Functional and Structural Characterization of ebola Virus Glycoprotein (1976–2015) – An In-silico study

Article *in* International Journal of Biomathematics · July 2017

DOI: 10.1142/S179352451750108X

CITATIONS
10

READS
245

9 authors, including:

Behzad Dehghani
Shiraz University of Medical Sciences
46 PUBLICATIONS **184** CITATIONS
SEE PROFILE

Farzaneh Ghasabi
Shiraz HIV/AIDS Research Center
12 PUBLICATIONS **58** CITATIONS
SEE PROFILE

Tayebeh Hashempour
45 PUBLICATIONS **274** CITATIONS
SEE PROFILE

Hassan Joulaei
Shiraz University of Medical Sciences
120 PUBLICATIONS **1,026** CITATIONS
SEE PROFILE

International Journal of Biomathematics
Vol. 10, No. 8 (2017) 1750108 (23 pages)

DOI: 10.1142/S179352451750108X

Functional and structural characterization of Ebola virus glycoprotein (1976–2015) — An *in silico* study

Behzad Dehghani*,‡, Farzane Ghasabi*,¶, Tayebeh Hashempoor*,§, Hassan Joulaei*,§§, Zahra Hasanshahi*,¶¶, Mehrdad Halaji*,†,‖, Nazanin Chatrabnous*,‡‡, Zahra Mousavi*,†† and Javad Moayedi*,**

*Shiraz HIV/AIDS Research Center
Institute of Health, Shiraz University of
Medical Sciences, Shiraz, Iran

†Department of Microbiology
School of Medicine, Isfahan University of
Medical Sciences, Isfahan, IR Iran
‡dehghanibehzad@gmail.com
§thashem@sums.ac.ir
¶farzanegh92@gmail.com
‖mehrdad.md69@gmail.com
**javad.moayedi88@gmail.com
††zahra.mousavei@gmail.com
‡‡nazanin@sums.ac.ir
§§Joulaee@sums.ac.ir
¶¶zahrahasanshahi48@yahoo.com

Received 19 February 2017
Accepted 26 June 2017
Published 21 September 2017

Ebola virus (EBOV) is the causative agent of a severe hemorrhagic fever disease associated with high mortality rates in humans. This virus has five strains of which Zaire Ebola virus (ZEBOV) is the first and most important strain. It can be transmitted through contact with contaminated surfaces and objects. The genome of EBOV codes one nonstructural and seven structural proteins consisting of two forms of glycoprotein (GP): soluble glycoprotein (sGP) and GP (spike). In this paper, we attempted to characterize and predict physicochemical properties, B-cell epitopes, mutation sites, phosphorylation sites, glycosylation sites, and different protein structures of EBOV GP to provide comprehensive data about changes of this GP during a 40-years course (1976–2015). GP sequences were obtained from NCBI gene bank from 1976–2015. Expasy'sProtParam, PROTSCALE, immuneepitope, Bepipred, BcePred, ABCpred, VaxiJen, DISPHOS, NetPhos, and numerous programs were used to predict and analyze all sequences. More variety of mutations were found in 2015 sequences and mutations were related to huge changes in B-cell epitopes, phosphorylation and glycosylation sites. In addition, our results determined different sites of disulfide bonds and an important mutation related to IgE epitope as well as four potent B-cell epitopes (380–387, 318–338, 405–438 and

§Corresponding author.

434–475). In this study, we suggested the effect of mutations on GP properties, six positions for disulfide bonds and four phosphorylation sites for protein kinase C enzyme. Selected sequences were shown different sites for O-link and N-link glycosylation. A mutation that changed GP to an allergen protein and has a significant role in vaccine designing as well as four potent B-cell epitopes in GP protein were found.

Keywords: EBOV; GP protein; bioinformatics; epitopes mapping; post-modification sits; structural analysis.

Mathematics Subject Classification 2010: 92C40

1. Introduction

Filoviridae family consists of two genera, Marburg virus and Ebola virus, which have likely evolved from a common ancestor [1]. The Ebola virus (EBOV) is non-segmented, pleomorphic and negative-sense RNA virus [2]. They have an individual filamentous morphology with a uniform diameter of 80 nm and variable length of up to 14 mm [3].

The EBOV is the causative agent of a severe hemorrhagic fever disease associated with mortality rates ranging from 23–90% [4, 5]. There are five different members of the EBOV genus; Zaire Ebola virus (ZEBOV), Sudan Ebola virus (SUDV), Tai Forest Ebola virus (TAFV), Bundibugyo Ebola virus (BDBV) and Reston Ebola virus (RESTV) [6]. The deadliest of the five EBOV strain is ZEBOV and the case fatality ratio (CFR) for the Zaire strain ranges from 61–89%. However, recorded clinical outcome for Guinea, Liberia and Sierra Leone has been constantly estimated at 70.8% [7]. As a high contagious disease it can transmit through direct contact from a person who has developed signs and symptoms of illness such as a fever, sore throat and muscle pain. After that, vomiting, diarrhea and rash usually follow, in company with decreased function of the liver and kidneys [8–10].

The genome of each virion is about 19 kb in length, and codes structural and non-structural proteins. The gene order is as follows: 3′–leader–NP–VP35–VP40–GP/sGP–VP30–VP24–L–trailer–5′ [11]. The leader and trailer regions are not transcribed, but carry important signals that control transcription, replication and packaging of the genome into new virions [12].

VP24 is necessary for the correct assembly of a functional nucleocapsid, along with VP35 and NP are sufficient to form nucleocapsid structures, as a lack of VP24 leads to reduced transcription/translation of VP30 [13, 14]. EBOV encodes two forms of glycoproteins (GPs): the first one is a full-length 150–170 kDa protein which is expressed at the cell surface that is called GP, and the second form is the small (60–70 kDa protein), non-structural, and dimeric soluble form which is called soluble glycoprotein (sGP). Furin or related proteases are the major host cell proteases for GP cleavage into GP1 and GP2. GP1 is involved in attachment of the virus to host cells, but GP2 mediates fusion of viral and host membranes [15–17]. EBOV GP is highly glycosylated, which happens in heavily glycosylated mucin-like domain

(MLD), with approximately half of its apparent molecular weight attributed to the attached N-linked and O-linked carbohydrates. MLD is implicated in cell rounding in cultured cells that is suggested to have a role in filovirus pathogenesis [18].

Vaccine development is a critically important strategy to control Ebola infection outbreaks and many vaccine candidates based on EBOV GP were introduced recently DNA vaccine, replication-incompetent chimpanzee recombinant adenoviruses (rAds), and live-attenuated recombinant Vesicular Stomatitis viruses (rVSVs) are some vaccine candidates which showed remarkable achievement in clinical trials [19, 20]. Although several vaccine candidates were introduced, there is not any protective vaccine against EBOV infection yet.

EBOV employed two strategies to avoid host immunity and both of them related to GP function. First one, GP can alter target-cell function and use this evolutionary advantage in immune evasion. Second one, GP showed cytotoxic effects on macrophage and endothelial cells [20–22].

In addition, EBOV can alter cell surface expression of adhesion proteins to evade host immunity [23]. Based on significant roles of GP, this protein was considered as a target for a protective vaccine and other clinical interventions. EBOV was an interesting and important subject for bioinformatics researches, some previous research used bioinformatics tools to improve our knowledge of virus genome and structure [24], especially EBOV GP showed itself as a remarkable protein for understanding the virus life cycle and targeting prevention mechanisms.

Here we employed bioinformatics tools to define structure, modification sites, B-cell epitopes, and finding major changes in all the ZEBOV GP sequences (1976–2015) available in the NCBI.

2. Material and Methods

2.1. ZEBOV GP *sequence analysis*

For bioinformatics analysis, all GP sequences ($n = 98$) from 1976–2015 and one reference sequence (NC_002549) were obtained from NCBI databank at http://www.ncbi.nlm.nih.gov. All sequences were listed in Table 1.

2.2. *Amino acid changing, phylogenetic trees, and signal peptide prediction*

The CLC sequence viewer version Beta (RNA Workbench 4) was used for analysis of the mutations, and homology among sequences. The alignment of the translated peptides was generated using CLUSTAL X software version 1.81 to confirm the reliability of phylogenetic tree and phylogenetic trees was constructed by neighbor-joining method 100 times. To predict signal peptide predication, "SignalP" 4.1 server [25] at http://www.cbs.dtu.dk/services/SignalP/, "SignalBLAST" [26] at http://sigpep.services.came.sbg.ac.at/signalblast.html, and "predisi" [27] at http://www.predisi.de/were used.

Table 1. List of all reference numbers of sequences was used in this research and 16 selected sequences for bioinformatics analysis.

1976 (4seq)	Selected
AF086833.2	NC_002549.1
NC_002549.1	KM655246.1
KM655246.1	
KC242801.1	
1994 (1 seq)	
KC242792.1	KC242792.1
1995 (2 seq)	
KC242796.1	KC242796.1
KC242799.1	
1996 (5 seq)	
KC242793.1	KC242795.1
KC242794.1	
KC242795.1	
KC242797.1	
KC242798.1	
2002 (1 seq)	
KC242800.1	KC242800.1
2007 (7 seq)	
KC242784.1-90.1	KC242788.1
2014 (58 seq)	
KM233060-118	KM233097
2015 (20 seq)	
KT357813.1-	
833.1	KT357814.1
	KT357817.1
	KT357818.1
	KT357820.1
	KT357821.1
	KT357830.1

2.3. *Glycoprotein physicochemical properties*

Prediction of instability index (II), aliphatic index, theoretical isoelectric point (pI), and grand average of hydropathy (GRAVY) was done by "Expasy'sProtParam" [28] at http://expasy.org/tools/protparam.html. Also to compute the number of codons, bulkiness, polarity, refractivity, recognition factors, hydrophobicity, transmembrane tendency, percent buried residues, percent accessible residues, average area buried, average flexibility, relative mutability and number of amino acids; "PROTSCALE" [28] at http://us.expasy.org/tools/protscale.html was employed.

2.4. *Immunoinformatics analysis*

All methods of "immuneepitope" [29–33] http://tools.immuneepitope.org/tools/bcell/iedb_input online program were performed to find B-cell epitopes sites. "BcePred" [34] software at http://www.imtech.res.in/raghava/bcepred/was run

on sequences to detect B-cell epitopes based on polarity in addition to properties that were analyzed using previous server. "ABCpred" software [35] at http://www.imtech.res.in/raghava/abcpred/predicted 16 meric B-cell epitopes. The prediction of allergenic properties were done by "AlgPred" [36] at http://www.imtech.res.in/raghava/algpred/submission.html and protective antigens and subunit vaccines prediction was worked by "VaxiJen" software [37] at http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html.

2.5. *Disulfide bonds, glycosylation, and phosphorylation sites prediction*

Disulfide bonds positions were expected using "DiANNA" [38] at http://clavius.bc.edu/∼clotelab/DiANNA/, "SCRATCH" [39] at http://scratch.proteomics.ics.uci.edu/.

"DISPHOS" [40] at http://www.dabi.temple.edu/disphos/pred.html and "NetPhos" [41] at http://www.cbs.dtu.dk/services/NetPhos/ were used to predict phosphorylation sites.

Determination of kinase specific phosphorylation sites in eukaryotic proteins was calculated by "NetPhosK" [42] at http://www.cbs.dtu.dk/services/NetPhosK/. "NetNGlyc" [43] at http://www.cbs.dtu.dk/services/NetNGlyc/and "GlycoEP" [44] at http://www.imtech.res.in/raghava/glycoep/submit.html were employed for N-glycosylation sites and O-glycosylation sites prediction.

2.6. *Secondary structure prediction*

"SOPMA" [45] at http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html, "Phyre" server [46] at http://www.sbg.bio.ic.ac.uk/phyre, were reliable programs that we used for secondary structure prediction for selected sequences.

2.7. *Tertiary structure*

"I-TASSER" [47] at http://zhanglab.ccmb.med.umich.edu/I-TASSER, "Phyre^2" program [46] at http://www.sbg.bio.ic.ac.uk/∼phyre2/html, and "(PS)2-v2" Server [48] at http://ps2v2.life.nctu.edu.tw were utilized to predict tertiary structure of selected sequences. All predicted 3D structures were evaluated for the stereochemistry, reliability, and quality by "Qmean" [49] at http://swissmodel.expasy.org/qmean/cgi/index.cgi and "Rammpage" at http://mordred.bioc.cam.ac.uk/∼rapper/rampage.php.

3. Results

3.1. *Sequence analysis and phylogeny*

The sequences were analyzed and arranged by CLC sequence viewer. According to high similarity in amino acid sequences and based on e-values, 16 sequences were selected. We used abbreviations instead of isolated names for easier reading (Table 2).

Table 2. List of the abbreviations that were used in this study.

Abbreviations	Sequence number	Abbreviations	Sequence number
Seq Ref	NC_002549.1	Seq 2014	KM233097
Seq 1976	KM655246.1	Seq 2015 (1)	KT357820.1
Seq 1994	KC242792.1	Seq 2015 (2)	KT357818.1
Seq 1995	KC242796.1	Seq 2015 (3)	KT357821.1
Seq 1996	KC242795.1	Seq 2015 (4)	KT357814.1
Seq 2002	KC242800.1	Seq 2015 (5)	KT357830.1
Seq 2007	KC242788.1	Seq 2015 (6)	KT357817.1

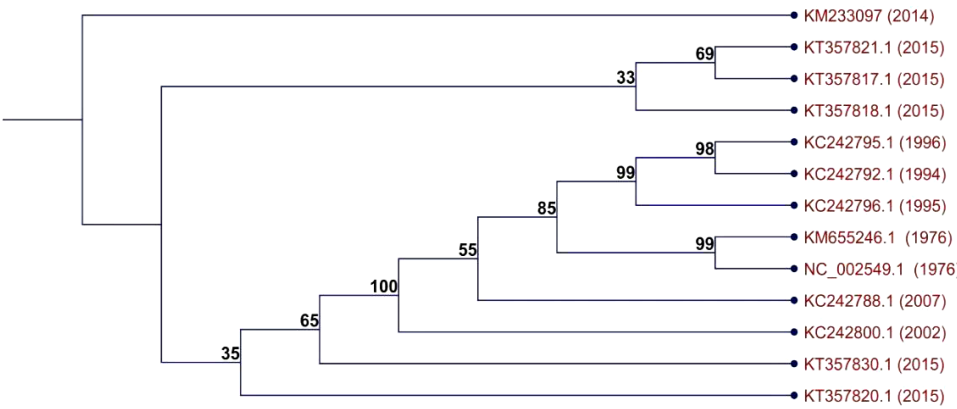

Fig. 1. Phylogenetic tree of 16 selected sequences by using neighbor-joining method.

The result of phylogenetic tree of sequences was displayed in Fig 1. Sixteen selected sequences were divided into two major clades that upper clade include only one sequence (KM233097). The second one divided into two major clusters that upper cluster included three sequences of 2015 and lower cluster includes other sequences and reference sequence.

Comparison analysis for 16 selected isolates showed multiple changes (Table 3). The major changes were observed in 82, 262, 310, 359, 378, 405, 411, 422, 430 and 441 sites.

3.2. *Physicochemical properties*

"Protparam" outcomes were summarized in Table 4. Theoretical pI range for the selected sequences was between 5.9–6.3 that determined this protein can be an acidic protein.

"Protscale" results shown similarity for all sequences, except in polarity that showed different positions in some sequences. Seq 2014, Seq 2015 (1), Seq 2015 (2), and Seq 2015 (5) had similar maximum polarity position. Also Seq 2015 (3), Seq

Table 3. Mutations sites for 15 sequences compared with Ref sequence.

Mutations	Ref Seq	1976	1994	1995	1996	2002	2007	2014	2015	2015 (2)	2015 (3)	2015 (4)	2015 (5)	2015 (6)
75 (V- A)										*				
82 (A-V)								*	*	*	*	*	*	*
262 (T-A)						*	*	*	*	*	*	*	*	*
310 (V-A)			*	*	*	*	*	*	*	*	*	*	*	*
314 (G -R)			*	*	*									
315(A -P)									*	*	*	*	*	*
317 (N-D)						*								
331(G-E)						*	*	*	*	*	*	*	*	*
336 (T-N)									*	*	*	*	*	*
342 (M-V)							*							
355 (N-S)						*								
359 (E-K)								*	*	*	*	*	*	*
366 (T-L)						*								
368 (L-P)					*									
376 (Q-R)			*											
377 (S-P)			*	*			*	*						
378 (L-P)			*	*		*	*	*	*	*	*	*	*	*
379 (T-I)			*											
380 (T-I)										*				
382 (P-T)								*						
384 (P-Q)						*								
389 (H-Y)							*							
398 (I-T)							*							

(*Continued*)

Table 3. (*Continued*)

Mutations	Ref Seq	1976	1994	1995	1996	2002	2007	2014	2015	2015 (2)	2015 (3)	2015 (4)	2015 (5)	2015 (6)
405 (E-G)								*	*	*	*	*	*	*
410 (R-S)											*	*		*
411 (T-A)								*	*	*	*	*	*	*
414 (D-A)			*		*						D-G			
417 (A-T)			*		*									
422 (S-P)			*	*	*	*	*	*	*	*	*	*	*	*
429 (P-H)							*							
430 (P-L)			*	*	*			*	*	*	*	*	*	*
435 (T-I)						*								
439 (K-E)											*	*		
440 (S-G)			*	*	*									
441(T-A)						*	*	*	*	*	*	*	*	*
443 (F-S)						*	*	*	*	*	*	*	*	*
446 (P-L)								*	*	*	*	*	*	*
453 (Q-P)							*							
455 (H-Y)						*								
474 (A-T)					*									
475 (S-G)						*								
483 (T-A)						*								
499 (T-A)				*										
503 (A-V)								*	*	*	*	*	*	*
539 (K-I)													*	
544 (I-T)			*	*	*		*							

Table 4. Physicochemical properties of all selected sequences by using Protparam. Seq 2002 showed different statement in some properties.

Properties	pI	Half-life in *E. coli*	Instability index	Class	Aliphatic index	GRAVY
Seq Ref	6.16	10 h	38.36	Stable	75.77	−0.38
Seq 1976	6.16	10 h	38.36	Stable	75.77	−0.38
Seq 1994	6.43	10 h	36.94	Stable	76.64	−0.379
Seq 1995	6.24	10 h	37.87	Stable	75.64	−0.393
Seq 1996	6.33	10 h	37.14	Stable	76.07	−0.385
Seq 2002	**5.97**	10 h	**40.79**	**Unstable**	76.08	−0.383
Seq 2007	6.24	10 h	37.09	Stable	74.48	−0.406
Seq 2014	6.3	10 h	37.45	Stable	76.35	−0.378
Seq 2015 (1)	6.3	10 h	37.45	Stable	76.35	−0.378
Seq 2015 (2)	6.3	10 h	37.2	Stable	76.64	−0.374
Seq 2015 (3)	6.12	10 h	37.32	Stable	76.35	−0.367
Seq 2015 (4)	6.04	10 h	37.83	Stable	76.35	−0.372
Seq 2015 (5)	6.21	10 h	37.45	Stable	76.78	−0.366
Seq 2015 (6)	6.21	10 h	37.26	Stable	76.35	−0.372

2015 (4) and Seq 2015 (6) had similar polarity positions. A cleavage site was found after amino acid 32 for all sequences whereas it was a conserve site. "SignalP" did not find any site for GP but predisi and Signal-BLAST were predicted a position for signal peptide (aa1–aa32).

3.3. *Immunoinformatics analysis*

Results of five methods for B-cell epitopes prediction by "immuneepitope" were summarized in Table 5 that showed similarity among all sequences by using surface accessibility and antigenicity methods. But sequences showed different positions by beta-turn, flexibility, and hydrophilicity methods. Results of "Bepipred", "BcePred", "ABCpred", "VaxiJen" and "AlgPred" were listed in Table 6.

"Bepipred" results were determined same positions for all sequences. Also prediction by "BcePred" showed six potent B-cell epitopes positions that were found in all selected sequences. "ABCpred" established two start positions (479 and 406) for 16 meric B-cell epitopes in majority of sequences; start positions were 479 and 319 for Seq 1994 and Seq 1996; 79 and 406 for Seq 2002. Analysis by "VaxiJen" concluded all sequences were probable antigen. Using "AlgPred", we found Seq Ref, Seq 1976, Seq 1994, Seq 1995, and Seq 1996 were non-allergen proteins but other sequences were predicted as allergen proteins.

3.4. *Post-modification prediction*

Using "SCRATCH", five disulfide bonds were found in all sequences, but it was different from the result of "DiANNA" which found seven positions for disulfide bonds that were different in selected sequences (Table 7).

"DISPHOS" could not find any phosphorylation sites, except one serine (aa418) in three sequences (Seq 1994, Seq 1995 and Seq 1996); however, "NetPhos" found

Table 5. B-cell epitopes prediction by using immuneepitope and based on five methods.

Sequences	Chou and Fasman beta-turn prediction	Emini surface accessibility prediction	Karplus and Schulz flexibility prediction	Kolaskar and Tongaonkar antigenicity	Parker hydrophilicity prediction
Seq Ref	327–333, 381–388, 636–643	405–414, 496–502	262–268, 318–325, 434–442	180–186, 664–669, 670–676	636–642, 467–473, 409–419
Seq 1976	327–333, 381–388, 636–643	405–414, 496–502	262–268, 318–325, 434–442	180–186, 664–669, 670–676	636–642, 467–473, 409–419
Seq 1994	327–333, 381–388, 636–643	405–414, 496–502	262–268, 318–325, 434–442	180–186, 664–669, 670–676	636–642, 467–473, 409–419
Seq 1995	327–333, 381–388, 636–643	405–414, 496–502	262–268, 318–325, 434–442	180–186, 664–669, 670–676	636–642, 467–473, 409–419
Seq 1996	327–333, 381–388, 636–643	405–414, 496–502	262–268, 318–325, 434–442	180–186, 664–669, 670–676	636–642, 467–473, 409–419
Seq 2002	327–333, 381–388, 636-643	405–414, 496–502	318–325, 372–379, 380–387	180–186, 664–669, 670–676	636–642, 467–473, 409–419
Seq 2007	327–333, 381–388, 636–643	405–414, 496–502	264–271, 318–325, 434–440	180–186, 664–669, 670–676	636–642, 467–473, 409–419
Seq 2014	208–215, 381–387, 636–645	405–414, 496–502	264–271, 318–325, 434–440	180–186, 664–669, 670–676	412–420, 332–338, 636–643
Seq 2015 (1)	208–215, 381–387, 636–645	405–414, 496–502	264–271, 318–325, 434–440	180–186, 664–669, 670–676	412–420, 332–338, 636–643
Seq 2015 (2)	208–215, 381–387, 636–645	405–414, 496–502	264–271, 318–325, 434–440	180–186, 664–669, 670–676	412–420, 332–338, 636–643
Seq 2015 (3)	208–215, 381–387, 636–645	405–414, 496–502	264–271, 318–325, 434–440	180–186, 664–669, 670–676	331–338, 433–467, 636–643
Seq 2015 (4)	208–215, 381–387, 636–645	405–414, 496–502	264–271, 318–325, 434–440	180–186, 664–669, 670–676	410–473, 332–338, 636–643
Seq 2015 (5)	208–215, 381–387, 636–645	405–414, 496–502	264–271, 318–325, 434–440	180–186, 664–669, 670–676	410–473, 332–338, 636–643
Seq 2015 (6)	208–215, 381–387, 636–645	405–414, 496–502	264–271, 318–325, 434–440	180–186, 664–669, 670–676	410–473, 332–338, 636–643

Table 6. Bepipred, BcePred, ABCpred, VaxiJen and AlgPred results for all selected sequences.

	Bepipred, Threshold: 1.5	BcePred	ABCpred	VaxiJen		AlgPred
Seq Ref	204–212, 323–336, 378–389, 413–438, 448–460, 467–474, 636–642	11–16, 262–269, 292–305, 322–338, 382–389, 408–422, 431–439, 450–475	406 and 479	0.4919	antigen	NON-ALLERGEN
Seq 1976	*	*	406 and 479	0.4919	antigen	NON-ALLERGEN
Seq 1994	*	*	319 and 479	0.4951	antigen	NON-ALLERGEN
Seq 1995	*	*	406 and 479	0.5005	antigen	NON-ALLERGEN
Seq 1996	*	*	319 and 479	0.5038	antigen	NON-ALLERGEN
Seq 2002	*	*	79 and 406	0.4871	antigen	**ALLERGEN**
Seq 2007	*	*	406 and 479	0.4834	antigen	**ALLERGEN**
Seq 2014	*	*	406 and 479	0.5037	antigen	**ALLERGEN**
Seq 2015 (1)	*	*	406 and 479	0.5037	antigen	**ALLERGEN**
Seq 2015 (2)	*	*	406 and 479	0.5052	antigen	**ALLERGEN**
Seq 2015 (3)	*	*	406 and 479	0.4999	antigen	**ALLERGEN**
Seq 2015 (4)	*	*	406 and 479	0.4978	antigen	**ALLERGEN**
Seq 2015 (5)	*	*	406 and 479	0.5082	antigen	**ALLERGEN**
Seq 2015 (6)	*	*	406 and 479	0.4979	antigen	**ALLERGEN**

Table 7. Disulfide bond predicted positions for selected sequences by using SCRATCH and DiANNA online softwares. Amino acids 53, 108, 121, 135, 147 and 609 were predicted as the most involved positions in disulfide bonds.

	SCRATCH					DiANNA						
Sequences	53–108	135–147	511–556	601–608	609–670	53	108	121	147	511	556	672
Seq Ref	*	*	*	*	*	53–672	108–670	121–135	147–608	511–601	556–609	
Seq 1976	*	*	*	*	*	53–672	108–670	121–135	147 -608	511–601	556–609	
Seq 1994	*	*	*	*	*	53–601	108–511	121–135	147–670		556–608	609–672
Seq 1995	*	*	*	*	*	53–672	108–609	121–135	147–670	511–601	556–608	
Seq 1996	*	*	*	*	*	53–672	108–670	121–135	147–608	511–609	556–601	
Seq 2002	*	*	*	*	609–672	53–601	108–609	121–135	147–670	511–672	556–608	
Seq 2007	*	*	*	*	609–672	53–601	108–609	121–135	147–670	511–672	556–608	
Seq 2014	*	*	*	*	*	53–601	108–609	121–135	147 -670	511–672	556–608	
Seq 2015 (1)	*	*	*	*	*	53–601	108–609	121–135	147–670	511–672	556–608	
Seq 2015 (2)	*	*	*	*	*	53–672	108–670	121–135	147–608	511–609	556–601	
Seq 2015 (3)	*	*	*	*	*	53–601	108–511	121–135	147–608		556–609	670–672
Seq 2015 (4)	*	*	*	*	*	53–672	108–670	121–135	147–608	511–601	556 -609	
Seq 2015 (5)	*	*	*	*	*	53–672	108–670	121–135	147–608	511–609	556–601	
Seq 2015 (6)	*	*	*	*	*	53–601	108–609	121–135	147–670	511–672	556–608	

Table 8. Results of phosphorylation sites prediction by DISPHOS, NetPhos and NetPhosK programs. NetPhosK software determined three phosphorylation positions which were similar in all sequences.

Sequences	DISPHOS			NetPhos 2.0 server	NetPhosK, Threshold: 0.85 PKC
Seq Ref	ser:0	thr:0	tyr:0	Ser: 27 Thr: 22 Tyr: 6	168–216–298
Seq 1976	ser:0	thr:0	tyr:0	Ser: 27 Thr: 22 Tyr: 6	168–216-298
Seq 1994	ser:1	thr:0	tyr:0	Ser: 23 Thr: 23 Tyr: 6	168–216–298
Seq 1995	ser:0	thr:0	tyr:0	Ser: 24 Thr: 22 Tyr: 6	168–216–298
Seq 1996	ser:1	thr:0	tyr:0	Ser: 22 Thr: 23 Tyr: 6	168–216–298
Seq 2002	ser:0	thr:0	tyr:0	Ser: 24 Thr: 24 Tyr: 8	168–216–298–263
Seq 2007	ser:0	thr:0	tyr:0	Ser: 28 Thr: 23 Tyr: 7	168–216–298–263
Seq 2014	ser:0	thr:0	tyr:0	Ser: 26 Thr: 20 Tyr: 7	168–216–298–263
Seq 2015 (1)	ser:0	thr:0	tyr:0	Ser: 26 Thr: 20 Tyr: 7	168–216–298–263
Seq 2015 (2)	ser:0	thr:0	tyr:0	Ser: 26 Thr: 19 Tyr: 7	168–216–298–263
Seq 2015 (3)	ser:0	thr:0	tyr:0	Ser: 26 Thr: 18 Tyr: 7	168–216–298–263
Seq 2015 (4)	ser:0	thr:0	tyr:0	Ser: 26 Thr: 19 Tyr: 7	168–216–298–263
Seq 2015 (5)	ser:0	thr:0	tyr:0	Ser: 25 Thr: 20 Tyr: 7	168–216–298–263
Seq 2015 (6)	ser:0	thr:0	tyr:0	Ser: 26 Thr: 20 Tyr: 7	168–216–298–263

numerous phosphorylation sites. Using "NetPhosK", three phosphorylation sites were found for Protein Kinase C (PKC), also this program found one extra position in eight sequences (Table 8).

Four N-linked glycosylation sites were predicted by "NetNGlyc" for all sequences; also several positions (238, 257, 317, 333, 386, 462 and 563) were calculated by "GlycoEP". To find O-linked glycosylation sites, "GlycoEP" was used and numerous sites were predicted. Finally, different patterns among selected sequences was found (Table 9).

3.5. *Secondary structure*

Percentages of secondary structure constituents for all selected sequences were summarized in Table 10 and schematic displayed in Fig. 2. According to the analysis, the majority of secondary structures of GP were random coil and the percentage of extended strand and alpha helix was equal.

3.6. *Tertiary structures and validation*

Tertiary structures of 16 selected sequences were predicted by three well-known softwares. The results of each program were validated by "Q-mean software" that showed "I-TASSER" is better software to detect tertiary structure of EBOV GP (Fig. 3). For final structure, Q-score and Z-score were 4.3 and −3.1, respectively. Ramachandran plots by Rammpage for 3D-structure were presented in Fig. 4. The average of residues in favored region and residues in allowed region were 446 (66.2%) and 165 (24.5%).

Table 9. N-link and O-link glycosylation sites were predicted by NetNGlyc and GlycoEP. Results showed a number of glycosylation sites for selected sequences.

	NetNGlyc							GlycoEP N-link	GlycoEP O-link									
Sequences	40	204	238	257	317	386	454	238–257–317–333–386–462–563	326	327	380	420	425	449	373	374	372	450
Seq Ref	*	*	*	*	*	*	*	*	*	*	*	*	*	*				
Seq 1976	*	*	*	*	*	*	*	*	*	*	*	*	*	*				
Seq 1994	*	*	*	*	*	*	*	*	*	*	*	*	*	*				
Seq 1995	*	*	*	*	*	*	*	*	*	*	*	*	*	*				
Seq 1996	*	*	*	*	*	*	*	*	*	*	*	*	*	*				
Seq 2002	*	*	*	*		*	*	*	*			*	*	*				
Seq 2007	*	*	*	*	*	*	*	*			*		*	*	*			*
Seq 2014	*	*	*	*	*	*	*	*			*	*	*			*	*	
Seq 2015 (1)	*	*	*	*	*	*	*	*			*	*	*			*	*	
Seq 2015 (2)	*	*	*	*	*	*	*	*				*	*					
Seq 2015 (3)	*	*	*	*	*	*	*	*			*	*	*			*	*	
Seq 2015 (4)	*	*	*	*	*	*	*	*			*	*	*			*	*	
Seq 2015 (5)	*	*	*	*	*	*	*	*			*	*	*			*	*	
Seq 2015 (6)	*	*	*	*	*	*	*	*			*	*	*			*	*	

Table 10. The percentage of secondary structure for all selected sequences that showed the majority of structure was random coil.

Sequences	Alpha helix (%)	Extended strand (%)	Beta turn (%)	Random coil (%)
Seq Ref	22.93	24.70	8.58	43.79
Seq 1976	22.93	24.70	8.58	43.79
Seq 1994	24.26	23.52	8.14	44.08
Seq 1995	23.82	23.96	8.28	43.93
Seq 1996	24.26	23.52	8.14	44.08
Seq 2002	23.82	23.96	8.28	43.93
Seq 2007	22.93	23.96	8.28	44.82
Seq 2014	22.93	25.15	7.99	43.93
Seq 2015 (1)	22.93	25.15	7.99	43.93
Seq 2015 (2)	23.08	25.15	7.84	43.93
Seq 2015 (3)	23.08	25.15	7.84	43.93
Seq 2015 (4)	23.08	25.15	7.84	43.93
Seq 2015 (5)	22.34	25.44	8.14	44.08
Seq 2015 (6)	22.78	25.15	7.99	44.08

4. Discussion

Generally, EBOV species are determined by considering position (498–501), if it is R-T/A-R-R the subtype is surly Zaire. For all selected sequences this position was RTRR except in Seq (1996) that was R-A-R-R.

According to alignment analysis, we found a large number of mutations in selected sequences. In comparison with other years, 2015 sequences showed more variety of mutations. Mentioned mutations can be effective in structural and functional property of GP and resistance to treatment in EBOV infections.

Successive human-to-human transmission events and adaptation to the human host were considered as the causes of the increase in the number of mutation in recent years [50].

The pI prediction showed GP is an acidic protein as there are a large number of acidic amino acids in GP structure. The half-time of stability of recombinant protein in *E. coli* as a host was more than 10 h. II showed an estimate of the stability of protein in a test tube, this index was around 37–38 for all sequences except Seq 2014, and confirmed majority of selected sequences are stable. Stability of this protein was established by Lingemann *et al.* that showed expression of GP protein in the respiratory tract of African green monkeys [51]. In addition, Marzi *et al.* in 2016 showed the stability of GP in 7 in RhCMV/EBOV-GP infected primary rhesus fibroblasts (RFs) until at least passage 7 that was related to structural features especially disulfide bonds [52].

Aliphatic index relates to amino acids with aliphatic side chains (alanine, valine, isoleucine and leucine), and determines thermostability of proteins and it is related to high a-helical content of GP protein that has profound impact on folding and unfolding processes [53].

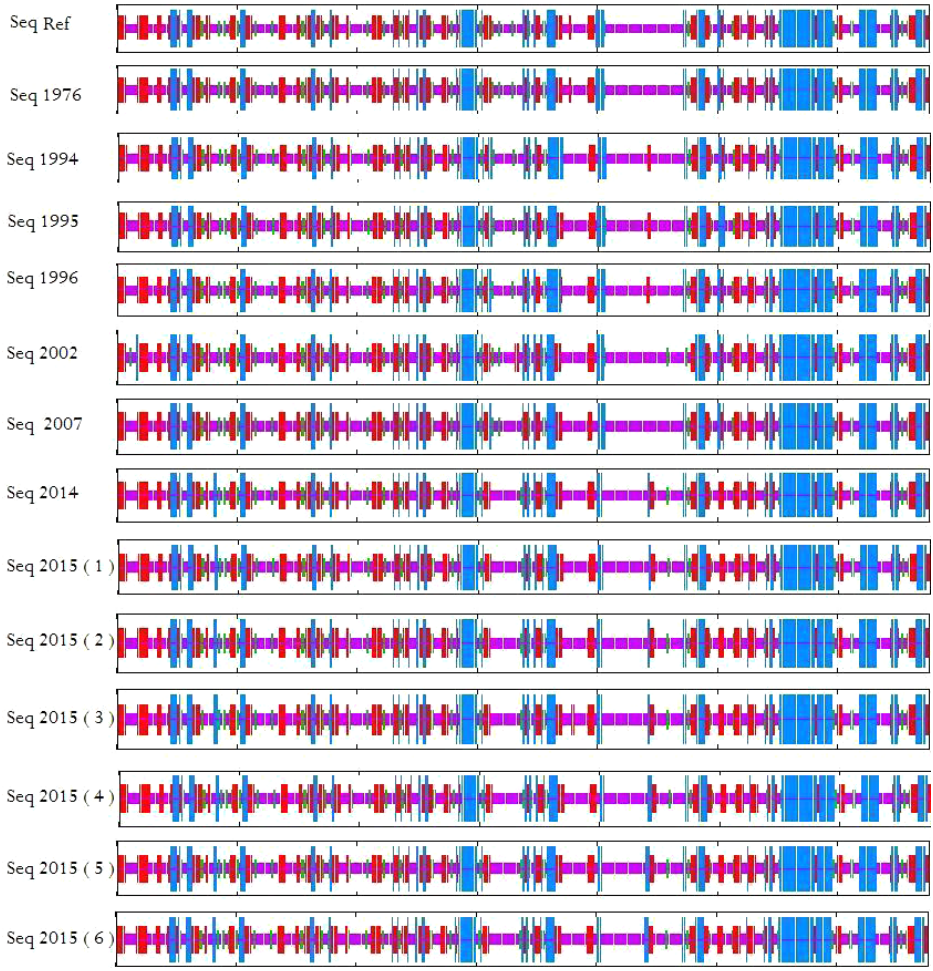

Fig. 2. (Color online) Secondary structure prediction for selected sequences, Red region is extended strand, Blue is alpha helix, Green is beta turn, and Purple is random coil.

Aliphatic index was more than 75 for all sequences, so they were considered as stable proteins. GRAVY was around -3 and -4, so all proteins were the most hydrophobic proteins.

Lee and Saphire in 2009 described four discontinuous sections (residues 33–69, 95–104, 158–167 and 176–189) that form a hydrophobic, semicircular surface also, GP_2 that contains the hydrophobic internal fusion loop, which is responsible for fusion of the viral and host cell membranes [54].

In most parts of "Protparam" results, Seq 2002 showed different properties that may be due to some individual mutations in this sequence [317 (N to D), 355 (N to S), 366 (T to L), 384 (P to Q), 435 (T to I), 455 (H to Y), 475 (S to G), 483 (T to A)].

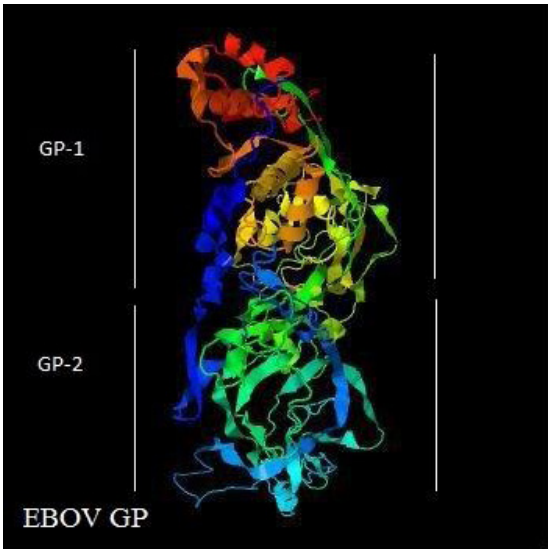

Fig. 3. 3D structure of EBOV GP by using I-TASSER program. Our results showed that I-TASSER could provide more reliable structure compared to other software.

According to our results, we found six cysteine positions (53, 108, 121, 135, 147 and 609) for GP protein. Disulfide bonds play an important role in stability of GP protein and assembling a heterodimer (GP1 + GP2) at the surface of nascent virions [54]. Barrientos *et al.* in 2004 found six cysteines for GP (53, 306, 135, 108, 121 and 147) [55]. Also Falzarano *et al.* in 2006 determined cysteines 53 and 306 as conserve positions for disulfide bonds [56]. Cys-53 was confirmed as disulfide bond by Jeffers *et al.* [57]. In addition, Lee and Saphire mentioned two cysteines (53, 609) for GP protein [54].

Phosphorylation of EBOV proteins is vital to regulate transcription and replication cycle. It was demonstrated that ectodomain of GP is phosphorylated, and serine residues between amino acid 260 and 273 were predicted as the main region for phosphorylation. There were three conserved phosphorylation positions for PKC and there was one more position which was serine 263 [58]. This serine existed in all sequences, but this position was considered as a phosphorylation site just in sequences with substitution in aa262 (threonine–alanine). So this mutation has the effect of recognizing aa263 as a phosphorylation site.

Prediction of phosphorylation sites by "DISPHOS" presented just one serine for two selected sequences (Seq 1994, Seq 1996). All sequences had a conserve serine in aa418 but only in two sequences this serine was predicted as a phosphorylation site. This situation may be due to mutation in aa417 (alanine–threonine) and the effect of this mutation in phosphorylation of aa418.

Previous studies found numerous functions for GP glycosylation. This process is necessary for conformational integrity and correct folding of GP [59].

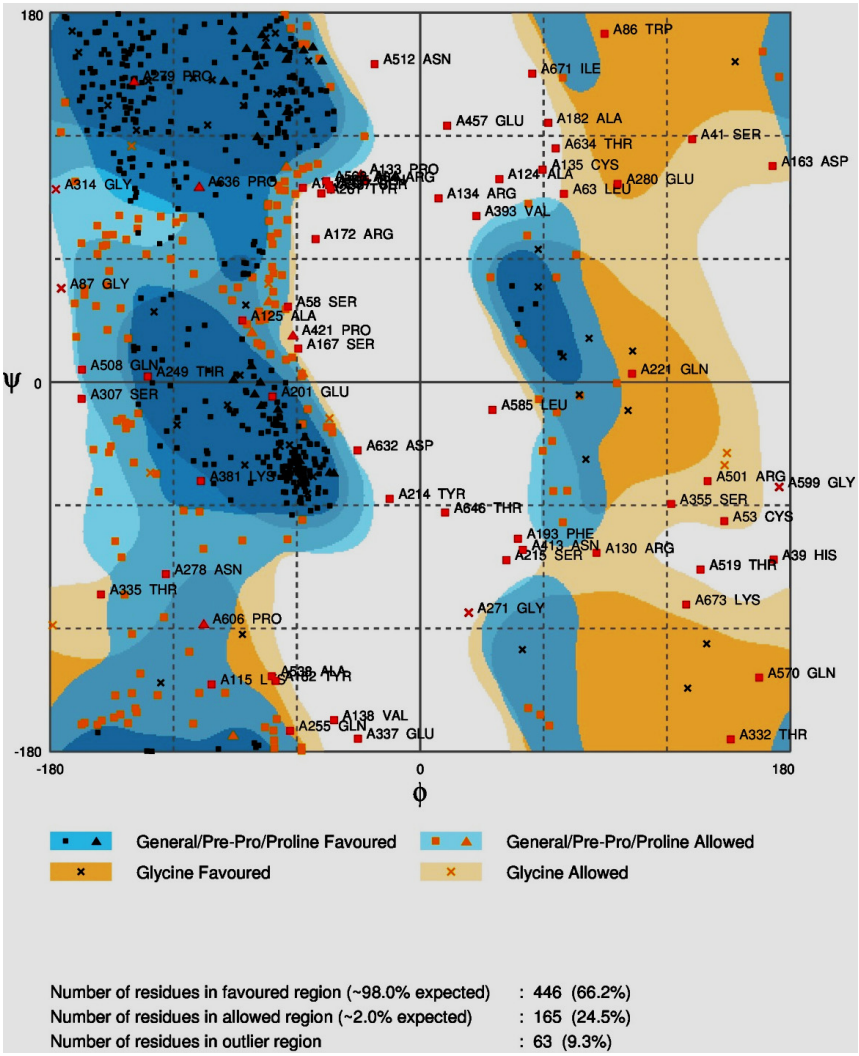

Fig. 4. Results of Ramachandran plot for GP 3D structure by Rammpage online software. The most number of residues were in favored region.

There are some glycosylation sites in GP. Glycosylation masks epitopes which can lead to neutralizing protective epitopes. Combining all ideas about GP glycosylation indicated this modification has an important role in the protection of virus from the host immune response [60–62].

Jeffers *et al.* in 2002 found eight positions (40, 204, 238, 257, 277, 296 and 563) for glycosylation [57]. Also Falzarano *et al.* in 2006 determined six N-linked glycosylation sites (40, 204, 228, 238, 257 and 268) for GP [56]. N-linked glycosylation sites prediction were done using two reliable programs which showed four conserve positions for all sequences (40, 204, 238 and 257).

In Asn-Xaa-Ser/Thr sequon, Asn is glycan, aa317 in all sequences was N (Asn) except in Seq 2002 (asparagine to aspartic acid). So this position was omitted in Seq 2002. O-linked glycosylation prediction showed several positions for selected sequences. Results verified different O-linked glycosylation sites in selected sequences as mutation occurred exactly in that site or around the glycosylation sites. The Seq Ref showed six O-linked glycosylation sites; however, Seq 2015 (2) showed just two sites. aa326 and aa327 in both sequences were threonine and serine; these sites were predicted as glycosylation sites for Seq Ref but not for Seq 2015 (2). We found a mutation in aa331 in Seq 2015 (2) that was the nearest mutation to these sites. Also aa449 was a glycosylation site for Seq Ref but not for Seq 2015 (2), and the nearest site was in aa446. In Seq 2015 (2), we found a mutation (threonine–isoleucine) in aa380, so this position was omitted in this sequence. We found an important mutation in aa262 (threonine–alanine) in some selected sequences that made a potent IgE epitope and changed them to allergen proteins. We suggest this adverse effect must be considered in further Ebola GP DNA vaccine. Some potential regions for antibody epitopes of GP were found by several studies.

Also it was demonstrated that GP-specific monoclonal antibodies or polyclonal antiserum could provide considerable protection against EBOV infection in mice. Animal model studies demonstrated that both cellular and humoral immunities can protect models against EBOV infection. Based on animal model researches, IgG response was suggested for recovery in human survivors. Wilson *et al.* in 2000 found three potent epitopes (401–417, 389–405 and 447–493) for GP by using GP monoclonal antibodies that all were protective against EBOV [63]. In 2007, Shahhosseini *et al.* determined that aa311–369 was recognized by majority of monoclonal antibodies; it suggested the high immunogenicity of this region [64]. Also in 2007, Takada *et al.* found four epitopes for GP (171–190, 331–350, 391–410 and 462–501) [65]. Bale *et al.* in 2012 found five regions (32–40, 42–52, 550–564, 503–556) that have interaction with 16F6 and KZ52 and two novel Fabs [66]. In 2014, Becquart *et al.* found five linear immunogenic regions (41–67, 93–127, 201–251, 301–359 and 381–411) for GP by ELISA method [67]. In 2014, Wang *et al.* tested a fragment (393–556) derived from ZEBOV GP which could induce humoral immune response strongly [62].

Among different regions, these five regions (32–67, 311–369, 380–417, 450–501 and 503–569) were the most reported. According to previous research and this study, four potent B-cell epitopes were found in GP (380–387, 318–338, 405–438 and 434–475).

Our analysis proved that humoral immunity can be effective against Ebola infection and maybe this is an approach to designing a potent vaccine. Secondary structure prediction showed that the major part of GP was random coil and after that alpha helix and extended strand had high percentages. The majority of B-cell epitopes, N-linked glycosylation and phosphorylation positions were placed in random coil. Among disulfide bond positions, aa53 and aa609 were placed in alpha helix,

and aa135 and aa609 were placed in strand and other sites were located in coil. Tertiary structure by using several programs was mapped and results established that I-TASSER is the most reliable software for 3D-structure prediction. A reliable tertiary structure is needed to identify interaction between GP and membrane compounds.

5. Conclusion

This research determined 10 major substitution sites for GP in the course of 40 years that caused many changes in modification sites, epitopes and structure of ZEBOV GP. Our results found five serine amino acids (418, 168, 216, 298 and 263) as phosphorylation sites for GP and four potent B-cell epitopes. Six cysteines were the most common positions for disulfide bonds. Prediction of O-linked and N-linked glycosylation showed 10 amino acids with high ranks for glycosylation. A region (259–271, TIYASGKRSNTTG) was considered as IgE epitope for some sequences that had a mutation (aa262, T to A); this region in reference sequence was TIYTSGKRSNTTG, and this instance explained the effect of just one mutation in a huge change in an indicator of protein. An overall result is the collection of comprehensive data about GP that can be useful for further studies in Ebola infection and designing a potent vaccine.

Acknowledgments

The authors would like to acknowledge Shiraz University of Medical Sciences for support (Project Code: 95-01-59-13084) and Professor Jean Oliver Krugman for manuscript editing.

References

[1] Y. Suzuki and T. Gojobori, The origin and evolution of Ebola and Marburg viruses, *Molec. Biol. Evol.* **14**(8) (1997) 800–806.
[2] M. Kiley *et al.*, Filoviridae: A taxonomic home for Marburg and Ebola viruses? *Intervirology* **18**(1–2) (1982) 24–32.
[3] H. Feldmann, C. Will, M. Schikore, W. Slenczka and H.-D. Klenk, Glycosylation and oligomerization of the spike protein of Marburg virus, *Virology* **182**(1) (1991) 353–356.
[4] K. Johnson, J. Lange, P. Webb and F. Murphy, Isolation and partial characterisation of a new virus causing acute haemorrhagic fever in Zaire, *Lancet* **309**(8011) (1977) 569–571.
[5] Team RoaWIS, Ebola haemorrhagic fever in Sudan, 1976, *Bull. World Health Organ.* **56**(2) (1978) 247.
[6] E. M. Leroy *et al.*, Fruit bats as reservoirs of Ebola virus, *Nature* **438**(7068) (2005) 575–576.
[7] G. Chowell and H. Nishiura, Transmission dynamics and control of Ebola virus disease (EVD): A review, *BMC Med.* **12**(1) (2014) 196.
[8] D. Gatherer, The 2014 Ebola virus disease outbreak in West Africa, *J. Gen. Virol.* **95**(8) (2014) 1619–1624.

[9] N. K. Jaax *et al.*, Lethal experimental infection of rhesus monkeys with Ebola-Zaire (Mayinga) virus by the oral and conjunctival route of exposure, *Arch. Pathol. Lab. Med.* **120**(2) (1996) 140–155.
[10] S. Schou and A. K. Hansen, Marburg and Ebola virus infections in laboratory non-human primates: A literature review, *Comparative Med.* **50**(2) (2000) 108–123.
[11] E. Mühlberger, Filovirus replication and transcription, *Future Viral.* **2**(2) (2007) 205–215.
[12] H. Feldmann, H.-D. Klenkand and A. Sanchez, Molecular biology and evolution of filoviruses, in *Unconventional Agents and Unclassified Viruses* (Springer, 1993), pp. 81–100.
[13] T. Hoenen *et al.*, Infection of naive target cells with virus-like particles: Implications for the function of Ebola virus VP24, *J. Virol.* **80**(14) (2006) 7260–7264.
[14] Y. Huang, L. Xu, Y. Sun and G. J. Nabel, The assembly of Ebola virus nucleocapsid requires virion-associated proteins 35 and 24 and post-translational modification of nucleoprotein, *Molec. Cell.* **10**(2) (2002) 307–316.
[15] G. Simmons, R. J. Wool-Lewis, F. Baribaud, R. C. Netter and P. Bates, Ebola virus glycoproteins induce global surface protein down-modulation and loss of cell adherence, *J. Virol.* **76**(5) (2002) 2518–2528.
[16] V. E. Volchkov, H. Feldmann, V. A. Volchkova and H.-D. Klenk, Processing of the Ebola virus glycoprotein by the proprotein convertase furin, *Proc. Natl. Acad. Sci.* **95**(10) (1998) 5762–5767.
[17] A. Yonezawa, M. Cavrois and W. C. Greene, Studies of Ebola virus glycoprotein-mediated entry and fusion by using pseudotyped human immunodeficiency virus type 1 virions: Involvement of cytoskeletal proteins and enhancement by tumor necrosis factor alpha, *J. Virol.* **79**(2) (2005) 918–926.
[18] N. J. Lennemann, B. A. Rhein, E. Ndungo, K. Chandran, X. Qiu and W. Maury, Comprehensive functional analysis of N-linked glycans on Ebola virus GP1, *MBio.* **5**(1) (2014) e00862–13.
[19] D. Falzarano, T. W. Geisbert and H. Feldmann, Progress in filovirus vaccine development: Evaluating the potential for clinical use, *Exp. Rev. Vacc.* **10**(1) (2011) 63–77.
[20] C. Nunes-Alves, Viral infection: New options to fight Ebola, *Nat. Rev. Microbiol.* **13**(6) (2015) 327.
[21] N. Sullivan, Z.-Y. Yang and G. J. Nabel, Ebola virus pathogenesis: Implications for vaccines and therapies, *J. Virol.* **77**(18) (2003) 9733–9737.
[22] Y. Vidal, *How to Prevent the Spread of Ebola: Effective Strategies to Reduce Hospital Acquired Infections* (Lara Publications, 2015).
[23] J. R. Francica, A. Varela-Rohena, A. Medvec, G. Plesa, J. L. Riley and P. Bates, Steric shielding of surface epitopes and impaired immune recognition induced by the Ebola virus glycoprotein, *PLoS Pathog.* **6**(9) (2010) e1001098.
[24] A. Moattari, B. Dehghani, N. Khodadad and F. Tavakoli, *In silico* functional and structural characterization of H1N1 influenza a viruses hemagglutinin, 2010–2013, Shiraz, Iran, *Acta Biotheor.* **63**(2) (2015) 183–202.
[25] T. N. Petersen, S. Brunak, G. von Heijne and H. Nielsen, SignalP 4.0: Discriminating signal peptides from transmembrane regions, *Nat. Meth.* **8**(10) (2011) 785–786.
[26] K. Frank and M. J. Sippl, High-performance signal peptide prediction based on sequence alignment techniques, *Bioinformatics* **24**(19) (2008) 2172–2176.
[27] K. Hiller, A. Grote, M. Scheer, R. Münch and D. Jahn, PrediSi: Prediction of signal peptides and their cleavage positions, *Nucleic Acids Res.* **32**(Suppl. 2) (2004) W375–W379.

[28] E. Gasteiger *et al.*, *Protein Identification and Analysis Tools on the ExPASy Server* (Springer, 2005).
[29] P. Chou and G. D. Fasman, Amino acid sequence, *Adv. Enzymol. Relat. Areas Molec. Biol.* **47** (2009) 45.
[30] E. A. Emini, J. V. Hughes, D. Perlow and J. Boger, Induction of hepatitis a virus-neutralizing antibody by a virus-specific synthetic peptide, *J. Virol.* **55**(3) (1985) 836–839.
[31] P. Karplus and G. Schulz, Prediction of chain flexibility in proteins, *Naturwissenschaften* **72**(4) (1985) 212–213.
[32] J. E. Larsen, O. Lund and M. Nielsen, Improved method for predicting linear B-cell epitopes, *Immunome Res.* **2**(1) (2006) 2.
[33] J. Parker, D. Guo and R. Hodges, New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: Correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites, *Biochemistry* **25**(19) (1986) 5425–5432.
[34] S. Saha and G. Raghava (eds.), BcePred: Prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties, in *Int. Conf. Artificial Immune Systems* (Springer, 2004).
[35] S. Saha and G. Raghava, Prediction of continuous B-cell epitopes in an antigen using recurrent neural network, *Proteins Struct. Funct. Bioinform.* **65**(1) (2006) 40–48.
[36] S. Saha and G. Raghava, AlgPred: Prediction of allergenic proteins and mapping of IgE epitopes, *Nucleic Acids Res.* **34**(Suppl. 2) (2006) W202–W209.
[37] I. A. Doytchinova and D. R. Flower, VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines, *BMC Bioinform.* **8**(1) (2007) 4.
[38] F. Ferrè and P. Clote, DiANNA 1.1: An extension of the DiANNA web server for ternary cysteine classification, *Nucleic Acids Res.* **34**(Suppl. 2) (2006) W182–W185.
[39] J. Cheng, A. Z. Randall, M. J. Sweredoski and P. Baldi, SCRATCH: A protein structure and structural feature prediction server, *Nucleic Acids Res.* **33**(Suppl. 2) (2005) W72–W726.
[40] L. M. Iakoucheva *et al.*, The importance of intrinsic disorder for protein phosphorylation, *Nucleic Acids Res.* **32**(3) (2004) 1037–1049.
[41] N. Blom, S. Gammeltoft and S. Brunak, Sequence and structure-based prediction of eukaryotic protein phosphorylation sites, *J. Molec. Biol.* **294**(5) (1999) 1351–1362.
[42] N. Blom, T. Sicheritz-Pontén, R. Gupta, S. Gammeltoft and S. Brunak, Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence, *Proteomics* **4**(6) (2004) 1633–1649.
[43] R. Gupta and S. Brunak (eds.), Prediction of glycosylation across the human proteome and the correlation to protein function, *Pac. Symp. Biocomput.* **7**(1) (2001) 310–322.
[44] J. S. Chauhan, A. Rao and G. P. Raghava, *In silico* platform for prediction of N-, O- and C-glycosites in eukaryotic protein sequences, *PLoS One* **8**(6) (2013) e67008.
[45] C. Geourjon and G. Deleage, SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, *Comput. Appl. Biosci.* **11**(6) (1995) 681–684.
[46] L. A. Kelley and M. J. Sternberg, Protein structure prediction on the web: A case study using the Phyre server, *Nat. Protocols* **4**(3) (2009) 363–371.
[47] A. Roy, A. Kucukural and Y. Zhang, I-TASSER: A unified platform for automated protein structure and function prediction, *Nat. Protocols* **5**(4) (2010) 725–738.
[48] C.-C. Chen, J.-K. Hwang and J.-M. Yang, (Ps)2: Protein structure prediction server, *Nucleic Acids Res.* **34**(Suppl. 2) (2006) W152–W157.

[49] P. Benkert, S. C. Tosatto and D. Schomburg, QMEAN: A comprehensive scoring function for model quality assessment, *Proteins Struct. Funct. Bioinform.* **71**(1) (2008) 261–277.
[50] W. E. Diehl *et al.*, Ebola virus glycoprotein with increased infectivity dominated the 2013–2016 epidemic, *Cell* **167**(4) (2016) 1088–1098.
[51] M. Lingemann *et al.*, Attenuated human parainfluenza virus type 1 expressing Ebola virus glycoprotein GP administered intranasally is immunogenic in African Green Monkeys, *J. Virol.* **91**(10) (2017) e02469–16.
[52] A. Marzi *et al.*, Cytomegalovirus-based vaccine expressing Ebola virus glycoprotein protects nonhuman primates from Ebola virus infection, *Sci. Rep.* **6** (2016) 1–10.
[53] W. Weissenhorn, L. J. Calder, S. A. Wharton, J. J. Skehel and D. C. Wiley, The central structural feature of the membrane fusion protein subunit from the Ebola virus glycoprotein is a long triple-stranded coiled coil, *Proc. Natl. Acad. Sci.* **95**(11) (1998) 6032–6036.
[54] J. E. Lee and E. O. Saphire, Ebola virus glycoprotein structure and mechanism of entry, *Future Virol.* **4**(6) (2009) 621–635.
[55] L. G. Barrientos, A. M. Martin, P. E. Rollin and A. Sanchez, Disulfide bond assignment of the Ebola virus secreted glycoprotein SGP, *Biochem. Biophys. Res. Commun.* **323**(2) (2004) 696–702.
[56] D. Falzarano *et al.*, Structure-function analysis of the soluble glycoprotein, sGP, of Ebola virus, *Chem. Biochem.* **7**(10) (2006) 1605–1611.
[57] S. A. Jeffers, D. A. Sanders and A. Sanchez, Covalent modifications of the Ebola virus glycoprotein, *J. Virol.* **76**(24) (2002) 12,463–12,472.
[58] C. Sänger, E. Mühlberger, B. Lötfering, H.-D. Klenk and S. Becker, The Marburg virus surface protein GP is phosphorylated at its ectodomain, *Virology* **295**(1) (2002) 20–29.
[59] W. Dowling *et al.*, Influences of glycosylation on antigenicity, immunogenicity, and protective efficacy of Ebola virus GP DNA vaccines, *J. Virol.* **81**(4) (2007) 1821–1837.
[60] M.-A. de La Vega, G. Wong, G. P. Kobinger and X. Qiu, The multiple roles of sGP in Ebola pathogenesis, *Viral Immunol.* **28**(1) (2015) 3–9.
[61] S.-R. Jun *et al.*, Ebola virus comparative genomics, *FEMS Microbiol. Rev.* **39**(5) (2015) 764–778.
[62] Y. Wang, Z. Liu and Q. Dai, A highly immunogenic fragment derived from Zaire Ebola virus glycoprotein elicits effective neutralizing antibody, *Virus Res.* **189** (2014) 254–261.
[63] J. A. Wilson *et al.*, Epitopes involved in antibody-mediated protection from Ebola virus, *Science* **287**(5458) (2000) 1664–1666.
[64] S. Shahhosseini, D. Das, X. Qiu, H. Feldmann, S. M. Jones and M. R. Suresh, Production and characterization of monoclonal antibodies against different epitopes of Ebola virus antigens, *J. Virol. Meth.* **143**(1) (2007) 29–37.
[65] A. Takada, H. Ebihara, H. Feldmann, T. W. Geisbert and Y. Kawaoka, Epitopes required for antibody-dependent enhancement of Ebola virus infection, *J. Infect. Diseases* **196**(Suppl. 2) (2007) S347–S356.
[66] S. Bale *et al.*, Structural basis for differential neutralization of Ebola viruses, *Viruses* **4**(4) (2012) 447–470.
[67] P. Becquart, T. Mahlakõiv, D. Nkoghe and E. M. Leroy, Identification of continuous human B-cell epitopes in the VP35, VP40, nucleoprotein and glycoprotein of Ebola virus, *PLoS One* **9**(6) (2014) e96360.