Int. Res. J. of Science & Engineering, Volume 9 (3) 2021
ISSN: 2322-0015

SJIF Impact Factor 6.70

ORIGINAL ARTICLE OPEN ACCESS

Glycosylation pattern in the glycoproteins of the Makona variant of Ebola virus

Mahale Kiran[1]*, Patole Milind[1] and Walhe Rajan[2]*

[1]National Centre for Cell Science, SP Pune University Campus, Pune 411007, India.
[2]Department of Microbiology, MES Abasaheb Garware College, Pune 411004, India.
*Correspondence: kkmahale@gmail.com, rawalhe@rediffmail.com

Manuscript Details

Received :08.05.2020
Accepted: 29.06.2021
Published: 30.06.2021

Available online on https://www.irjse.in
ISSN: 2322-0015

Editor: Dr. Arvind Chavhan

Cite this article as:
Mahale Kiran, Patole Milind and Walhe Rajan. Glycosylation pattern in the glycoproteins of the Makona variant of Ebola virus, *Int. Res. Journal of Science & Engineering*, 2021, Volume 9(3): 71-76.

Abstract

Surface glycoprotein of viruses like Ebola virus are very important, since they are the targets of many vaccination strategies. The current study focuses on the glycoprotein sequences of the West African Makona variant (2104-16) of Ebola virus. The protein sequences of the glycoprotein of 1,184 genomes and their glycosylation potential showed that the diversity of the protein sequences in the 2014-16 outbreak was limited and did not disturb the glycosylation potential of these proteins to a great extent.

Keywords: Ebola Virus, Makona, Vaccine, N-glycosylation, O-glycosylation

1. Introduction

Ebola virus disease (EVD) is caused by ebolavirus (genus *Ebolavirus*, family *Filoviridae*) and causes severe viral haemorrhagic fever. Although it is a rare fever, it is highly fatal. The fatality can range from 25 to 90%, with an average fatality of nearly 50% (https://www.who.int/news-room/fact-sheets/detail/ebola-virus-disease). Due to its lethality, dearth of appropriate therapeutics and being bioweapon potential Ebola virus is classified as category A pathogen [1]. Although the disease was first seen in 1976 in the African continent with intermittent occurrences, the most devastating outbreak was seen in the 2014-16 epidemic of Ebola virus. The total death was 11,310 (recorded on 13-Apr-2016; http://www.cdc. gov/vhf/ ebola/outbreaks/2014-west-africa/case-counts.html).

This outbreak was caused by Ebola virus (EBOV; member of *Zaire ebolavirus* species), one of the 5 members of genus *Ebolavirus* [2]. The 5 members of the genus *Ebolavirus* are Taï Forest virus (TAFV; formerly Côte d'Ivoire), Reston virus (RESTV), Sudan virus (SUDV), Bundibugyo virus (BDBV) and Ebola virus (EBOV; member of *Zaire ebolavirus* species). A comparison of the genomic sequences of the 5 members of *Ebolavirus* has been carried out in a previous study [3]. This study is focused on the West African variant of the 2014-16 variant, the Makona variant. This variant was found to have spread in the three West African countries: Sierra Leone, Guinea and Liberia. The variant was named after the Makona river, bordering these three West African countries [4].

Genome:

The ebolavirus is a filamentous, enveloped, negative strand RNA virus [5]. The genome is linear and about 19kb in size. Genome organization of Ebolavirus is similar with that of rhabdoviruses and paramyxoviruses [6]. It has 7 protein coding genes lying in between the two terminal non-transcribed 3′ leader (*Ldr*) and 5′ trailer (*Tlr*) sequences. The 7 genes are *NP* (Nucleoprotein), *VP35* (Polymerase cofactor VP35), *VP40* (Matrix protein VP40), *GP* (coding for 3 glycoproteins: sGP, ssGP, preGP), *VP30* (Hexameric zinc-finger protein VP30), *VP24* (Membrane-associated protein VP24) and *L* (coding for RNA polymerase).

Viral glycoprotein (GP) mediates the entry of Ebolavirus to host cell [7]. The gene of interest for this study is *GP* gene, which codes for its surface glycoprotein. The glycoprotein is present in the viral envelope that helps in the entry of the virus into the host cell. Glycoprotein is important because it induces the host to produce neutralizing antibodies. This also makes these viral surface glycoproteins the most eligible vaccine candidates for a given virus. Human immunodeficiency Virus (HIV) and Simian immunodeficiency Virus (SIV) also enter host cells by GP mediated fusion reactions [8]. A remarkable thing about the transcription of *GP*-gene is its ribosomal slippage/transcriptional stuttering [9] leading to the synthesis of three GP proteins (described in detail in the Materials and Methods section; Table 1). The three proteins are sGP (small secreted protein, with 364 amino acids), ssGP (super small secreted glycoprotein, with 297 amino acids) and preGP protein (with 676 amino acids).

Glycosylation of GP proteins:

Enveloped viruses have evolved along with hosts and adapted in many ways to deal host cell defense [10]. Glycosylation is widespread post translational change for many important biological roles [11]. The glycoprotein has great potential for adding sugar moieties to its amino acids, thereby increasing the diversity of the protein. Any change in the sequence of the protein, especially the amino acids directly involved in glycosylation, lead to an additional diversity in its glycosylation.

The translation of GP-gene mRNAs occurs on polysomes attached to endoplasmic reticulum and the protein is processed in Golgi bodies, where it undergoes glycosylation by an enzymatic reaction [12]. Glycosylation is a reaction, wherein a carbohydrate (or glycan) is attached to a hydroxyl or other functional group of another molecule (a glycosyl acceptor) in order to form a glycoconjugate.

The well-studied motifs for N-glycosylation and C-mannosylation are N{P}[ST]{P} and WXXW motifs respectively [13,14]. The details of the pattern search in the glycoprotein of the virus is explained in the Materials and Methods section.

2. Materials and Method

A total of 1,184 genomes of the Makona variant (West African variant of 2014-16) of Ebola virus were selected for this study from the NCBI nucleotide database (https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=186536). The related details are mentioned in Table 2. The annotation information was used to extract the *GP*-gene nucleotide sequences from each of these genomes and was used as the primary source of information for further analysis The length of each of these nucleotide sequences is 2,030 nucleotides.

Note that the length of the sequence is not a triplet of 3. The virus employs a special mechanism known as ribosomal slippage to synthesize 3 different proteins from the same sequence. The *GP*-gene contains a polyadenosine stretch making it susceptible to ribosomal slippage.

Table 1. Transcripts and proteins of *GP*-gene

Protein	Frame	Transcripts	Transcript Length	Protein Length
sGP	Frame +1	1-1095	1095	364
ssGP	Frame +2	1-885 and 887-895	885 + 0009 = 0894	295 + 002 = 297
preGP	Frame -1	1-885 and 885-2030	885 + 1146 = 2031	295 + 381 = 676

Table 2. Nucleotide sequences and protein sequences of *GP*-gene from 1184 genomes; number of unique nucleotide sequences and protein sequences

Sequence Name	Number of unique sequences	Accession number of the major representative (no. of sequences represented)
GP-gene	168	KM233036.1 (551), KR817121.1 (105)
sGP	50	KM233036.1 (959)
ssGP	35	KM233036.1 (986)
preGP	101	KM233036.1 (781)

Table 3. Motifs for N-glycosylation and C-mannosylation

Motif	Position 1	Position 2	Position 3	Position 4
N{P}[ST]{P}	N Asparagine N-Glycosylation site	{P} Any amino acid, but not Proline	[ST] Serine or Threonine	{P} Any amino acid, but not Proline
WXXW	W Tryptophan C-mannosylation site	X Any amino acid	X Any amino acide	W Tryptophan

The unedited transcript (+1 frame) of the glycoprotein gene leaves 7 adenosine residues at the slippage site (near the 886th residue), and synthesizes a soluble form of the glycoprotein (sGP), which is 364 amino acids long. The slippage at +2 frame leads to missing out of the 886th adenosine residue and transcribes only the 1-885 residues and 887-895 residues. This slippage leads to an early encountering of a stop codon and leads to a smaller soluble glycoprotein (ssGP) of 297 amino acids. The slippage in -1 frame leads to the repeated transcription of the 885th residue leading to a transcription of 1-885 and 885-2030 and forming a 2031 long nucleotide transcript (a multiple of 3). This leads to a 676 long amino acid protein, known as preGP protein ($GP_{1,2}$; or virion spike protein). All the transcript generation were done using in-house Python version 2.8 scripts.

Glycosylation pattern: The protein sequences were scanned for the motif N{P}[ST]{P} and WXXW motif for verifying the N-glycosylation and C-mannosylation using in-house Python scripts. The details of the motif are described in Table 3.

3. Results and Discussion

Vigerust and Shepherd [15] have also reviewed the significant role of glycosylation of several pathogenic viruses on virulence and immune evasion. Iraqi et al [16] have studied the role of GP1 N- glycosylation on the protecting ligands to immune receptor.

Table 4: An overview of the glycosylation pattern (O-glycosylation, N-glycosylation, C-mannosylation) in the preGP protein of 10 representatives of the Makona variant of Ebola virus. Numbers in the brackets in the 2nd column represent the number of viral sequences that are identical to the representative.

Sr	preGP acc	[O-Glycosyl]	N-Glycosyl pattern (residue number of the 1st amino acid, followed by the sequence of tetrapeptide)	C-mannosyl
1	KT357814.1 (31)	48S + 71T = 119	040_NSTL, 204_NATE, 228_NETE, 238_NLTY, 257_NETI, 268_NTTG, 296_NLTR, 317_NISG, 333_NTTN, 346_NSSA, 386_NSTH, 413_NDST, 436_NTSE, 454_NYSE, 462_NNTH, 563_NETT, 618_NITD	288_WAFW, 645_WTGW, 648_WRQW
2	KM233036.1 (781)	47S + 71T = 118	040_NSTL, 204_NATE, 228_NETE, 238_NLTY, 257_NETI, 268_NTTG, 296_NLTR, 317_NISG, 333_NTTN, 346_NSSA, 386_NSTH, 413_NDST, 436_NTSK, 454_NYSE, 462_NNTH, 563_NETT, 618_NITD	288_WAFW, 645_WTGW, 648_WRQW
3	KJ660346.2 (43)	47S + 71T = 118	040_NSTL, 204_NATE, 228_NETE, 238_NLTY, 257_NETI, 268_NTTG, 296_NLTR, 317_NISG, 333_NTTN, 346_NSSA, 386_NSTH, 413_NDST, 436_NTSK, 454_NYSE, 462_NNTH, 563_NETT, 618_NITD	288_WAFW, 645_WTGW, 648_WRQW
4	KJ660348.2 (4)	47S + 71T = 118	040_NSTL, 204_NATE, 228_NETE, 238_NLTY, 257_NETI, 268_NTTG, 296_NLTR, 317_NISG, 333_NTTN, 346_NSSA, 386_NSTH, 413_NDST, 436_NTSK, 454_NYSE, 462_NNTH, 563_NETT, 618_NITD	645_WTGW, 648_WRQW
5	KR105301.1 (55)	47S + 71T = 118	040_NSTL, 204_NATE, 228_NETE, 238_NLTY, 257_NETI, 268_NTTG, 296_NLTR, 317_NISG, 333_NTTN, 346_NSSA, 386_NSTH, 413_NDST, 436_NTSK, 454_NYSE, 462_NNTH, 563_NETT, 618_NITD	288_WAFW, 645_WTGW, 648_WRQW
6	KR653304.1 (1)	48S + 71T = 119	040_NSTL, 204_NATE, 228_NETE, 238_NLTY, 257_NETI, 268_NTTG, 296_NLTR, 317_NISG, 333_NTTN, 346_NSSA, 386_NSTH, 413_NDST, 436_NTSK, 454_NYSE, 462_NNTH, 563_NETT, 618_NITD	288_WAFW, 645_WTGW, 648_WRQW
7	KP759733.1 (6)	47S + 71T = 118	040_NSTL, 204_NATE, 228_NETE, 238_NLTY, 257_NETI, 268_NTTG, 296_NLTR, 317_NISG, 333_NTTN, 346_NSSA, 386_NSTH, 413_NDST, 436_NTSK, 454_NYSE, 462_NNTH, 563_NETT, 618_NITD	288_WAFW, 645_WTGW, 648_WRQW
8	KP759759.1 (2)	47S + 71T = 118	040_NSTL, 204_NATE, 228_NETE, 238_NLTY, 257_NETI, 268_NTTG, 296_NLTR, 317_NISG, 333_NTTN, 346_NSSA, 386_NSTH, 413_NDST, 436_NTSK, 454_NYSE, 462_NNTH, 563_NETT, 618_NITD	288_WAFW, 645_WTGW, 648_WRQW
9	KR105291.1 (26)	47S + 70T = 117	040_NSTL, 204_NATE, 228_____, 238_NLTY, 257_NETI, 268_NTTG, 296_NLTR, 317_NISG, 333_NTTN, 346_NSSA, 386_NSTH, 413_NDST, 436_NTSK, 454_NYSE, 462_NNTH, 563_NETT, 618_NITD	288_WAFW, 645_WTGW, 648_WRQW
10	KR653268.1 (4)	47S + 70T = 117	040_NSTL, 228_NETE, 228_____, 238_NLTY, 257_NETI, 268_NTTG, 296_NLTR, 317_NISG, 333_NTTN, 346_NSSA, 386_NSTH, 413_NDST, 436_NTSK, 454_NYSE, 462_NNTH, 563_NETT, 618_NITD	288_WAFW, 645_WTGW, 648_WRQW

The total number of nucleotide sequences of the *GP*-gene utilized for this study are 1,184. However, the unique number of nucleotide sequences is only 168. The nucleotide sequence of the *GP*-gene of KM233036.1 (https://www.ncbi.nlm.nih.gov/nuccore/KM233036.1, Makona-EM106 isolated from Sierra Leone on 02-Jun-2014) alone is the representative of another 550 Makona variants (Table 4). The protein sequences show further reduction in the number of unique sequences due to codon redundancy. The number of unique preGP sequences is just 101, with KM233036.1 representing 781 protein sequences of other Makona variants.

Although *GP*-gene encodes 3 proteins by ribosomal slippage, the longest protein preGP represents all the glycosylation related modifications. The first 295 amino acids are common to all three proteins. The short unique parts in the two shorter proteins, sGP and ssGP, do not have any patterns for glycosylation. Hence, preGP gives a fair idea about the glycosylation of the receptor binding domain of the GP proteins.

O-glycosylation:
Serine and Threonine are two amino acids where the O-glycosylation can take place. The major representative of the sequences, KM233036.1, shows 47 serines and 71 threonines totaling to 118 number of O-glycosylation sites. This number can range from 116 (KR105281.1) to 120 (KR653230.1) in other Makona virus genomes. The total number of serines in preGP can range from 46 (KT589390.1) to 48 (KT357814.1 representing 31 other sequences). The number of threonines can range from 69 (KR105281.1) to 73 (KR653230.1).

The preGP sequence of KR105281.1 has the least number of serines and threonines (47S + 69T) totalling to 116 O-glycosylation sites. However, we do not see any major change in O-glycosylation pattern in the protein sequences of the Makona variant.

N-glycosylation:
Although N-glycosylation occurs on the Aspargine (Asn, N) amino acid, the surrounding pattern of amino acids is very important for glycosylation. Most of the preGP sequences have 38 Aspargine residues. Only 17 such Asparagine residues qualify for N-glycosylation because of the tetrapeptide sequence pattern in which it is found at the 1st position. Nearly 90% of these protein sequences have the same pattern of 17 glycosylation tetrapeptides similar to the master representative genome (see Table 4). Some patterns also show amino acid changes, 436_NTSK > 436_NTSE, as seen in KT357814.1 and 30 other viruses.

A total of 30 viruses do not have the 228_NETE motif, since the necessary amino acids following 228th N amino residue are absent.

C-mannosylation:
It is the attachment of an alpha-mannopyranose to a tryptophan (Trp, W) amino acid via a C-C linkage. The sequence WXXW, in which the first Trp becomes mannosylated, has been suggested as a consensus motif for the modification. However, roughly 1/3 rd of known sites are found to follow this motif pattern. We see 3 such motifs in the preGP protein of the Makona variant: 288_WAFW, 645_WTGW, 648_WRQW. One must note that the last 2 motifs are in patch of 7 amino acids, 3 tryptophans separated by 2 amino acids. Although these 3 motifs of WXXW are seen in all the variants, there are 4 virus samples that do not show the 1st motif at residue 288 (KJ660348.2 and 3 other viruses).

The overall study shows that more than 90% of the viruses found in the West African variant have very similar glycosylation pattern, besides their conservation in sequence.

4. Conclusion

An understanding of the protein diversity of the glycoprotein is very important in the design of vaccines targeted against the viral glycoprotein. The understanding of the glycosylation is equally important for the same. Such studies are also important for designing antibody cocktails for treating already infected patients with the virus.

The protein sequences of the glycoprotein of 1,184 genomes and their glycosylation potential showed that the diversity of the protein sequences in the 2014-16 Ebola outbreak was very limited, as seen from a very low number of unique protein sequences. The low diversity of the proteins also did not disturb the glycosylation potential of these proteins to a great extent. Such a finding encourages the designing of vaccines from the genome sequences of the samples found at the beginning of any outbreak

Conflict of interest

No conflict of interest influenced in this research.

5. References

1. Martinez O, Ndungo E, Tantral L, Miller EM, Leung LW, Chandran K, Basler CF. A Mutation in the Ebola Virus Envelope Glycoprotein Restricts Viral Entry in a Host Species- and Cell-Type-Specific Manner. Journal of Virology. 2013; 87 (6) :3324 –3334.
2. Kuhn JH, Becker S, Ebihara H, Geisbert TW, Johnson KM, Kawaoka Y, et al. Proposal for a revised taxonomy of the family Filoviridae: classification, names of taxa and viruses, and virus abbreviations. Archives of virology. 2010; 155 (12): 2083-2103.
3. Mahale KN, Patole MS. The crux and crust of ebolavirus: analysis of genome sequences and glycoprotein gene. Biochemical and biophysical research communications. 2015; 463 (4): 756-761.
4. Kuhn JH, Andersen KG, Baize S, Bào Y, Bavari S, et al. Nomenclature-and database-compatible names for the two Ebola virus variants that emerged in Guinea and the Democratic Republic of the Congo in 2014. Viruses. 2014; 6 (11): 4760-4799.
5. Ito H, Watanabe S, Takada A, Kawaoka Y. Ebola Virus Glycoprotein: Proteolytic Processing, Acylation, Cell Tropism, and Detection of Neutralizing Antibodies. Journal of Virology. 2019; 75 (3): 1576–1580.
6. Sanchez A, Kiley MP, Holloway BP, Auperin DD. Sequence analysis of the Ebola virus genome: organization, genetic elements, and comparison with the genome of Marburg virus. Virus research. 1993; 29 (3): 215-240.
7. Manicassamy B, Rong L. Expression of Ebolavirus glycoprotein on the target cells enhances viral entry. Virology Journal. 2009; 6 (75):1-15.
8. Lin G, Simmons G, Po¨hlmann S, Baribaud F, Ni H, Leslie G, Haggarty BS, Bates P, Weissman D, Hoxie JA, Doms RW. Differential N-Linked Glycosylation of Human Immunodeficiency Virus and Ebola Virus Envelope Glycoproteins Modulates Interactions with DC-SIGN and DC-SIGNR. Journal of Virology. 2003;77 (2): 1337–1346.
9. Cook JD, Lee JE. The secret life of viral entry glycoproteins: moonlighting in immune evasion. PLoS pathogens. 2013; 9 (5): e1003258.
10. Bagdonaite L, Wandall HH. Global aspects of viral glycosylation. Glycobiology. 2018;28 (7): 443–467.
11. Watanabe Y. Bowden TA, Wilson IA, Crispina M. Exploitation of glycosylation in enveloped virus pathobiology. BBA - General Subjects. 2019; 1863: 1480–1497.
12. Volchkov VE, Feldmann H, Volchkova VA, Klenk HD. Processing of the Ebola virus glycoprotein by the proprotein convertase furin. Proceedings of the National Academy of Sciences. 1998; 95 (10): 5762-5767.
13. Apweiler R, Hermjakob H, Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochimica et Biophysica Acta (BBA)-General Subjects. 1999; 1473 (1): 4-8.
14. Krieg J, Hartmann S, Vicentini A, GlaÌˆsner W, Hess D, Hofsteenge J. Recognition signal for C-mannosylation of Trp-7 in RNase 2 consists of sequence Trp-xx-Trp. Molecular biology of the cell. 1998; 9 (2): 301-309.
15. Vigerust D, Shepherd VL. Virus glycosylation: role in virulence and immune interactions. Trends in Microbiology. 2019; 15 (5): 211-218.
16. Iraqi M, Edri A, Greenshpan Y, Kundu K, Bolel P. et al. N-Glycans Mediate the Ebola Virus-GP1 Shielding of Ligands to Immune Receptors and Immune Evasion. Frontiers in cellular and infection Microbiology. 2020; 10 (48): 1-11.