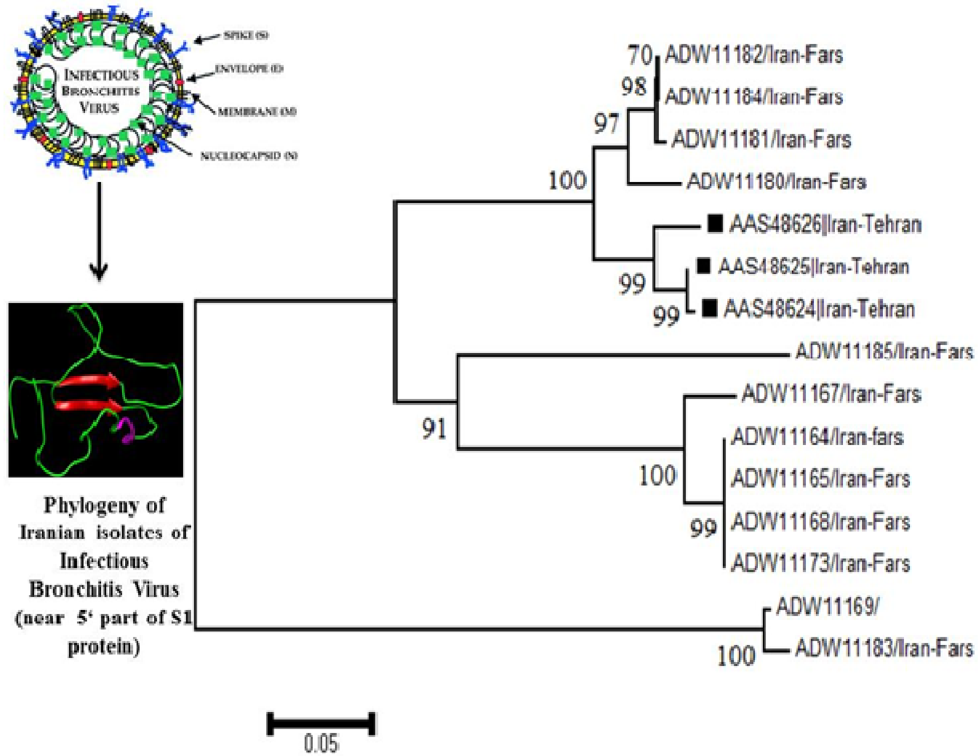

Figure 3. Phylogenetic tree constructed with MEGA 5.3 software based on the amino acid sequences, illustrating the relationships in part of S1 glycoprotein of IBV isolated from two different provinces of Iran. S1 glycoprotein from Tehran is marked by ■ . The sequence information at the tips of the branches includes the accession numbers for the sequences and the source province.

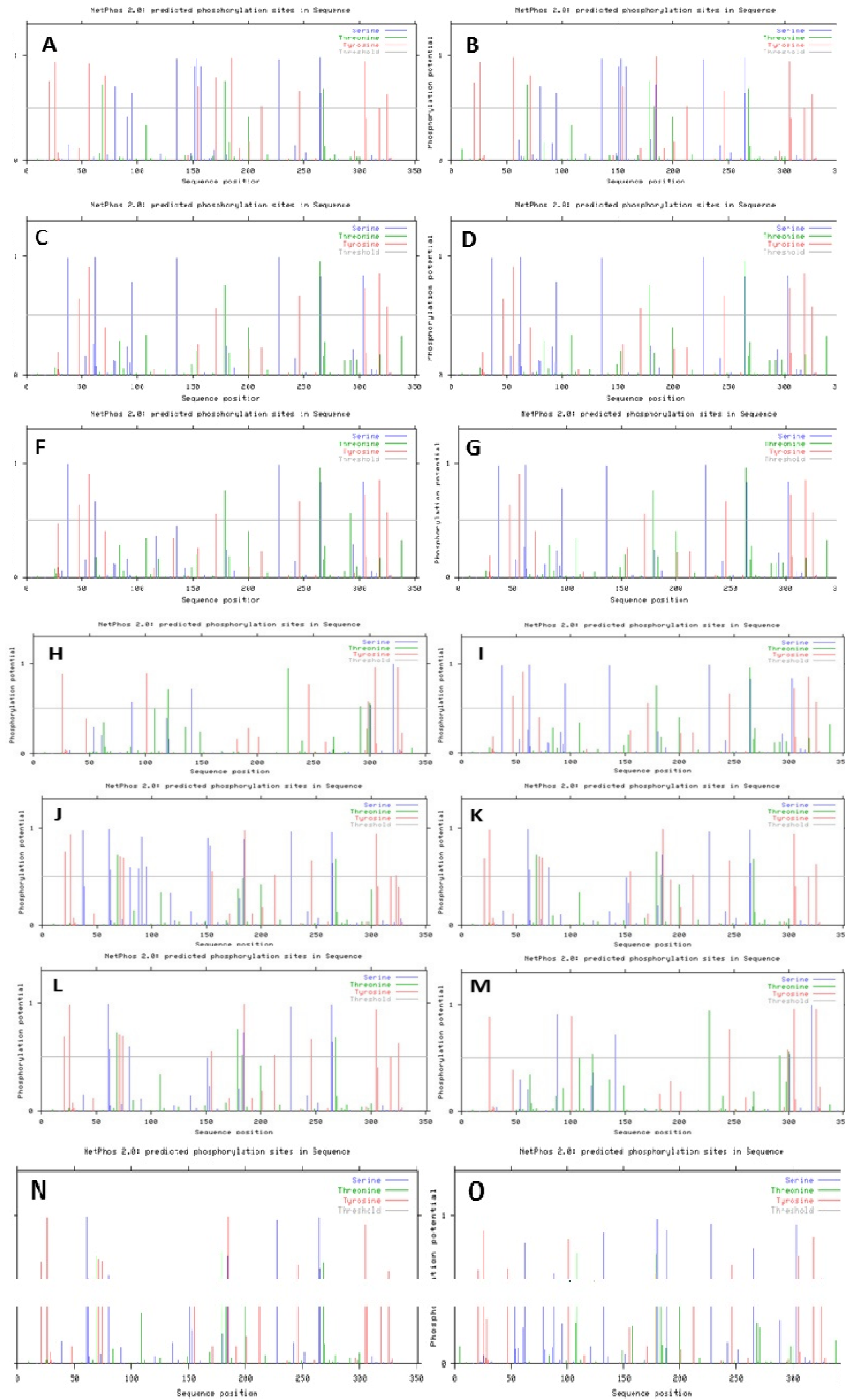

Figure 4. Plots of potential phosphorylation sites; A to M plots: Fars isolate; N and O plots: the S1 glycoprotein of Tehran isolates

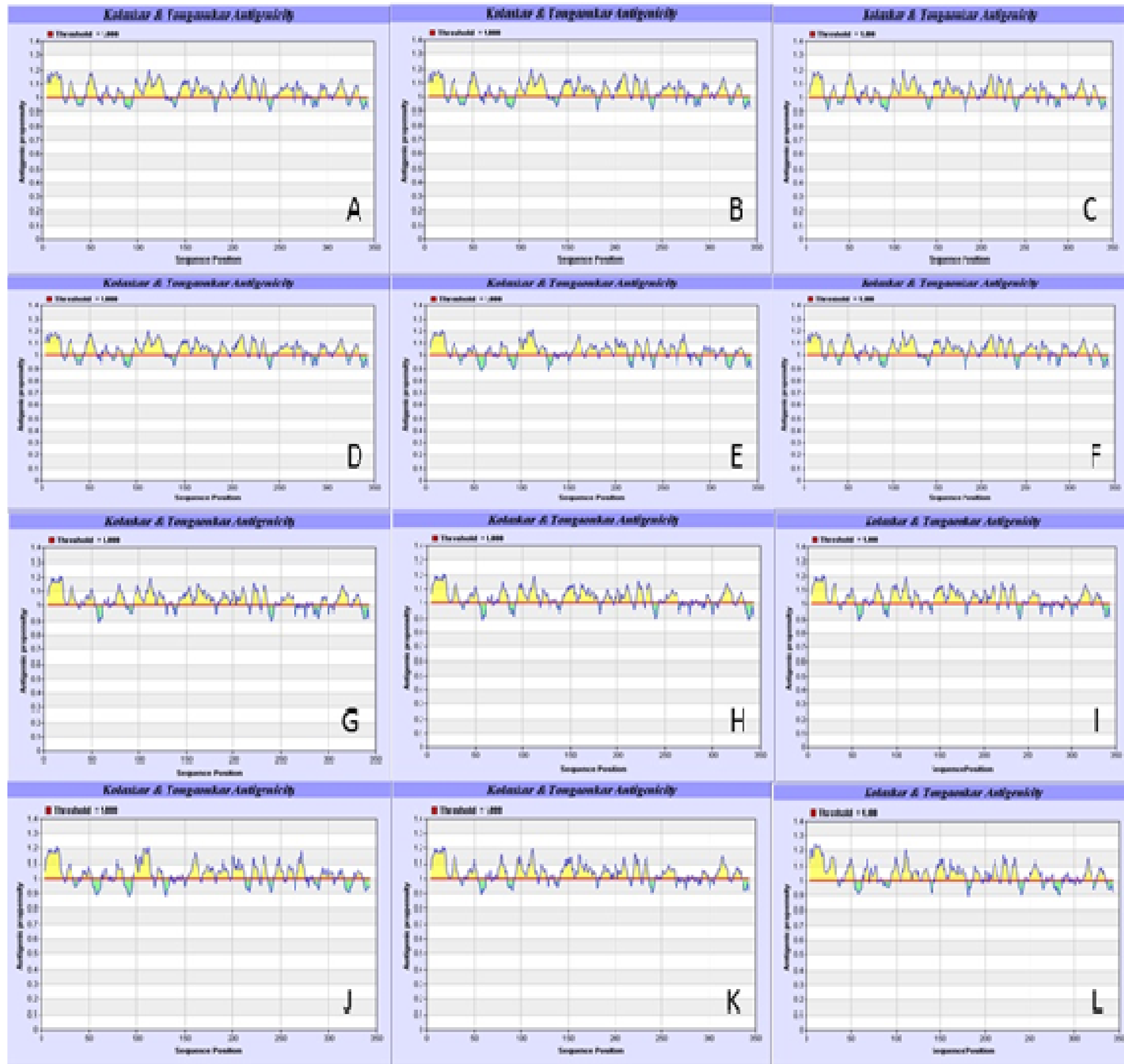

Figure 5. Antigenicity prediction plots for part of the S1 glycoprotein in Iranian infectious bronchitis virus isolates using Kolaskar-Tongaonkar algorithm. Regions with antigenic propensity scale of >1 are antigenic regions. Threshold, average, maximum, and minimum antigenicity are 1, 1.033, 1.215, and 0.880, respectively. Window size and center position were 7 and 4, respectively. A to L: ACs from ADW11164 to ADW11185. The bottom figures depict Tehran isolate S1 glycoprotein, ACs AAS48624 (M) and AAS48626 (N). Threshold, average, maximum, and minimum antigenicity are 1, 1.036, 1.196, and 0.895, respectively.

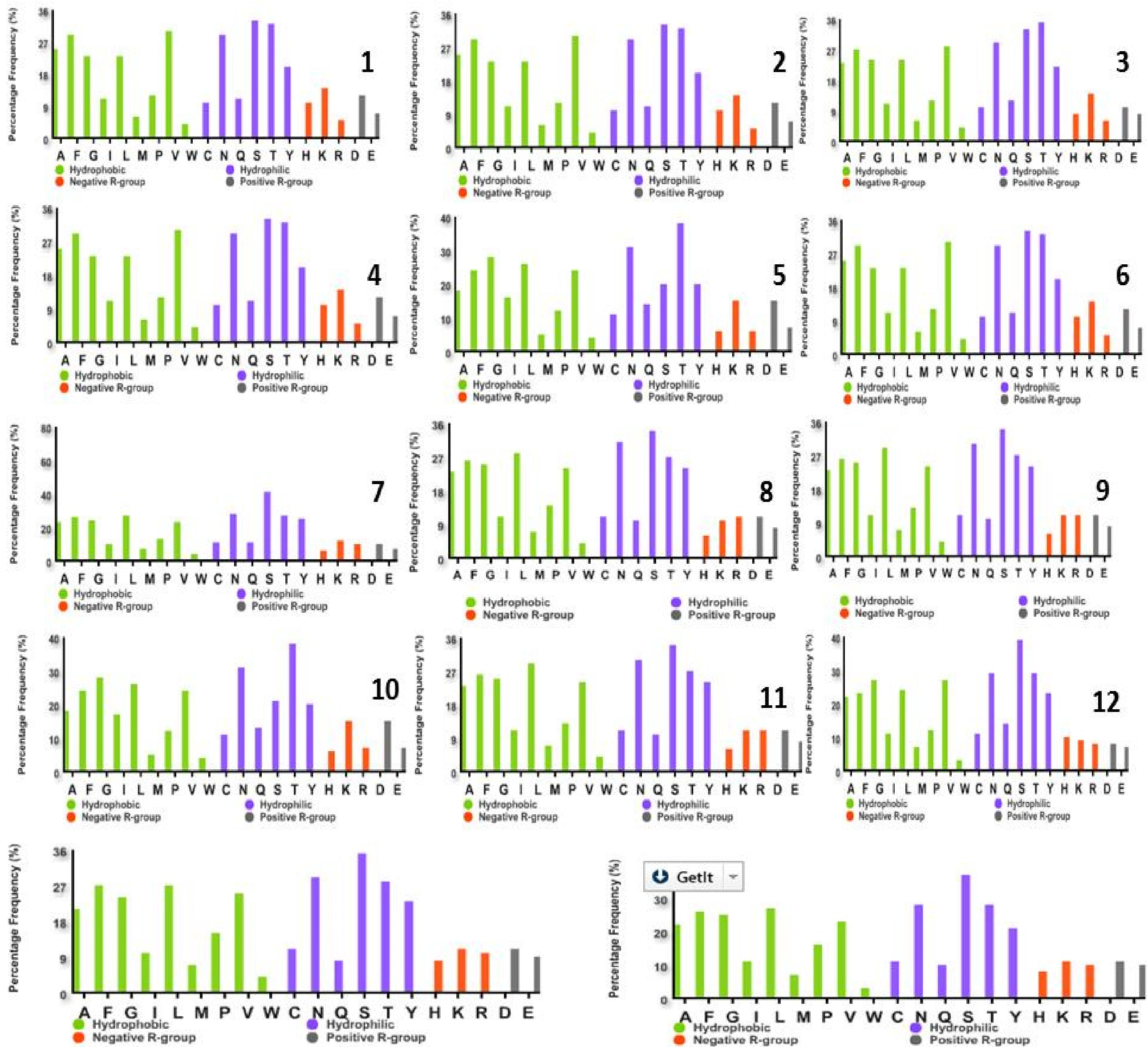

Figure 6. Amino acid distribution. Frequency of hydrophobic, hydrophilic, positive, and negative R-groups (1 to 12). The bottom left (AAS48624) and right (AAS48626) panels show sequences of S1 glycoprotein for Tehran and Fars isolates, AC ADW11164 to ADW11185.

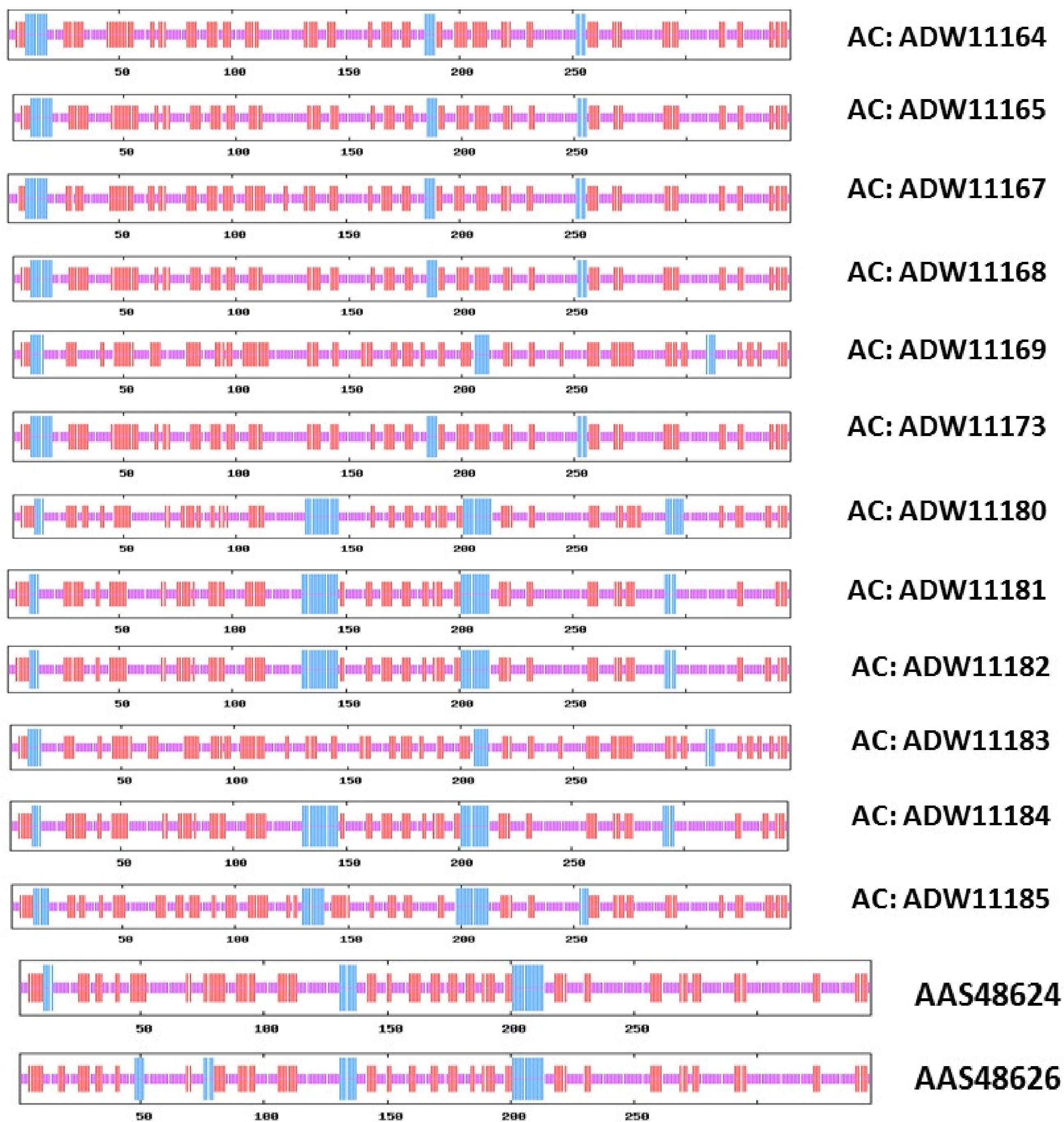

Figure 7. GOR IV secondary structure prediction method (graphic visualization of the prediction); blue: α-helix; red: extended strand (β-strand); violet: other states (coils)

Palmitoylation Sites. The strains of ACs AAS48624, AAS48625, and AAS48626 had 3 palmitoylation sites. On the other hand the ADW11164, ADW11165, ADW11167, ADW11168, ADW11169, ADW11173, ADW11180, ADW11181, ADW11182, ADW11183, and ADW11184 (but not ADW11185) all possessed 2 palmitoylation sites (Table 4). It should be noted that 13, 16, 65, and 163 seemed to be conserved sites for palmitoylation in almost all the isolates.

DISCUSSION

The previous studies on molecular characteristics of complete or part of S1 glycoprotein sequence of 793/B IBV in Massachusetts and other serotypes revealed a high degree of variation and complexity for this protein (Abro et al., 2012). Modifications such as glycosylation, palmitoylation, and phosphorylation have enormous effects on folding, transportation, replication, assembly, infectivity, pathogenicity, immunogenicity, and virulence of the relevant virus particles (Veit and Schmidt, 2006). In the present study computational and in silico prediction approaches by public databases, as well as the bioinformatics tools were used to analyze the S1 sequences of 793/B serotype of IBV strains isolated from two provinces of Iran. Our findings showed that there were high similarities and phylogenetic diversities among the IBV strains isolated from the two different provinces. These might result from the common mutations within the S1 gene of the virus. Moreover, it was observed that the isolates from the two different provinces had low homology. There was much higher homology among the viruses isolated during the same period of time and from the same geographical origin. These results are to some extent similar to the previous studies that assessed the different isolates, as well as their phylogeny and similarities (Ladman et al., 2006; Wei et al., 2008). Analysis of the different parts in S1 may indicate different levels of homology and similarity, especially if it includes different HVRs (Wang et al., 1993; Dolz et al., 2008). Antigenicity analysis of the 5' part in the S1 protein revealed the presence of shared epitopes. This can be helpful in preparing the vaccines with high cross-protection efficacy. Several articles have reported the level of homology in the S1 gene/protein or a part of it and the level of cross-protection (Cook et al., 2001; Meir et al., 2004). Our results showed that there is a higher chance of suitable cross protection between the strains isolated during the same time and from the same geographical region. The vaccines against S1 glycoprotein of IBV strains that differ from the other strains in only a few amino acids provide a significant cross-protection (Abdel-Moneim et al., 2006). There are reports on the different IBV strains from other known IBVs isolated from various geographical regions based on analysis of S1 sequence (Shimazaki et al., 2008). Conserved palmitoylation sites introduced in this study may reveal some special features of the 793/B strains (Table 4). The palmitoylation and glycosylation sites have been evaluated in influenza virus (Melkonian et al., 1999) and Human Immunodeficiency Virus-1 (Zhang et al., 2000; Keyvani et al., 2016). The findings of all these studies show that these changes have impacts on viral biology and functions. An average of 12 N-glycosylation sites and some conserved sites were found in the partial protein sequence analysis of S1 protein in Tehran and Fars isolates. Abro et al. investigated the S glycoprotein for N-glycosylation and found 30 and 35 N-glycosylation sites in the Massachusetts and QX-like strains, respectively. Furthermore, the N-glycosylation sites of ASN-Xaa-SER/THR-55, -147, -200, and -545 were present in the spike gene of QX-like strains. They finally concluded that most of the N-glycosylation sites were conserved within the genotype of the mentioned strains, except in the strain CK/SWE/082066/10 that had lost ASN-Xaa-SER/THR-533.

According to the findings of the current study, analysis of the potential phosphorylation sites revealed that the potential phosphorylated peptides in the S1 glycoprotein were different regarding the number and location. These results are consistent with the previous study on Massachusetts and QX-like strains (Abro et

al., 2012). The primary structure analysis in IBV isolates demonstrated that all the parameters had variable values, although the differences are not significant. These results are as well repeated for the secondary structures. There is a discrepancy between the studies on these values and the overall numbers of structures, which depends on either partial or complete evaluation of the S (or S1) protein (Abro et al., 2012). The findings of our study suggest that the characteristics of the Iranian IBVs might have changed during time and in different geographical regions. Analysis of the genomic/proteomic data and understanding the structure and function of important proteins in the viruses are necessary. These evaluations can provide essential information for vaccine development, diagnostic methods, potential anti-viral therapy, viral characteristics, pathogenicity, and epidemiological practices. However, in order to change these predicted findings and achievements to real knowledge, it is recommended to validate these results by proper laboratory experiments.

Ethics

I hereby declare all ethical standards have been respected in preparation of the submitted article.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

Abdel-Moneim, A.S., El-Kady, M.F., Ladman, B.S., Gelb, J., 2006. S1 gene sequence analysis of a nephropathogenic strain of avian infectious bronchitis virus in Egypt. Virol J 3, 78.

Abro, S.H., Ullman, K., Belak, S., Baule, C., 2012. Bioinformatics and evolutionary insight on the spike glycoprotein gene of QX-like and Massachusetts strains of infectious bronchitis virus. Virol J 9, 211.

Adzhar, A., Gough, R.E., Haydon, D., Shaw, K., Britton, P., Cavanagh, D., 1997. Molecular analysis of the 793/B serotype of infectious bronchitis virus in Great Britain. Avian Pathol 26, 625-640.

Bochkov, Y.A., Batchenko, G.V., Shcherbakova, L.O., Borisov, A.V., Drygin, V.V., 2006. Molecular epizootiology of avian infectious bronchitis in Russia. Avian Pathol 35, 379-393.

Boroomand, Z., Razeghian, I., Asasi, K., 2011. Isolation and identification of a new isolate of avian infectious bronchitis virus IRFIBV32 and a study of its pathogenicity. Online J Vet Res 15, 366–380.

Dadmanesh M., Mohammad Mehdi Ranjbar, Seyed Moayed Alavian, Khodayar Ghorban. Sequencing and phylogenetic study of partial NS3 gene of Iranian GB virus C/Hepatitis G Virus (HGV) originated from hemodialysis patients in Tehran. Hepatitis Monthly .2015. Hepat Mon. 2015 Mar; 15(3): e24173.

Cavanagh, D., Davis, P.J., Cook, J.K., 1992. Infectious bronchitis virus: evidence for recombination within the Massachusetts serotype. Avian Pathol 21, 401-408.

Cook, J.K., Chesher, J., Baxendale, W., Greenwood, N., Huggins, M.B., Orbell, S.J., 2001. Protection of chickens against renal damage caused by a nephropathogenic infectious bronchitis virus. Avian Pathol 30, 423-426.

Dolz, R., Pujols, J., Ordonez, G., Porta, R., Majo, N., 2008. Molecular epidemiology and evolution of avian infectious bronchitis virus in Spain over a fourteen-year period. Virology 374, 50-59.

Draper, J.M., Xia, Z., Smith, C.D., 2007. Cellular palmitoylation and trafficking of lipidated peptides. J Lipid Res 48, 1873-1884.

Ingrell, C.R., Miller, M.L., Jensen, O.N., Blom, N., 2007. NetPhosYeast: prediction of protein phosphorylation sites in yeast. Bioinformatics 23, 895-897.

Ivanov, K.I., Puustinen, P., Merits, A., Saarma, M., Makinen, K., 2001. Phosphorylation down-regulates the RNA binding function of the coat protein of potato virus A. J Biol Chem 276, 13530-13540.

Kant, A., Koch, G., van Roozelaar, D.J., Kusters, J.G., Poelwijk, F.A., van der Zeijst, B.A., 1992. Location of antigenic sites defined by neutralizing monoclonal antibodies on the S1 avian infectious bronchitis virus glycopolypeptide. J Gen Virol 73 (Pt 3), 591-596.

Keyvani, H., Ahmadi, N.A., Ranjbar M. M., Ataei Kachooei, S., Ghorban, Kh., Dadmanesh, M., 2016. Immunoinformatics Study of gp120 of Human Immunodeficiency Virus Type 1 Subtype CRF35_AD Isolated from Iranian Patients. Arch of Clin In Dis. 11, 4.

Kusters, J.G., Niesters, H.G.M., Lenstra, J.A., Horzinek, M.C., van der Zeijst, B.A.M., 1989. Phylogeny of antigenic variants of avian coronavirus IBV. Virology 169, 217-221.

Ladman, B.S., Loupos, A.B., Gelb, J., Jr., 2006. Infectious bronchitis virus S1 gene sequence comparison is a better predictor of challenge of immunity in chickens than serotyping by virus neutralization. Avian Pathol 35, 127-133.

Linder, M.E., Deschenes, R.J., 2007. Palmitoylation: policing protein stability and traffic. Nat Rev Mol Cell Biol 8, 74-84.

Meir, R., Rosenblut, E., Perl, S., Kass, N., Ayali, G., Perk, S., et al., 2004. Identification of a novel nephropathogenic infectious bronchitis virus in Israel. Avian Dis 48, 635-641.

Melkonian, K.A., Ostermeyer, A.G., Chen, J.Z., Roth, M.G., Brown, D.A., 1999. Role of lipid modifications in targeting proteins to detergent-resistant membrane rafts. Many raft proteins are acylated, while few are prenylated. J Biol Chem 274, 3910-3917.

Poorbaghi, S., Mohammadi, A., Asasi, K., 2011. Molecular detection of avian infectious bronchitis virus serotypes from clinically suspected broiler chicken flocks in Fars province of Iran. Pak Vet J 32, 93-96.

Ranjbar, M.M., Ghorban, K., Alavian, S.M., Keyvani, H., Dadmanesh, M., Roayaei Ardakany, A., Motedayen, M.H., Sazmand, A., 2014. GB virus C/Hepatitis G virus envelope glycoprotein E2: computational Molecular features and Immunoinformatic study. Hepatitis Month. 13(12):e15342. DOI:10.5812/hepatmon.15342.

Ranjbar, M.M., Ahmadi, N.A., Ghorban, K., Ghalyanchi Langeroudi, A., Dadmanesh, M., Amini, H., et al., 2015. Immnoinformatics: novel view in understanding of immune system function, databases and prediction of immunogenic epitopes. koomesh. 17 (1):18-26.

Saif, Y.M., Fadly, A.M., Glisson, J.R., McDougald, L.R., Nolan, L.K., Swayne, D.E., 2011. Diseases of Poultry, Wiley.

Seyfi Abad Shapouri, M.R., Mayahi, M., Assasi, K., Charkhkar, S., 2004. A survey of the prevalence of infectious bronchitis virus type 4/91 in Iran. Acta Vet Hung 52, 163-166.

Shen, C., Guo, Y., Cheng, A., Wang, M., Zhou, Y., Lin, D., et al., 2009. Characterization of subcellular localization of duck enteritis virus UL51 protein. Virol J 6, 92.

Shimazaki, Y., Horiuchi, T., Harada, M., Tanimura, C., Seki, Y., Kuroda, Y., et al., 2008. Isolation of 4/91 type of infectious bronchitis virus as a new variant in Japan and efficacy of vaccination against 4/91 type field isolate. Avian Dis 52, 618-622.

Vasfi Marandi, M., Keyvani Amineh, H., Akbari Azad, G., 2007. Molecular analysis of three Iranian isolates belonged to 793/B serotype of infectious bronchitis viruses. J Vet Res 62, 69-80.

Veit, M., Schmidt, M.F., 2006. Palmitoylation of influenza virus proteins. Berl Munch Tierarztl Wochenschr 119, 112-122.

Vigerust, D.J., Shepherd, V.L., 2007. Virus glycosylation: role in virulence and immune interactions. Trends Microbiol 15, 211-218.

Wan, J., Roth, A.F., Bailey, A.O., Davis, N.G., 2007. Palmitoylated proteins: purification and identification. Nat Protoc 2, 1573-1584.

Wang, L., Junker, D., Collisson, E.W., 1993. Evidence of Natural Recombination within the S1 Gens of Infectious Bronchitis Virus. Virology 192, 710-716.

Wei, Z.J., Wei, P., Mo, M.L., Li, M., Wei, T.C., Li, K.R., 2008. [Genetic variation of S1 gene hypervariable region I of infectious bronchitis viruses isolated in different periods in Guangxi]. Bing Du Xue Bao 24, 126-132.

Yan, Q., 2008. Bioinformatics databases and tools in virology research: an overview. In Silico Biol 8, 71-85.

Zhang, J., Pekosz, A., Lamb, R.A., 2000. Influenza virus assembly and lipid raft microdomains: a role for the cytoplasmic tails of the spike glycoproteins. J Virol 74, 4634-4644.