



IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY



VOLUME : 4 ISSUE : 09 Print / Issue Publication Date: 10-Mar-2020



ISSN : 2455-2143



DOI : 10.33564/IJEAST.2020.v04i09.033

Indexed In



WWW.IJEAST.COM

editor@ijeast.com



CLASSIFICATION OF SPECIES OF EBOLA-VIRUS, BASED ON THE PHYSICO-CHEMICAL-PROPERTIES, CONSERVED AMINO-ACIDS, GENE-STRUCTURAL INFORMATION AND INTERPRETING ITS SIGNIFICANCE

Divya Peddapalyam
Msc. Biotechnology
Garden City University
Bengaluru, Karnataka, India

Reetika Mishra
Msc. Biotechnology
Garden City University
Bengaluru, Karnataka, India

Dr. Madhu Malleshappa
Msc, PhD, Post-Doc
Garden city University
Bengaluru, Karnataka, India

Abstract— Ebola Infection malady ia an uncommon and Destructive ailment which for the most part two people And non human primates, for example monkeys, gorillas, Chimpanzees. Among the five types of species i.e; Zaire, Sudan, Taiforest, Bundibugyo impacts the people and Reston Ebola virus known to cause disease in non human Primates and pigs. This Ebola infection was first found in 1976 close to ebola water way in majority rule republic Of congo. For Ebola virus sickness casuality rate is 90%. 318 individuals tainted and 280 passings are brought About by this infection with death pace of 88%.

I. OBJECTIVES

The extensive target of the subsequent module is that to create a compact, data rich outline of grouping information., illustrate the disparity between a gathering of sequences, Usage of arrangements as models to test hypohthesis, as well as to know whether this model of occasions precisely reflect known organic proof. The primary goal of the third module is to comprehend the essential ideas in quality finding, for example, relationship of protein and nucleotide groupings/exons/introns/coding arrangements/open perusing outlines/agreement properties of exon-intron fringes. Novel highlights of the program incorporate the ability to anticipate different qualities in a succession, to manage fractional just as complete qualities, and to foresee steady arrangements of qualities happening on either or both DNA strands. In the enormous scale investigation of gene, the regular procedure is totally inactivate every quality or over express it. In each case coming about phenotype may not be instructive. The loss of numerous proteins is deadly and this reveals to us that protein is fundamental however donot determine what protein really does. After forecast of quality structure, related with this protein we can research its Structure, Function, diseases, mutations and by utilizing this data we can fix numerous diseases. It utilizes factual example recognizable proof and succession similitude comparision, in which first technique utilizes every single imaginable ways to deal with concentrate the quality structure which incorporates advertiser region, start and end arrangements of intron and exon. As the closeness depends on the

evolution, either our grouping is homologous or not, this procedure depends on the comparability which exploit on the way that if the succession is similar, it will have a similar capacity.

II. METHODOLOGY

ProtParam registers different physico-substance properties that can be reasoned from a protein succession. No extra data is required about the protein under thought. The protein can either be indicated as a Swiss-Prot/TrEMBL promotion number or ID, or in type of a crude arrangement. Blank area and numbers are disregarded. In the event that you give the promotion number of a Swiss-Prot/TrEMBL passage, you will be incited with a delegate page that enables you to choose the part of the succession on which you might want to play out the investigation. The decision incorporates a choice of develop chains or peptides and spaces from the Swiss-Prot highlight table (which can be picked by tapping on the situations), just as the likelihood to enter start and end position in two boxes.

III. EXTINCTION COEFFICIENT

The Extinction coefficient shows how much light a protein assimilates at a specific wavelength. It is helpful to have an estimation of this coefficient for following a protein which a spectrophotometer when filtering it. It is conceivable to assess the molar Extinction coefficient of a protein from information of its amino corrosive creation. From the molar extinction coefficient of tyrosine, tryptophan, and cystine (cysteine doesn't ingest considerably at wavelengths >260 nm, while cystine does) at a given wavelength, the termination coefficient of a denatured protein can be figured. Two tables are delivered by ProtParam, the first demonstrating the processed qualities dependent on the supposition that all cysteine deposits show up as half cystines, and the subsequent one accepting that no cysteine shows up as half cystine. Formula for calculating Extinction coefficient is given below.

$$E(\text{Prot}) = \text{Numb}(\text{Tyr}) * \text{Ext}(\text{Tyr}) + \text{Numb}(\text{Trp}) * \text{Ext}(\text{Trp}) + \text{Numb}(\text{Cystine}) * \text{Ext}(\text{Cystine})$$
$$\text{Ext}(\text{Tyr}) = 1490,$$
$$\text{Ext}(\text{Trp}) = 5500,$$
$$\text{Ext}(\text{Cystine}) = 125;$$



IV. ALIPHATIC INDEX

The Aliphatic list of a protein is characterized as the relative volume involved by aliphatic side chains (alanine, valine, isoleucine, and leucine). It might be viewed as a positive factor for the expansion of thermostability of globular proteins

$$\text{Aliphatic index} = X(\text{Ala}) + a * X(\text{Val}) + b * (X(\text{Ile}) + X(\text{Leu}))$$

Where,

X(Ala) = Mole percent of Alanine

X(Val) = Mole percent of valine

X(Ile) = Mole percent of Isoleucine

X(Leu) = Mole percent of leucine

V. GRAND AVERAGE OF HYDROPHATICITY

The Grand Average of hydrophathy (GRAVY) esteem for a peptide or protein is determined as the entirety of hydrophathy estimations of all the amino acids, isolated by the number of deposits in the succession

VI. INVIVO HALF-LIFE

The half-life is an expectation of the time it takes for half of the measure of protein in a cell to vanish after its blend in the cell. The expectation is given for three creatures (human, yeast, and E. coli), yet it is conceivable to extrapolate the outcome to comparative living beings. ProtParam gauges the half-life by taking a gander at the N-terminal amino corrosive of the grouping under investigation.

VII. INSTABILITY INDEX

```
# @export instaIndex
# @title Compute the instability index of a protein
sequence
# @description This function calculates the instability
index proposed by Guruprasad (1990). This index predicts
the stability of a protein based on its amino acid
composition, a protein whose instability index is smaller
than 40 is predicted as stable, a value above 40 predicts that
the protein may be unstable.
# @param seq An amino-acids sequence
# @return The computed instability index for a given
amino-acids sequence
# @references Guruprasad K, Reddy BV, Pandit MW
(1990). "Correlation between stability of a protein and its
dipeptide composition: a novel approach for predicting in
vivo stability of a protein from its primary sequence".
Protein Eng. 4 (2): 155 - 61. doi:10.1093/protein/4.2.155
# @examples
# # COMPARED TO ExPASy INSTAINDEX
# # http://web.expasy.org/protparam/
# # SEQUENCE:
QWGRRC CGWGPGRRYCVRWC
# # The instability index (II) is computed to be
83.68
#
# instaIndex(seq
"QWGRRC CGWGPGRRYCVRWC")
# # [1] 83.68
```

```
#
instaIndex <- function(seq) {
# Setting the Guruprasad scale
guruprasad <-
c(
WW = 1,
WC = 1,
WM = 24.68,
WH = 24.68,
WY = 1,
WF = 1,
WQ = 1,
WN = 13.34,
WI = 1,
WR = 1,
WD = 1,
WP = 1,
WT = -14.03,
WK = 1,
WE = 1,
WV = -7.49,
WS = 1,
WG = -9.37,
WA = -14.03,
WL = 13.34,
CW = 24.68,
CC = 1,
CM = 33.6,
CH = 33.6,
CY = 1,
CF = 1,
CQ = -6.54,
CN = 1,
CI = 1,
CR = 1,
CD = 20.26,
CP = 20.26,
CT = 33.6,
CK = 1,
CE = 1,
CV = -6.54,
CS = 1,
CG = 1,
CA = 1,
CL = 20.26,
MW = 1,
MC = 1,
MM = -1.88,
MH = 58.28,
MY = 24.68,
MF = 1,
MQ = -6.54,
MN = 1,
MI = 1,
MR = -6.54,
MD = 1,
MP = 44.94,
MT = -1.88,
MK = 1,
ME = 1,
MV = 1,
```



MS = 44.94,
MG = 1,
MA = 13.34,
ML = 1,
HW = -1.88,
HC = 1,
HM = 1,
HH = 1,
HY = 44.94,
HF = -9.37,
HQ = 1,
HN = 24.68,
HI = 44.94,
HR = 1,
HD = 1,
HP = -1.88,
HT = -6.54,
HK = 24.68,
HE = 1,
HV = 1,
HS = 1,
HG = -9.37,
HA = 1,
HL = 1,
YW = -9.37,
YC = 1,
YM = 44.94,
YH = 13.34,
YY = 13.34,
YF = 1,
YQ = 1,
YN = 1,
YI = 1,
YR = -15.91,
YD = 24.68,
YP = 13.34,
YT = -7.49,
YK = 1,
YE = -6.54,
YV = 1,
YS = 1,
YG = -7.49,
YA = 24.68,
YL = 1,
FW = 1,
FC = 1,
FM = 1,
FH = 1,
FY = 33.6,
FF = 1,
FQ = 1,
FN = 1,
FI = 1,
FR = 1,
FD = 13.34,
FP = 20.26,
FT = 1,
FK = -14.03,
FE = 1,
FV = 1,
FS = 1,

FG = 1,
FA = 1,
FL = 1,
QW = 1,
QC = -6.54,
QM = 1,
QH = 1,
QY = -6.54,
QF = -6.54,
QQ = 20.26,
QN = 1,
QI = 1,
QR = 1,
QD = 20.26,
QP = 20.26,
QT = 1,
QK = 1,
QE = 20.26,
QV = -6.54,
QS = 44.94,
QG = 1,
QA = 1,
QL = 1,
NW = -9.37,
NC = -1.88,
NM = 1,
NH = 1,
NY = 1,
NF = -14.03,
NQ = -6.54,
NN = 1,
NI = 44.94,
NR = 1,
ND = 1,
NP = -1.88,
NT = -7.49,
NK = 24.68,
NE = 1,
NV = 1,
NS = 1,
NG = -14.03,
NL = 1,
IW = 1,
IC = 1,
IM = 1,
IH = 13.34,
IY = 1,
IF = 1,
IQ = 1,
IN = 1,
II = 1,
IR = 1,
ID = 1,
IP = -1.88,
IT = 1,
IK = -7.49,
IE = 44.94,
IV = -7.49,
IS = 1,
IG = 1,
IA = 1,



IL = 20.26,
RW = 58.28,
RC = 1,
RM = 1,
RH = 20.26,
RY = -6.54,
RF = 1,
RQ = 20.26,
RN = 13.34,
RI = 1,
RR = 58.28,
RD = 1,
RP = 20.26,
RT = 1,
RK = 1,
RE = 1,
RV = 1,
RS = 44.94,
RG = -7.49,
RA = 1,
RL = 1,
DW = 1,
DC = 1,
DM = 1,
DH = 1,
DY = 1,
DF = -6.54,
DQ = 1,
DN = 1,
DI = 1,
DR = -6.54,
DD = 1,
DP = 1,
DT = -14.03,
DK = -7.49,
DE = 1,
DV = 1,
DS = 20.26,
DG = 1,
DA = 1,
DL = 1,
PW = -1.88,
PC = -6.54,
PM = -6.54,
PH = 1,
PY = 1,
PF = 20.26,
PQ = 20.26,
PN = 1,
PI = 1,
PR = -6.54,
PD = -6.54,
PP = 20.26,
PT = 1,
PK = 1,
PE = 18.38,
PV = 20.26,
PS = 20.26,
PG = 1,
PA = 20.26,
PL = 1,

TW = -14.03,
TC = 1,
TM = 1,
TH = 1,
TY = 1,
TF = 13.34,
TQ = -6.54,
TN = -14.03,
TI = 1,
TR = 1,
TD = 1,
TP = 1,
TT = 1,
TK = 1,
TE = 20.26,
TV = 1,
TS = 1,
TG = -7.49,
TA = 1,
TL = 1,
KW = 1,
KC = 1,
KM = 33.6,
KH = 1,
KY = 1,
KF = 1,
KQ = 24.68,
KN = 1,
KI = -7.49,
KR = 33.6,
KD = 1,
KP = -6.54,
KT = 1,
KK = 1,
KE = 1,
KV = -7.49,
KS = 1,
KG = -7.49,
KA = 1,
KL = -7.49,
EW = -14.03,
EC = 44.94,
EM = 1,
EH = -6.54,
EY = 1,
EF = 1,
EQ = 20.26,
EN = 1,
EI = 20.26,
ER = 1,
ED = 20.26,
EP = 20.26,
ET = 1,
EK = 1,
EE = 33.6,
EV = 1,
ES = 20.26,
EG = 1,
EA = 1,
EL = 1,
VW = 1,



VC = 1,
 VM = 1,
 VH = 1,
 VY = -6.54,
 VF = 1,
 VQ = 1,
 VN = 1,
 VI = 1,
 VR = 1,
 VD = -14.03,
 VP = 20.26,
 VT = -7.49,
 VK = -1.88,
 VE = 1,
 VV = 1,
 VS = 1,
 VG = -7.49,
 VA = 1,
 VL = 1,
 SW = 1,
 SC = 33.6,
 SM = 1,
 SH = 1,
 SY = 1,
 SF = 1,
 SQ = 20.26,
 SN = 1,
 SI = 1,
 SR = 20.26,
 SD = 1,
 SP = 44.94,
 ST = 1,
 SK = 1,
 SE = 20.26,
 SV = 1,
 SS = 20.26,
 SG = 1,
 SA = 1,
 SL = 1,
 GW = 13.34,
 GC = 1,
 GM = 1,
 GH = 1,
 GY = -7.49,
 GF = 1,
 GQ = 1,
 GN = -7.49,
 GI = -7.49,
 GR = 1,
 GD = 1,
 GP = 1,
 GT = -7.49,
 GK = -7.49,
 GE = -6.54,
 GV = 1,
 GS = 1,
 GG = 13.34,
 GA = -7.49,
 GL = 1,
 AW = 1,
 AC = 44.94,

AM = 1,
 AH = -7.49,
 AY = 1,
 AF = 1,
 AQ = 1,
 AN = 1,
 AI = 1,
 AR = 1,
 AD = -7.49,
 AP = 20.26,
 AT = 1,
 AK = 1,
 AE = 1,
 AV = 1,
 AS = 1,
 AG = 1,
 AA = 1,
 AL = 1,
 LW = 24.68,
 LC = 1,
 LM = 1,
 LH = 1,
 LY = 1,
 LF = 1,
 LQ = 33.6,
 LN = 1,
 LI = 1,
 LR = 20.26,
 LD = 1,
 LP = 20.26,
 LT = 1,
 LK = -7.49,
 LE = 1,
 LV = 1,
 LS = 1,
 LG = 1,
 LA = 1,
 LL = 1,
 'NA' = 1

```
) # Divide the amino acid sequence in dipeptides
aa <- aaCheck(seq)
dp <- lapply(aa, function(aa) {
  apply(embed(aa, 2)[, 2:1], 1, paste0, collapse = "")
})
# Apply the formula:
# (10/L)*sum(DIWV(XiYi+1) for each dipeptide)
# Return the index value rounded to 2 decimals
gp <- lapply(dp, function(dp) {
  (10 / (length(dp) + 1)) * sum(guruprasad[dp], na.rm = TRUE)
})
return(unlist(gp))
```

VIII. MULTIPLE SEQUENCE ALIGNMENT

MODULE-2

We care about the grouping arrangements in the computational science since it gives scientists helpful data about various viewpoints. For instance, it can educate us regarding the advancement of the living beings, we can see which locales of a quality (or its determined protein) are defenseless to change and which can have one buildup supplanted by another without evolving capacity, we can think about Homologous



qualities and can reveal paralogs and Orthologs qualities that are developmental related. In issues, for example, the development of a transformative tree dependent on arrangement information, or in protein designing, where a different arrangement of related groupings may regularly yield the most supportive data on the plan of another protein, an atomic scholar must think about multiple arrangements all the while. A numerous grouping arrangement (MSA) organizes protein successions into a rectangular exhibit with the objective that buildups in a given section are homologous (gotten from a single situation in a hereditary grouping), superposable (in an unbending nearby basic arrangement) or assume a typical practical job. In spite of the fact that these three criteria are basically comparable for firmly related proteins, arrangement, structure and capacity separate over transformative time and various criteria may bring about various arrangements. Physically refined arrangements keep on being better than simply mechanized techniques; there is in this manner a persistent exertion to improve the organic exactness of MSA apparatuses. Also, the high computational expense of most guileless calculations inspires upgrades in speed and memory utilization to suit the quick increment in accessible succession information. The ClustalW calculation has three Important Phases. They are

Stage I: All sets of arrangements are adjusted independently ascertain a Distance Matrix dependent on the level of befuddles each pair of groupings.

Stage II: The guide tree is developed from the separation framework utilizing the Neighbor Joining calculation.

Stage III: The successions are dynamically adjusted after the guide tree.

MODULE-3

Genscan is utilized for anticipating the areas and exon-intron structures of qualities in genomic groupings from an assortment of creatures. This server can acknowledge arrangements up to 1 million base sets (1 Mbp) long. On the off chance that you experience difficulty with the web server or on the off chance that you have an enormous number of groupings to process, demand a nearby duplicate of the program. OMICS_01494 was created by Chris Burge in the examination gathering of Samuel Karlin, Department of Mathematics, Stanford University. OMICS_10494 is uninhibitedly accessible for scholastic use. Executables are right now accessible for the accompanying Unix stages: Intel/Linux, Sun/Solaris, Intel/Solaris, SGI/Irix, DEC/Tru64, and IBM/AIX. Distinguishes total exon/intron structures of qualities in genomic DNA. OMICS_01494 utilizes a homogeneous fifth request Markov model of noncoding areas and a three intermittent (inhomogeneous) fifth request Markov model of coding districts. Highlights of the program incorporate the ability to foresee various qualities in a grouping, to manage halfway just as complete qualities, and to anticipate predictable arrangements of qualities happening on either or both DNA strands

IX. LITERATURE REVIEW

Ebolavirus has a place with the request Mononegavirales and the family Filoviridae. Its RNA genome encodes the accompanying 9 protein items: Spike glycoprotein (GP), Small secreted glycol-protein, Second secreted Glyco-protein, Nucleoprotein (NP), RNA-subordinate RNA polymerase (L), Membrane-related protein (VP24), Minor nucleoprotein (VP30), Polymerase cofactor (VP35), and Matrix protein (VP40). The GP transcript can be altered, and the quality item can be handled by host protease, offering ascend to 4 elective types of

quality items: GP1,2; GP1,2delta; sGP and ssGP. Host furin can sever the longest item interpreted from altered GP mRNA and create GP1,2, which comprises of 2 peptide chains associated by a disulfide bond, GP1 and GP2. GP1,2 is gathered on the layer of Ebolavirus and intercedes cell passage. GP1,2delta is the handled item after evacuation of the C-terminal transmembrane locale of GP1,2 by host ADAM17. Different results of the GP quality, sGP and ssGP are interpreted from the unedited mRNA and then again altered mRNA, respectively. These items share the N-terminal 295 buildups with GP1,2, however vary in their short tails (69 and 3 deposits, separately). GP1,2delta, sGP and ssGP may keep the killing antibodies from restricting GP1,2 on the infection surface, adding to the insusceptible avoidance of the infection. Notwithstanding filling in as basic parts, the Ebolavirus proteins assume numerous jobs in the infection life cycle. GP intercedes cell section and layer combination between the infection and the host cell. NP encapsidates the genome and shields it from nucleases. VP30 is a translation hostile to eliminator and directs the switch among interpretation and replication. VP35 goes about as a cofactor of the polymerase, and VP40 may likewise assume a job in genome replication and interpretation. VP24 and VP35 take an interest in viral nucleocapsid assembly, and VP40 is basic for infection growing and gathering. What's more, GP, VP24, VP30, VP35 and VP40 associate with different host proteins to finish the viral life cycle and to stifle the host insusceptible reaction. In the present examination, we anticipate the 3D structure and utilitarian locales for Ebolavirus protein areas that are not yet portrayed. Also, we think about successions of Ebolavirus proteins' communicating accomplices from RESTV-safe primates with those from RESTV-powerless monkeys. Raised arrangement difference for GP and VP35's collaboration accomplices recommends that these 2 viral proteins might be in charge of host particularity in RESTV. At last, we think about the protein groupings from various Ebolavirus species to distinguish places that are moderated among human pathogenic species yet extraordinary in non-pathogenic (RESTV-explicit transformations). Mapping of these RESTV-explicit transformations and known utilitarian destinations to the 3D structures uncovers bunches of RESTV-explicit changes on the surfaces of GP, VP35 and VP24. These bunches don't cover with the known useful locales and may propose novel connection destinations with host proteins. Based on this review we decided to study physico-chemical propertys of Ebolavirus along with Gene structural information and sequence homology to interpret significant aspects on Ebola virus. Ebola Virus Disease (EVD) is a rare and deadly disease in people and nonhuman primates. The viruses that cause EVD are located mainly in sub-Saharan Africa. People can get EVD through direct contact with an infected animal (bat or nonhuman primate) or a sick or dead person infected with Ebola virus. The U.S. Food and Drug Administration (FDA) has approved the Ebola vaccine rVSV-ZEBOV (tradename "Ervebo") for the prevention of EVD. The rVSV-ZEBOV vaccine has been found to be safe and protective against only the Zaire ebolavirus species of ebolavirus. Ebola virus disease (EVD), one of the deadliest viral diseases, was discovered in 1976 when two consecutive outbreaks of fatal hemorrhagic fever occurred in different parts of Central Africa. The first outbreak occurred in the Democratic Republic of Congo (formerly Zaire) in a village near the Ebola River, which gave the virus its name. The second outbreak occurred in what is now South Sudan, approximately 500 miles (850 km) away. Initially, public health officials assumed these outbreaks were a single event associated with an infected person who traveled between the two locations. However, scientists later discovered that the two outbreaks were caused by two genetically distinct viruses: Zaire ebolavirus and Sudan ebolavirus. After this discovery, scientists concluded that the virus came from two



different sources and spread independently to people in each of the affected areas. Viral and epidemiologic data suggest that Ebola virus existed long before these recorded outbreaks occurred. Factors like population growth, encroachment into forested areas, and direct interaction with wildlife (such as bushmeat consumption) may have contributed to the spread of the Ebola virus. Following the discovery of the virus, scientists studied thousands of animals, insects, and plants in search of its source (called reservoir among virologists, people who study viruses). Gorillas, chimpanzees, and other mammals may be implicated when the first cases of an EVD outbreak in people occur. However, they – like people – are “dead-end” hosts, meaning the organism dies following the infection and does not survive and spread the virus to other animals. Like other viruses of its kind, it is possible that the reservoir host animal of Ebola virus does not experience acute illness despite the virus being present in its organs, tissues, and blood. Thus, the virus is likely maintained in the environment by spreading from host to host or through intermediate hosts or vectors. African fruit bats are likely involved in the spread of Ebola virus and may even be the source animal (reservoir host). Scientists continue to search for conclusive evidence of the bat’s role in transmission of Ebola. The most recent Ebola virus to be detected, Bombali virus, was identified in samples from bats collected in Sierra Leone. The use of contaminated needles and syringes during the earliest outbreaks enabled transmission and amplification of Ebola virus. During the first outbreak in Zaire (now Democratic Republic of Congo – DRC), nurses in the Yambuku mission hospital reportedly used five syringes for 300 to 600 patients a day. Close contact with infected blood, reuse of contaminated needles, and improper nursing techniques were the source for much of the human-to-human transmission during early Ebola outbreaks. In 1989, Reston ebolavirus was discovered in research monkeys imported from the Philippines into the U.S. Later, scientists confirmed that the virus spread throughout the monkey population through droplets in the air (aerosolized transmission) in the facility. However, such airborne transmission is not proven to be a significant factor in human outbreaks of Ebola.⁴ The discovery of the Reston virus in these monkeys from the Philippines revealed that Ebola was no longer confined to African settings, but was present in Asia as well. By the 1994 Cote d’Ivoire outbreak, scientists and public health officials had a better understanding of how Ebola virus spreads and progress was made to reduce transmission through the use of face masks, gloves and gowns for healthcare personnel. In addition, the use of disposable equipment, such as needles, was introduced. During the 1995 Kikwit, Zaire (now DRC) outbreak, the international public health community played a strong role, as it was now widely agreed that containment and control of Ebola were paramount. Ebola virus disease (EVD) is a life-threatening viral disease with a fatality rate ranging from around 30% to 90%. The first EVD outbreak was reported in the 1970s in Zaire (now the Democratic Republic of the Congo). Until 2013, most outbreaks occurred in the Central Africa region, including Zaire, Sudan and Uganda. However, between March and October 2014, over 10 000 cases of EVD have been recorded in West Africa, such as in Guinea, Liberia, Sierra Leone, and Nigeria, and a few hospital or secondary infections of EVD have occurred in Spain and the United States of America. EVD is presently one of the world’s most feared diseases. In this literature review, we describe the epidemiology, clinical features, diagnosis, and treatment of EVD.

The virus is thought to be initially acquired by exposure to body fluids or tissue from infected animals, such as bats and non-human primates; however, the natural reservoir and mode of transmission to humans has not been confirmed. Laboratory testing of reservoir competence shows that successful infection is possible in bats and rodents, but not in plants or arthropods. Animal to human transmission may occur during hunting and consumption of the reservoir species or infected non-human primates. The practice of butchering or eating bush meat or food contaminated with bat faeces (three species of tree roosting bats have been implicated as a reservoir) is also thought to contribute. Human to human transmission occurs through contact with body fluids from infected patients. In early epidemics, the re-use of non-sterile injections was responsible for many healthcare associated transmissions. However, although this remains a risk, most cases result from close physical contact or contact with body fluids (such as sweat, blood, faeces, vomit, saliva, genital secretions, urine, and breast milk) of infected patients. In a study of viral shedding in various body fluids, Ebola virus was isolated from saliva, breast milk, stool, tears, and semen up to 40 days after the onset of illness, confirming the possibility of delayed sexual transmission. Virus may be found in urine during recovery, and the duration of this phenomenon needs further study. Infection through inhalation is possible in non-human primates, but there is no evidence for airborne transmission in humans. Outside endemic areas, Ebola virus infection is rare and is usually imported. Travellers from affected areas, and laboratory scientists and others working with potentially infected materials and animals, are at high risk.



RESULTS AND DISCUSSION MODULE-1

		Bundibugyo Ebola virus	Sudan Ebola virus	Reston Ebola virus	Zaire Ebola virus	Tai forest ebola virus
Number and composition of amino acids NUCLEO-PROTEIN	Alanine (Ala)	63(8.5%)	52(7.0%)	60(8.1%)	53(7.2%)	54(7.3%)
	Arginine (Arg)	31(4.2%)	27(3.7%)	39(5.3%)	33(4.5%)	29(3.9%)
	Asparagine (Asn)	43(5.8%)	34(4.6%)	40(5.4%)	33(4.5%)	45(6.1%)
	Aspartic acid (Asp)	57(7.7%)	59(8.0%)	59(8.0%)	59(8.0%)	48(6.5%)
	Cysteine (Cys)	03(0.4%)	03(0.4%)	03(0.4%)	03(0.4%)	03(0.4%)
	Glutamine (Gln)	51(6.9%)	47(6.4%)	52(7.0%)	53(7.2%)	49(6.6%)
	Glutamic acid (Glu)	56(7.6%)	58(7.9%)	59(8.0%)	59(8.0%)	63(8.5%)
	Glycine (Gly)	37(5.0%)	53(7.2%)	42(5.7%)	41(5.5%)	37(5.0%)
	Histidine (His)	25(3.4%)	25(3.4%)	28(3.8%)	30(4.1%)	30(4.1%)
	Iso-Leucine (Ile)	38(5.1%)	29(3.9%)	33(4.5%)	29(3.9%)	34(4.6%)
	Leucine (Leu)	62(8.4%)	77(10.4%)	74(10.0%)	67(9.1%)	64(8.7%)
	Lysine (Lys)	37(5.0%)	36(4.9%)	31(4.2%)	38(5.1%)	41(5.5%)
	Methionine (Met)	20(2.7%)	13(1.8%)	15(2.0%)	20(2.7%)	17(2.3%)
	Phenyl Alanine (Phe)	25(3.4%)	24(3.3%)	25(3.4%)	26(3.5%)	26(3.5%)
	Proline (Pro)	40(5.4%)	42(5.7%)	40(5.4%)	42(5.7%)	37(5.0%)
	Serine (Ser)	47(6.4%)	49(6.6%)	44(6.0%)	48(6.5%)	51(6.6%)
	Threonine (Thr)	41(5.5%)	39(5.3%)	32(4.3%)	38(5.1%)	49(6.9%)
	Tryptophan (Trp)	04(0.5%)	04(0.5%)	05(0.7%)	04(0.5%)	05(0.7%)
Tyrosine (Tyr)	22(3.0%)	22(3.0%)	24(3.2%)	21(2.8%)	21(2.8%)	
Valine (Val)	37(5.0%)	45(6.1%)	34(4.6%)	42(5.7%)	36(4.9%)	
Molecular weight		82905.24	81804.90	83452.62	83286.68	83308.50
Theoretical pI		4.90	4.73	4.91	4.98	5.13
Atomic composition		11487	11365	11557	11535	11541
Total number of positively charged residues		68	63	70	71	70
Total number of Negatively charged residues		113	117	118	118	111
Extinction coefficient assuming cystine residues		54905M ⁻¹ cm ⁻¹	54905M ⁻¹ cm ⁻¹	63385M ⁻¹ cm ⁻¹	53415M ⁻¹ cm ⁻¹	58915M ⁻¹ cm ⁻¹
Absorbance assuming cystine residues		0.662	0.671	0.760	0.641	0.707
Extinction coefficient with out cystine residues		54780M ⁻¹ cm ⁻¹	54780M ⁻¹ cm ⁻¹	63260M ⁻¹ cm ⁻¹	53290M ⁻¹ cm ⁻¹	58790M ⁻¹ cm ⁻¹
Absorbance with out cystine residues		0.661	0.670	0.758	0.640	0.706
Instability Index		45.60	38.60	46.18	50.31	46.88
Aliphatic Index		75.82	80.75	77.93	74.32	73.15
Grand Average of Hydrophaticity		-0.642	-0.565	-0.688	-0.691	-0.714

Table-1: The above table describes different physico-chemical properties associated with the nucleo-protein of Ebola virus species, in which all forms of nucleo-protein in all species of the Ebola virus species have the same number and composition of the amino acids on the whole but vary when compared with the individual amino acids. When we calculated the average isotope mass of the amino acid in the protein and one water molecule, the total molecular weight of the protein is estimated to be in this order: i.e., 83452.62 > 83308.50 > 83286.68 > 82905.24 > 81804.90. i.e., Reston Ebola virus with greater molecular weight and Sudan Ebola virus with the smaller molecular weight. The pI/Mw algorithm is mainly used to enhance a region in a 2-D gel to which a protein which is unmodified should be allowed to run, and point a region in which a modified form of protein should be found if the modifications are documented in the database. When we calculated the sums of different amino acid contributions assuming them as independent with out considering the secondary and tertiary structures, we observed similar values of extinction coefficient for Bundibugyo and Sudan Ebola virus with and without assuming cysteine residues with different absorbance. The nucleo-protein of Sudan Ebola virus is more stable when compared with the other forms of the virus species with a value of 38.60 as the value is less than 40 and Zaire Ebola virus is more unstable as it exceeds a value greater than 40. When we determined the positive factor which explains the increment phenomenon of the globular proteins and volume occupied relatively by the aliphatic side chains we observed 80.75 (Sudan Ebola virus) > 77.93 (Reston Ebola virus) > 75.82 (Bundibugyo Ebola virus) > 74.32 (Zaire Ebola virus) > 73.15 (Tai forest Ebola virus) and when we studied the phenomenon of the protein that is exhibiting the ability of repelling from the water, the hydrophobicity value of the Sudan Ebola virus (-0.565) is greater and for the Tai forest Ebola virus (-0.714) is less.



Number and composition of amino acids		Bundibugyo Ebola virus	Sudan Ebola virus	Reston Ebola virus	Zaire Ebola virus	Tai forest ebola virus
POLYMERASE COMPLEX PROTEIN	Alanine (Ala)	25(7.3%)	31(9.4%)	23(7.0%)	27(7.9%)	27(7.9%)
	Arginine (Arg)	16(4.7%)	13(4.0%)	13(4.0%)	18(5.3%)	18(5.3%)
	Asparagine (Asn)	10(2.9%)	12(3.6%)	14(4.3%)	14(4.1%)	15(4.4%)
	Aspartic acid (Asp)	18(5.3%)	20(6.1%)	22(6.7%)	17(5.0%)	21(6.2%)
	Cysteine (Cys)	07(2.1%)	05(1.5%)	07(2.1%)	08(2.4%)	08(2.3%)
	Glutamine (Gln)	18(5.3%)	18(5.5%)	15(4.6%)	23(6.8%)	20(5.9%)
	Glutamic acid (Glu)	20(5.9%)	19(5.8%)	17(5.2%)	20(5.9%)	21(6.2%)
	Glycine (Gly)	19(5.6%)	17(5.2%)	16(4.9%)	20(5.9%)	18(5.3%)
	Histidine (His)	07(2.1%)	07(2.1%)	05(1.5%)	07(2.1%)	08(2.3%)
	Iso-Leucine (Ile)	27(7.9%)	27(8.2%)	23(7.0%)	23(6.8%)	28(8.2%)
	Leucine (Leu)	28(8.2%)	23(7.0%)	29(8.8%)	25(7.4%)	28(8.2%)
	Lysine (Lys)	21(6.2%)	26(7.9%)	26(7.9%)	16(4.7%)	19(5.6%)
	Methionine (Met)	07(2.1%)	05(1.5%)	09(2.7%)	08(2.4%)	07(2.1%)
	Phenyl Alanine (Phe)	09(2.6%)	09(2.7%)	07(2.1%)	09(2.6%)	09(2.6%)
	Proline (Pro)	26(7.6%)	23(7.0%)	24(7.3%)	24(7.1%)	26(7.6%)
	Serine (Ser)	26(7.6%)	25(7.6%)	26(7.9%)	26(7.6%)	23(6.7%)
	Threonine (Thr)	30(8.8%)	20(6.1%)	23(7.0%)	28(8.2%)	24(7.0%)
Tryptophan (Trp)	03(0.9%)	03(0.9%)	03(0.9%)	03(0.9%)	03(0.9%)	
Tyrosine (Tyr)	05(1.5%)	07(2.1%)	09(2.7%)	06(1.8%)	05(1.5%)	
Valine (Val)	19(5.6%)	19(5.8%)	18(5.5%)	18(5.3%)	13(3.8%)	
Molecular weight		37399.85	36116.21	36409.75	37362.36	37732.92
Theoretical pI		6.67	7.05	6.95	6.19	5.94
Atomic composition		5286	5104	5135	5226	5296
Total number of positively charged residues		37	39	39	34	37
Total number of Negatively charged residues		38	39	39	37	42
Extinction coefficient assuming cystine residues		24325M ⁻¹ cm ⁻¹	27180M ⁻¹ cm ⁻¹	30285M ⁻¹ cm ⁻¹	25940M ⁻¹ cm ⁻¹	24450M ⁻¹ cm ⁻¹
Absorbance assuming cystine residues		0.650	0.753	0.832	0.694	0.648
Extinction coefficient with out cystine residues		23950M ⁻¹ cm ⁻¹	26930M ⁻¹ cm ⁻¹	29910M ⁻¹ cm ⁻¹	25440M ⁻¹ cm ⁻¹	23950M ⁻¹ cm ⁻¹
Absorbance with out cystine residues		0.640	0.746	0.821	0.681	0.635
Instability Index		50.17	45.04	46.31	47.32	46.50
Aliphatic Index		86.39	85.44	84.50	78.35	83.02
Grand Average of Hydrophaticity		-0.290	-0.369	-0.380	-0.409	-0.438

Table-2: The different physico-chemical properties of polymerase complex protein of the ebola virus is described in the above table in which, the total number shows approximation in its value (Similarity in total number of Amino acids showing similarity in two to three amino acid number) but differs in the individual amino acids. Average isotope mass on protein and one water molecule with respect to each amino acid is calculated, then the total molecular weight of the protein is obtained in this order i.e.; 37732.92 > 37399.85 > 37362.36 > 36409.75 > 36116.21. By this we can say that Tai forest Ebola virus having greater molecular weight and Sudan Ebola virus with smaller molecular weight. Compute pI/Mw algorithm is mainly used to enhance a region in a 2-D gel to which an a protein which is unmodified should allowed to run, and point a region in which modified form of protein should be found if the modifications are documented in the database. Assuming the different independent amino acid contributions with out considering the secondary and tertiary structures, we observed different values of Extinction coefficient with different absorbance values in both cases of assuming and non assuming cysteine residues. All the Polymerase complex proteins in all the Ebola virus species is not stable and this can be justified based upon the values we got i.e.; all the values obtained is greater than 40. When the positive factor which explains the increment phenomenon of the globular protein and volume occupied by the aliphatic side chains is determined the values are 78.35 (Zaire Ebolavirus) > 83.02 (Tai forest Ebolavirus) > 84.50 (Reston Ebolavirus) > 85.44 (Sudan Ebolavirus) > 86.39 (Bundibugyo Ebolavirus). The repelling capacity of the protein in Bundibugyo Ebolavirus (-0.290) is higher and the repelling capacity of protein in the Tai forest Ebola virus (-0.438) is less.



		Tai forest Ebola virus	Sudan Ebola virus	Reston Ebola virus	Zaire Ebola virus
Number and composition of amino acids MATRIX -PROTEIN	Alanine (Ala)	19(5.8%)	21(6.4%)	23(6.9%)	22(6.7%)
	Arginine (Arg)	12(3.7%)	10(3.1%)	11(3.3%)	11(3.4%)
	Asparagine (Asn)	16(4.9%)	11(3.4%)	13(3.9%)	14(4.3%)
	Aspartic acid (Asp)	18(5.5%)	19(5.8%)	20(6.0%)	17(5.2%)
	Cysteine (Cys)	02(0.6%)	02(0.6%)	02(0.6%)	02(0.6%)
	Glutamine (Gln)	12(3.7%)	15(4.6%)	14(4.2%)	11(3.4%)
	Glutamic acid (Glu)	08(2.5%)	07(2.1%)	06(1.8%)	09(2.8%)
	Glycine (Gly)	19(5.8%)	20(6.1%)	19(5.7%)	21(6.4%)
	Histidine (His)	07(2.1%)	08(2.5%)	09(2.7%)	07(2.1%)
	Iso-Leucine (Ile)	26(8.0%)	24(7.4%)	25(7.6%)	27(8.3%)
	Leucine (Leu)	35(10.7%)	35(10.7%)	36(10.9%)	33(10.1%)
	Lysine (Lys)	16(4.9%)	20(6.1%)	18(5.4%)	18(5.5%)
	Methionine (Met)	11(3.4%)	11(3.4%)	08(2.4%)	08(2.5%)
	Phenyl Alanine (Phe)	09(2.8%)	09(2.8%)	07(2.1%)	10(3.1%)
	Proline (Pro)	37(11.3%)	36(11.0%)	40(12.1%)	37(11.3%)
	Serine (Ser)	22(6.7%)	23(7.1%)	22(6.6%)	22(6.7%)
Threonine (Thr)	31(9.5%)	23(7.1%)	23(6.9%)	29(8.9%)	
Tryptophan (Trp)	02(0.6%)	02(0.6%)	02(0.6%)	02(0.6%)	
Tyrosine (Tyr)	07(2.1%)	08(2.5%)	08(2.4%)	06(1.8%)	
Valine (Val)	17(5.2%)	22(6.7%)	25(7.6%)	20(6.1%)	
Molecular weight		35525.19	35475.35	35820.66	35182.83
Theoretical pI		8.44	8.91	8.73	8.76
Atomic composition		5056	5064	5123	5026
Total number of positively charged residues		28	30	29	29
Total number of Negatively charged residues		26	26	26	26
Extinction coefficient assuming cystine residues		21555M ⁻¹ cm ⁻¹	23045M ⁻¹ cm ⁻¹	23045M ⁻¹ cm ⁻¹	20065M ⁻¹ cm ⁻¹
Absorbance assuming cystine residues		0.607	0.650	0.643	0.570
Extinction coefficient with out cystine residues		21430M ⁻¹ cm ⁻¹	22920M ⁻¹ cm ⁻¹	22920M ⁻¹ cm ⁻¹	19940M ⁻¹ cm ⁻¹
Absorbance with out cystine residues		0.603	0.646	0.640	0.567
Instability Index		40.13	41.76	41.41	40.39
Aliphatic Index		93.93	96.60	100.73	96.32
Grand Average of Hydrophaticity		-0.117	-0.063	-0.048	-0.052

Table-3: The matrix protein of all Ebola virus species shows approximation in the values illustrating total number and composition of the amino acids but completely varies in the individual amino acids. On the protein and Water molecule the average isotope mass with respect to each amino acid is calculated and the values are obtained to be 35820.66 > 35525.19 > 35475.35 > 35458.31 > 35182.83 i.e.; Reston Ebolavirus with the greater molecular weight and Zaire Ebolavirus with smaller molecular weight. Compute pI/Mw algorithm is mainly used to enhance a region in a 2-D gel to which an a protein which is unmodified should allowed to run, and point a region in which modified form of protein should be found if the modifications are documented in the database. Without considering the secondary and tertiary structures when independent amino acid contributions are studied we observed different values of the extinction coefficient with the different values of the absorbance in both the cases i.e.; assuming and non assuming the cysteine residues. The matrix protein of the all the species of the Ebola virus are not stable as the values are greater than 40. The increment phenomenon of the globular protein explaining the positive factor and volume occupied by the aliphatic side chains are studied then the obtained values are in this order 100.73 (Reston Ebolavirus) > 96.63 (Bundibugyo Ebolavirus) > 96.60 (Sudan Ebolavirus) > 96.32 (Zaire Ebolavirus) > 93.93 (Taiforest Ebolavirus). The Protein repelling capacity in the Bundibugyo Ebolavirus (-0.037) is greater and the repelling capacity of the Tai forest Ebola virus (-0.117) is less.



		Bundibugyo Ebola virus	Sudan Ebola virus	Reston Ebola virus	Zaire Ebola virus	Tai forest ebola virus
Number and composition of amino acids SECOND SECRETED GLYCO PROTEIN	Alanine (Ala)	14(4.6%)	20(6.3%)	15(4.5%)	16(5.4%)	17(5.6%)
	Arginine (Arg)	14(4.6%)	21(6.6%)	20(6.0%)	17(5.7%)	15(5.0%)
	Asparagine (Asn)	18(6.0%)	15(4.7%)	19(5.7%)	13(4.4%)	17(5.6%)
	Aspartic acid (Asp)	12(4.0%)	16(5.0%)	13(3.9%)	13(4.4%)	13(4.3%)
	Cysteine (Cys)	05(1.7%)	05(1.6%)	06(1.8%)	05(1.7%)	06(2.0%)
	Glutamine (Gln)	08(2.6%)	12(3.8%)	11(3.3%)	09(3.0%)	07(2.3%)
	Glutamic acid (Glu)	17(5.6%)	17(5.3%)	18(5.4%)	18(6.1%)	16(5.3%)
	Glycine (Gly)	21(7.0%)	21(6.6%)	24(7.3%)	24(8.1%)	21(7.0%)
	Histidine (His)	08(2.6%)	07(2.2%)	07(2.1%)	05(1.7%)	08(2.6%)
	Iso-Leucine (Ile)	11(3.6%)	16(5.0%)	11(3.3%)	14(4.7%)	14(4.6%)
	Leucine (Leu)	25(8.3%)	29(9.1%)	34(10.3%)	25(8.4%)	25(8.3%)
	Lysine (Lys)	18(6.0%)	16(5.0%)	19(5.7%)	16(5.4%)	19(6.3%)
	Methionine (Met)	03(1.0%)	02(0.6%)	04(1.2%)	01(0.3%)	03(1.0%)
	Phenyl Alanine (Phe)	21(7.0%)	21(6.6%)	17(5.1%)	20(6.7%)	22(7.3%)
	Proline (Pro)	20(6.6%)	18(5.7%)	23(6.9%)	17(5.7%)	18(6.0%)
	Serine (Ser)	16(5.3%)	22(6.9%)	25(7.6%)	20(6.7%)	17(5.6%)
	Threonine (Thr)	27(8.9%)	23(7.2%)	27(8.2%)	26(8.8%)	25(8.3%)
Tryptophan (Trp)	06(2.0%)	07(2.2%)	08(2.4%)	06(2.0%)	06(2.0%)	
Tyrosine (Tyr)	11(3.6%)	11(3.5%)	10(3.0%)	11(3.7%)	09(3.0%)	
Valine (Val)	27(8.9%)	19(6.0%)	20(6.0%)	21(7.1%)	24(7.9%)	
Molecular weight		34184.97	36146.05	37352.50	33391.87	34082.99
Theoretical pI		8.49	8.71	9.14	8.19	8.81
Atomic composition		4793	5065	5235	4681	4786
Total number of positively charged residues		32	37	39	33	34
Total number of Negatively charged residues		29	33	31	31	29
Extinction coefficient assuming cystine residues		49640M ⁻¹ cm ⁻¹	55140M ⁻¹ cm ⁻¹	59275M ⁻¹ cm ⁻¹	49640M ⁻¹ cm ⁻¹	46785M ⁻¹ cm ⁻¹
Absorbance assuming cystine residues		1.452	1.525	1.587	1.487	1.373
Extinction coefficient with out cystine residues		49390M ⁻¹ cm ⁻¹	54890M ⁻¹ cm ⁻¹	58900M ⁻¹ cm ⁻¹	49390M ⁻¹ cm ⁻¹	46410M ⁻¹ cm ⁻¹
Absorbance with out cystine residues		1.445	1.519	1.577	1.479	1.362
Instability Index		29.22	37.58	31.45	31.11	27.34
Aliphatic Index		77.05	78.81	75.08	77.10	79.04
Grand Average of Hydrophaticity		-0.275	-0.339	-0.440	-0.289	-0.220

Table-4: The second Secreted glycol-protein shows approximation in the total number and the composition of the amino acid and difference when compared individual amino acids. The average isotope mass on the protein and water molecule are examined and results are obtained in this order i.e.; 37352.50 > 36146.05 > 34184.97 > 34082.99 > 33391.87. By these studies we concluded that Reston Ebolavirus is having higher molecular weight and Zaire Ebolavirus with small molecular weight. Compute pI/Mw algorithm is mainly used to enhance a region in a 2-D gel to which a protein which is unmodified should allowed to run, and point a region in which modified form of protein should be found if the modifications are documented in the database. When independent amino acid contributions are studied with out considering the secondary and tertiary structure we observed different values of the extinction coefficient with different absorbance both assuming and not assuming the cysteine residues. When we studied the protein stability the secondary secreted protein of all the Ebola species, every organism shows its stability and this can be justified based on the values obtained and all the values obtained are less than 40. When the positive factor explaining the increment phenomenon of the globular proteins and aliphatic side chains are studied the values are obtained in this order i.e.; 79.04 (Tai forest Ebolavirus) > 78.81 (Sudan Ebolavirus) > 77.10 (Zaire Ebolavirus) > 77.05 (Bundibugyo Ebolavirus) > 75.08 (Reston Ebolavirus). The repelling capacity of the protein in the Taiforest Ebolavirus (-0.220) is greater and in the Reston Ebola virus (-0.440) is less.



		Bundibugyo Ebola virus	Sudan Ebola virus	Reston Ebola virus	Zaire Ebola virus	Tai forest ebola virus
Number and composition of amino acids SMALL SECRETED GLYCO PROTEIN	Alanine (Ala)	17(4.6%)	18(4.8%)	16(4.4%)	17(4.7%)	17(4.7%)
	Arginine (Arg)	23(6.2%)	24(6.5%)	23(6.3%)	21(5.8%)	24(6.6%)
	Asparagine (Asn)	17(4.6%)	17(4.6%)	17(4.6%)	13(3.6%)	17(4.7%)
	Aspartic acid (Asp)	12(3.2%)	15(4.0%)	14(3.8%)	13(3.6%)	13(3.6%)
	Cysteine (Cys)	08(2.1%)	08(2.2%)	08(2.2%)	08(2.2%)	09(2.5%)
	Glutamine (Gln)	18(4.8%)	17(4.6%)	16(4.4%)	17(4.7%)	17(4.7%)
	Glutamic acid (Glu)	19(5.1%)	23(6.2%)	20(5.4%)	22(6.0%)	17(4.7%)
	Glycine (Gly)	22(5.9%)	25(6.7%)	23(6.3%)	26(7.1%)	21(5.8%)
	Histidine (His)	08(2.1%)	10(2.7%)	08(2.2%)	04(1.1%)	07(1.9%)
	Iso-Leucine (Ile)	13(3.5%)	19(5.1%)	13(3.5%)	16(4.4%)	15(4.1%)
	Leucine (Leu)	29(7.8%)	33(8.9%)	34(9.3%)	34(9.3%)	30(8.2%)
	Lysine (Lys)	22(5.9%)	22(5.9%)	23(6.3%)	25(6.9%)	23(6.3%)
	Methionine (Met)	03(0.8%)	04(1.1%)	05(1.4%)	01(0.3%)	03(0.8%)
	Phenyl Alanine (Phe)	23(6.2%)	21(5.6%)	20(5.4%)	22(6.0%)	24(6.6%)
	Proline (Pro)	30(8.0%)	23(6.2%)	27(7.4%)	20(5.5%)	25(6.8%)
	Serine (Ser)	23(6.2%)	26(7.0%)	26(7.1%)	25(6.9%)	24(6.6%)
	Threonine (Thr)	37(9.9%)	25(6.7%)	34(9.3%)	34(9.3%)	35(9.6%)
Tryptophan (Trp)	08(2.1%)	08(2.2%)	09(2.5%)	08(2.2%)	08(2.2%)	
Tyrosine (Tyr)	13(3.5%)	14(3.8%)	10(2.7%)	12(3.3%)	11(3.0%)	
Valine (Val)	28(7.5%)	20(5.4%)	21(5.7%)	26(7.1%)	25(6.8%)	
Molecular weight		42471.48	42584.49	41744.64	41175.11	41655.71
Theoretical pI		9.41	8.98	9.31	9.20	9.55
Atomic composition		5959	5960	5852	5809	5854
Total number of positively charged residues		45	46	46	46	47
Total number of Negatively charged residues		31	38	34	35	30
Extinction coefficient assuming cystine residues		63870M ⁻¹ cm ⁻¹	65360M ⁻¹ cm ⁻¹	64900M ⁻¹ cm ⁻¹	62380M ⁻¹ cm ⁻¹	60890M ⁻¹ cm ⁻¹
Absorbance assuming cystine residues		1.504	1.535	1.555	1.515	1.462
Extinction coefficient with out cystine residues		63370M ⁻¹ cm ⁻¹	64860M ⁻¹ cm ⁻¹	64400M ⁻¹ cm ⁻¹	61880M ⁻¹ cm ⁻¹	60390M ⁻¹ cm ⁻¹
Absorbance with out cystine residues		1.492	1.523	1.543	1.503	1.450
Instability Index		39.79	39.34	31.30	34.79	34.57
Aliphatic Index		70.24	74.95	70.90	78.96	72.60
Grand Average of Hydrophaticity		-0.440	-0.469	-0.494	-0.321	-0.398

Table-5: The Small Secreted Glyco-Protein although shows approximation in total number and composition of amino acids, but varies when compared to individual amino acids. In the protein and one water molecule when we calculated the average isotope mass, the total molecular weight of the protein the order is obtained in this way i.e; 42584.49 > 42471.48 > 41744.64 > 41655.71 > 41175.11. Hence by this we can say that the molecular weight of the Sudan Ebolavirus is greater and the molecular weight of the Zaire Ebolavirus is smaller. Compute pI/Mw algorithm is mainly used to enhance a region in a 2-D gel to which an a protein which is unmodified should allowed to run, and point a region in which modified form of protein should be found if the modifications are documented in the database. When the contributions of the independent amino acids are studied with out considering the secondary and tertiary structure we observed different values of the extinction coefficient along with their absorbance values both with assuming and with out assuming cysteine residues. The small secreted glycoprotein in all species of the Ebola virus is stable as it has value less than 40. when we determined the positive factor which explains the increment phenomenon of the globular proteins and volume occupied relatively by the aliphatic side chains we observed the values in this order i.e; 78.96 (Zaire Ebolavirus) > 74.95 (Sudan Ebolavirus) > 72.60 (Taiforest Ebolavirus) > 70.90 (Reston Ebolavirus) > 70.24 (Bundibugyo Ebolavirus). and when we studied the phenomenon of the protein that is exhibiting the ability of repelling from the water, the hydrophobicity value of the Zaire Ebolavirus is more (-0.321) and Reston Ebolavirus (-0.494) is less .



		Bundibugyo Ebola virus	Sudan Ebola virus	Reston Ebola virus	Zaire Ebola virus	Tai forest ebola virus
Number and composition of amino acids SPIKE GLYCO-PROTEIN	Alanine (Ala)	34(5.0%)	43(6.4%)	41(6.1%)	48(7.1%)	41(6.1%)
	Arginine (Arg)	35(5.2%)	33(4.9%)	28(4.1%)	33(4.9%)	28(4.1%)
	Asparagine (Asn)	43(6.4%)	38(5.6%)	43(6.4%)	37(5.5%)	40(5.9%)
	Aspartic acid (Asp)	34(5.0%)	29(4.3%)	30(4.4%)	35(5.2%)	30(4.4%)
	Cysteine (Cys)	12(1.8%)	13(1.9%)	13(1.9%)	12(1.8%)	13(1.9%)
	Glutamine (Gln)	26(3.8%)	28(4.1%)	31(4.6%)	27(4.0%)	25(3.7%)
	Glutamic acid (Glu)	38(5.6%)	40(5.9%)	35(5.2%)	36(5.3%)	37(5.5%)
	Glycine (Gly)	40(5.9%)	51(7.5%)	49(7.2%)	53(7.8%)	47(7.0%)
	Histidine (His)	20(3.0%)	15(2.2%)	15(2.2%)	18(2.7%)	18(2.7%)
	Iso-Leucine (Ile)	39(5.8%)	48(7.1%)	39(5.8%)	42(6.2%)	40(5.9%)
	Leucine (Leu)	52(7.7%)	60(8.9%)	57(8.4%)	51(7.5%)	54(8.0%)
	Lysine (Lys)	26(3.8%)	27(4.0%)	28(4.1%)	30(4.4%)	31(4.6%)
	Methionine (Met)	06(0.9%)	06(0.9%)	10(1.5%)	04(0.6%)	07(1.0%)
	Phenyl Alanine (Phe)	29(4.3%)	24(3.6%)	22(3.2%)	30(4.4%)	32(4.7%)
	Proline (Pro)	56(8.3%)	46(6.8%)	51(7.5%)	35(5.2%)	49(7.2%)
	Serine (Ser)	36(5.3%)	47(7.0%)	55(8.1%)	48(7.1%)	41(6.1%)
	Threonine (Thr)	81(12.0%)	70(10.4%)	65(9.6%)	73(10.8%)	77(11.4%)
	Tryptophan (Trp)	14(2.1%)	14(2.1%)	14(2.1%)	14(2.1%)	14(2.1%)
	Tyrosine (Tyr)	15(2.2%)	16(2.4%)	17(2.5%)	15(2.2%)	12(1.8%)
	Valine (Val)	40(5.9%)	28(4.1%)	34(5.0%)	35(5.2%)	40(5.9%)
Molecular weight		75689.18	74594.18	74416.73	74464.46	74676.43
Theoretical pI		6.01	5.97	5.96	6.16	6.16
Atomic composition		10556	10434	10365	10375	10441
Total number of positively charged residues		61	60	56	63	59
Total number of Negatively charged residues		72	69	65	71	67
Extinction coefficient assuming cystine residues		100100M ⁻¹ cm ⁻¹	101590M ⁻¹ cm ⁻¹	103080M ⁻¹ cm ⁻¹	100100M ⁻¹ cm ⁻¹	95630M ⁻¹ cm ⁻¹
Absorbance assuming cystine residues		1.323	1.362	1.385	1.344	1.281
Extinction coefficient with out cystine residues		99350M ⁻¹ cm ⁻¹	100840M ⁻¹ cm ⁻¹	102330M ⁻¹ cm ⁻¹	99350M ⁻¹ cm ⁻¹	94880M ⁻¹ cm ⁻¹
Absorbance with out cystine residues		1.313	1.352	1.375	1.334	1.271
Instability Index		38.53	43.39	42.36	38.36	37.21
Aliphatic Index		74.69	80.68	75.92	75.77	77.46
Grand Average of Hydrophaticity		-0.466	-0.352	-0.404	-0.380	-0.320

Table-6: The above table describes different physico-chemical properties associated with the Spike glycoProtein of Ebola virus species, in which all forms of spike glycol-protein in all species of the ebola virus species have same number and composition of the aminoacids(approximate values) on the whole but varies when compared with the individual aminoacids. When we calculated average isotope mass of the aminoacid in the protein and one watermolecule the total molecular weight of the protein is estimated to be in this order i.e;75689.18 >7467.43 >74594.18 >74464.46 >74416.73,by this we concluded that the Bundibugyo Ebolavirus having greater molecular weight and Reston Ebolavirus with smaller molecular weight. Compute pI/Mw algorithm is mainly used to enhance a region in a 2-D gel to which an a protein which is unmodified should allowed to run, and point a region in which modified form of protein should be found if the modifications are documented in the database. when we calculated the sums of different aminoacid contributions assuming them as independent with out considering the secondary and tertiary structures, we observed different values of extinction coefficient with and with out assuming cysteine residues with different absorbance. The spike glycol-protein of the sudan ebol virus and reston Ebolavirus is more unstable when compared with the other forms of the virus species as values are greater than 40. when we determined the positive factor which explains the increment phenomenon of the globular proteins and volume occupied relatively by the aliphatic side chains we observed 80.68(Sudan Ebolavirus) >77.46(Tai Forest Ebolavirus) >75.92(Reston Ebolavirus) >75.77 (Zaire Ebolavirus) >74.69(Bundibugyo Ebolavirus) and when we studied the phenomenon of the protein that is exhibiting the ability of repelling from the water, the hydrophobicity value of the Taiforest Ebolavirus(-0.320) is more and for Bundibugyo Ebolavirus(-0.466), it is less.



		Bundibugyo Ebola virus	Sudan Ebola virus	Reston Ebola virus	Zaire Ebola virus	Tai forest ebola virus
MEMBRANE ASSOCIATED PROTEIN	Alanine (Ala)	19(7.6%)	15(6.0%)	15(6.0%)	16(6.4%)	20(8.0%)
	Arginine (Arg)	10(4.0%)	11(4.4%)	11(4.4%)	10(4.0%)	08(3.2%)
	Asparagine (Asn)	13(5.2%)	16(6.4%)	14(5.6%)	17(6.8%)	13(5.2%)
	Aspartic acid (Asp)	10(4.0%)	09(3.6%)	10(4.0%)	09(3.6%)	09(3.6%)
	Cysteine (Cys)	01(0.4%)	01(0.4%)	01(0.4%)	01(0.4%)	01(0.4%)
	Glutamine (Gln)	14(5.6%)	10(4.0%)	12(4.8%)	12(4.8%)	15(6.0%)
	Glutamic acid (Glu)	10(4.0%)	11(4.4%)	08(3.2%)	10(4.0%)	10(4.0%)
	Glycine (Gly)	11(4.4%)	11(4.4%)	13(5.2%)	13(5.2%)	13(5.2%)
	Histidine (His)	07(2.8%)	06(2.4%)	06(2.4%)	07(2.8%)	06(2.4%)
	Iso-Leucine (Ile)	17(6.8%)	18(7.2%)	17(6.8%)	19(7.6%)	17(6.8%)
	Leucine (Leu)	38(15.1%)	33(13.1%)	36(14.3%)	37(14.7%)	39(15.5%)
	Lysine (Lys)	16(6.4%)	13(5.2%)	13(5.2%)	15(6.0%)	17(6.8%)
	Methionine (Met)	08(3.2%)	09(3.6%)	09(3.6%)	09(3.6%)	08(3.2%)
	Phenyl Alanine (Phe)	11(4.4%)	10(4.0%)	13(5.2%)	11(4.4%)	09(3.6%)
	Proline (Pro)	09(3.6%)	13(5.2%)	12(4.8%)	09(3.6%)	10(4.0%)
	Serine (Ser)	20(8.0%)	19(7.6%)	21(8.4%)	21(8.4%)	18(7.2%)
	Threonine (Thr)	17(6.8%)	18(7.2%)	18(7.2%)	17(6.8%)	20(8.0%)
	Tryptophan (Trp)	05(2.0%)	05(2.0%)	05(2.0%)	05(2.0%)	05(2.0%)
Tyrosine (Tyr)	02(0.8%)	05(2.0%)	03(1.2%)	03(1.2%)	02(0.8%)	
Valine (Val)	13(5.2%)	18(7.2%)	14(5.6%)	10(4.0%)	11(4.4%)	
Molecular weight		28163.88	28276.97	28168.88	28218.85	27887.58
Theoretical pI		9.49	9.18	9.55	9.49	9.46
Atomic composition		4031	4034	4019	4023	3998
Total number of positively charged residues		26	24	24	25	25
Total number of Negatively charged residues		20	20	18	19	19
Extinction coefficient assuming cystine residues		30480M ⁻¹ cm ⁻¹	34950M ⁻¹ cm ⁻¹	31970M ⁻¹ cm ⁻¹	31970M ⁻¹ cm ⁻¹	30480M ⁻¹ cm ⁻¹
Absorbance assuming cystine residues		1.082	1.236	1.135	1.133	1.093
Extinction coefficient with out cystine residues		30480M ⁻¹ cm ⁻¹	34950M ⁻¹ cm ⁻¹	31970M ⁻¹ cm ⁻¹	31970M ⁻¹ cm ⁻¹	30480M ⁻¹ cm ⁻¹
Absorbance with out cystine residues		1.082	1.236	1.135	1.133	1.093
Instability Index		35.69	24.36	39.95	36.51	34.47
Aliphatic Index		108.05	106.02	104.50	104.94	107.69
Grand Average of Hydrophaticity		0.040	0.049	0.078	-0.013	0.028

Table-1: The above table describes different physico-chemical properties associated with the membrane associated protein of Ebola virus species, in which all forms of membrane associated protein in all species of the Ebola virus species have same number and composition of the amino acids (approximate values) on the whole but varies when compared with the individual amino acids. When we calculated average isotope mass of the amino acid in the protein and one water molecule the total molecular weight of the protein is estimated to be in this order i.e.; 28276.97 > 28218.85 > 28168.88 > 28163.88 > 27887.58. By this we can say that Sudan Ebolavirus has greater molecular weight and Taiforest Ebolavirus has smaller molecular weight. Compute pI/Mw algorithm is mainly used to enhance a region in a 2-D gel to which a protein which is unmodified should be allowed to run, and point a region in which modified form of protein should be found if the modifications are documented in the database. When we calculated the sums of different amino acid contributions assuming them as independent with out considering the secondary and tertiary structures, we observed similar values of extinction coefficient and absorbance in all species of Ebola virus with and with out assuming cysteine residues. The membrane associated protein of all virus species is more stable. When we determined the positive factor which explains the increment phenomenon of the globular proteins and volume occupied relatively by the aliphatic side chains we observed the values in this order i.e.; 108.05 (Bundibugyo Ebolavirus) > 107.69 (Taiforest Ebolavirus) > 106.02 (Sudan Ebolavirus) > 104.94 (Zaire Ebolavirus) > 104.50 (Reston Ebolavirus) and when we studied the phenomenon of the protein that is exhibiting the ability of repelling from the water, the hydrophobicity value of the Zaire Ebolavirus (-0.013) is smaller and Reston Ebolavirus (0.078) is more.



		Bundibugyo Ebola virus	Sudan Ebola virus	Reston Ebola virus	Zaire Ebola virus	Tai forest ebola virus
MINOR-NUCLEOPROTEIN	Alanine (Ala)	14(4.8%)	20(6.9%)	19(6.6%)	21(7.3%)	16(5.5%)
	Arginine (Arg)	26(9.0%)	23(8.0%)	23(8.0%)	25(8.7%)	23(8.0%)
	Asparagine (Asn)	05(1.7%)	10(3.5%)	13(4.5%)	06(2.1%)	06(2.1%)
	Aspartic acid (Asp)	19(6.6%)	18(6.2%)	18(6.3%)	16(5.6%)	15(5.2%)
	Cysteine (Cys)	07(2.4%)	06(2.1%)	07(2.4%)	08(2.8%)	08(2.8%)
	Glutamine (Gln)	18(6.2%)	16(5.6%)	17(5.9%)	18(6.2%)	17(5.9%)
	Glutamic acid (Glu)	18(6.2%)	16(5.6%)	16(5.6%)	21(7.3%)	19(6.6%)
	Glycine (Gly)	13(4.5%)	14(4.9%)	10(3.5%)	13(4.5%)	12(4.2%)
	Histidine (His)	09(3.1%)	05(1.7%)	09(3.1%)	09(3.1%)	08(2.8%)
	Iso-Leucine (Ile)	12(4.2%)	06(2.1%)	13(4.5%)	10(3.5%)	13(4.5%)
	Leucine (Leu)	35(12.1%)	36(12.5%)	35(12.2%)	32(11.1%)	33(11.4%)
	Lysine (Lys)	14(4.8%)	16(5.6%)	14(4.9%)	15(5.2%)	16(5.5%)
	Methionine (Met)	03(1.0%)	02(0.7%)	04(1.4%)	03(1.0%)	05(1.7%)
	Phenyl Alanine (Phe)	09(3.1%)	10(3.5%)	08(2.8%)	08(2.8%)	08(2.8%)
	Proline (Pro)	13(4.5%)	14(4.9%)	15(5.2%)	15(5.2%)	13(4.5%)
	Serine (Ser)	33(11.4%)	31(10.8%)	32(11.1%)	27(9.4%)	35(12.1%)
	Threonine (Thr)	22(7.6%)	23(8.0%)	16(5.6%)	19(6.6%)	18(6.2%)
	Tryptophan (Trp)	04(1.4%)	04(1.4%)	04(1.4%)	04(1.4%)	04(1.4%)
Tyrosine (Tyr)	04(1.4%)	03(1.0%)	03(1.0%)	04(1.4%)	04(1.4%)	
Valine (Val)	11(3.8%)	15(5.2%)	11(3.8%)	14(4.9%)	16(5.5%)	
Molecular weight		32839.07	32107.22	32400.65	32520.80	32600.14
Theoretical pI		8.46	8.89	8.46	8.40	8.76
Atomic composition		4600	4509	4538	4555	4582
Total number of positively charged residues		40	39	37	40	39
Total number of Negatively charged residues		37	34	34	37	34
Extinction coefficient assuming cystine residues		28335M ⁻¹ cm ⁻¹	26845M ⁻¹ cm ⁻¹	26845M ⁻¹ cm ⁻¹	28460M ⁻¹ cm ⁻¹	28460M ⁻¹ cm ⁻¹
Absorbance assuming cystine residues		0.863	0.836	0.829	0.875	0.873
Extinction coefficient with out cystine residues		27960M ⁻¹ cm ⁻¹	26470M ⁻¹ cm ⁻¹	26470M ⁻¹ cm ⁻¹	27960M ⁻¹ cm ⁻¹	27960M ⁻¹ cm ⁻¹
Absorbance with out cystine residues		0.851	0.824	0.817	0.860	0.858
Instability Index		59.23	48.85	53.74	52.08	57.49
Aliphatic Index		79.31	78.92	82.96	78.26	83.67
Grand Average of Hydrophaticity		-0.624	-0.551	-0.571	-0.607	-0.464

Table-1: The above table describes different physico-chemical properties associated with the minor nucleoprotein of Ebola virus species, in which all forms of minor nucleoprotein in all species of the Ebola virus species have same number and composition of the amino acids (approximate values) on the whole but varies when compared with the individual amino acids. When we calculated average isotope mass of the amino acid in the protein and one water molecule the total molecular weight of the protein is estimated to be in this order i.e.; 32839.07 > 32600.14 > 32520.80 > 21400.65 > 32107.22. By this we can say that Bundibugyo Ebolavirus has greater molecular weight and Sudan Ebolavirus has smaller molecular weight. Compute pI/Mw algorithm is mainly used to enhance a region in a 2-D gel to which a protein which is unmodified should be allowed to run, and point a region in which modified form of protein should be found if the modifications are documented in the database. When we calculated the sums of different amino acid contributions assuming them as independent with out considering the secondary and tertiary structures, we observed different values of extinction coefficient and absorbance in all species of Ebola virus with and with out assuming cysteine residues. The Minor Nucleoprotein of all virus species is more unstable. When we determined the positive factor which explains the increment phenomenon of the globular proteins and volume occupied relatively by the aliphatic side chains we observed the values in this order i.e.; 83.67 (Taiforest Ebolavirus) > 82.96 (Reston Ebolavirus) > 79.31 (Bundibugyo Ebolavirus) > 78.92 (Sudan Ebolavirus) > 78.26 (Zaire Ebolavirus) and when we studied the phenomenon of the protein that is exhibiting the ability of repelling from the water, the hydrophobicity value of the Taiforest Ebolavirus (-0.464) is greater and the value of the Bundibugyo Ebolavirus (-0.624) is less.



		Bundibugyo Ebola virus	Sudan Ebola virus	Reston Ebola virus	Zaire Ebola virus	Tai forest ebola virus
Number and composition of amino acids	Alanine (Ala)	121(5.5%)	122(5.5%)	115(5.2%)	119(5.4%)	127(5.7%)
	Arginine (Arg)	113(5.1%)	131(5.9%)	113(5.1%)	118(5.3%)	119(5.4%)
	Asparagine (Asn)	109(4.9%)	114(5.2%)	126(5.7%)	106(4.8%)	113(5.1%)
	Aspartic acid (Asp)	104(4.7%)	98(4.4%)	107(4.8%)	103(4.7%)	98(4.4%)
	Cysteine (Cys)	44(2.0%)	45(2.0%)	43(1.9%)	43(1.9%)	46(2.1%)
	Glutamine (Gln)	102(4.6%)	91(4.1%)	103(4.7%)	107(4.8%)	97(4.4%)
	Glutamic acid (Glu)	107(4.8%)	106(4.8%)	105(4.7%)	109(4.9%)	106(4.8%)
	Glycine (Gly)	102(4.6%)	107(4.8%)	108(4.9%)	101(4.6%)	103(4.7%)
	Histidine (His)	70(3.2%)	72(3.3%)	75(3.4%)	76(3.4%)	79(3.6%)
	Iso-Leucine (Ile)	148(6.7%)	153(6.9%)	144(6.5%)	147(6.6%)	151(6.8%)
	Leucine (Leu)	255(11.5%)	258(11.7%)	269(12.2%)	250(11.3%)	264(11.9%)
	Lysine (Lys)	121(5.5%)	101(4.6%)	116(5.2%)	113(5.1%)	107(4.8%)
	Methionine (Met)	28(1.3%)	38(1.7%)	34(1.5%)	38(1.7%)	25(1.1%)
	Phenyl Alanine (Phe)	106(4.8%)	97(4.4%)	102(4.6%)	116(5.2%)	100(4.5%)
	Proline (Pro)	105(4.8%)	110(5.0%)	98(4.4%)	102(4.6%)	107(4.8%)
	Serine (Ser)	191(8.6%)	186(8.4%)	184(8.3%)	184(8.3%)	188(8.5%)
	Threonine (Thr)	143(6.5%)	153(6.9%)	132(6.0%)	154(7.0%)	151(6.8%)
	Tryptophan (Trp)	30(1.4%)	29(1.3%)	30(1.4%)	29(1.3%)	29(1.3%)
	Tyrosine (Tyr)	93(4.2%)	89(4.0%)	97(4.4%)	87(3.9%)	85(3.8%)
Valine (Val)	118(5.3%)	110(5.0%)	111(5.0%)	110(5.0%)	115(5.2%)	
Molecular weight		251649.28	251294.24	252549.04	252724.47	250746.25
Theoretical pI		8.64	8.77	8.48	8.55	8.62
Atomic composition		35452	35386	35520	35538	35339
Total number of positively charged residues		234	232	229	231	226
Total number of Negatively charged residues		211	204	212	212	204
Extinction coefficient assuming cystine residues		306320M ⁻¹ cm ⁻¹	294860M ⁻¹ cm ⁻¹	312155M ⁻¹ cm ⁻¹	291755M ⁻¹ cm ⁻¹	289025M ⁻¹ cm ⁻¹
Absorbance assuming cystine residues		1.217	1.173	1.236	1.154	1.153
Extinction coefficient with out cystine residues		303570M ⁻¹ cm ⁻¹	292110M ⁻¹ cm ⁻¹	309530M ⁻¹ cm ⁻¹	289130M ⁻¹ cm ⁻¹	286150M ⁻¹ cm ⁻¹
Absorbance with out cystine residues		1.206	1.162	1.226	1.144	1.141
Instability Index		39.74	42.07	41.60	41.34	40.33
Aliphatic Index		92.08	92.48	92.57	89.80	94.07
Grand Average of Hydrophaticity		-0.218	-0.206	-0.242	-0.230	-0.191

Table-1: The above table describes different physico-chemical properties associated with the RNA dependent RNA Polymerase protein of Ebola virus species, in which all forms of RNA dependent RNA Polymerase in all species of the ebola virus species have same number and composition of the amino acids (approximate values) on the whole but varies when compared with the individual amino acids. When we calculated average isotope mass of the amino acid in the protein and one water molecule the total molecular weight of the protein is estimated to be in this order i.e.; 252724.47 > 252549.04 > 251649.28 > 251294.24 > 250746.25 i.e.; Zaire Ebolavirus with greater molecular weight and Taiforest Ebola virus with small molecular weight. Compute pI/Mw algorithm is mainly used to enhance a region in a 2-D gel to which an a protein which is unmodified should allowed to run, and point a region in which modified form of protein should be found if the modifications are documented in the database. when we calculated the sums of different amino acid contributions assuming them as independent with out considering the secondary and tertiary structures, we observed different values of extinction coefficient and absorbance in all species of Ebola virus with and with out assuming cysteine residues. The RNA dependent RNA polymerase of Bundibugyo Ebolavirus is more stable when compared to other virus species. when we determined the positive factor which explains the increment phenomenon of the globular proteins and volume occupied relatively by the aliphatic side chains we observed the values in this order i.e.; 94.07 (Taiforest Ebolavirus) > 92.57 (Reston Ebolavirus) > 92.48 (Sudan Ebolavirus) > 92.08 (Bundibugyo Ebolavirus) > 89.80 (Zaire Ebolavirus) and when we studied the phenomenon of the protein that is exhibiting the ability of repelling from the water, the hydrophobicity value of the Taiforest Ebolavirus (-0.191) is greater and the value of the Reston Ebolavirus (-0.242) is less



MODULE-2
 MULTIPLE SEQUENCE ALIGNMENT
 MATRIX PROTEIN

```

Sudan      MRRVTVPTAPPAYADIGYFMSMLPIKSSRAVSGIQQKQEVLPGMDFPSNSMRPVADDNID 60
Reston     MRRGVLPTAPPAYNDIAYPMSILPTRPSVIVNETKSDVLA VPGADVPSNSMRPVADDNID 60
TaiForest  MRRILPTAPPEYMEAVYPMRTMNSGADNTASGPNYTTTGVMTNDTPSNSLRPVADDNID 60
Bundibugyo MRRAILPTAPPEYIEAVYPMRTVSTSINSTASGPNFPAPDVMMSDTPSNSLRPVADDNID 60
Zaire     -----
Sudan      HTSHTPNGVASAFILEATVNVISGPKVLMKQIPIWLPLGLIADQKTYSFSTTAAIMLAS 120
Reston     HSSHTPSGVASAFILEATVNVISGPKVLMKQIPIWLPLGLVADQKIYSFDSTTAAIMLAS 120
TaiForest  HPSHTPNSVASAFILEAMVNVISGPKVLMKQIPIWLPLGLVSDQKTYSFSTTAAIMLAS 120
Bundibugyo HPSHTPNSVASAFILEAMVNVISGPKVLMKQIPIWLPLGLVADQKTYSFSTTAAIMLAS 120
Zaire     -----SAFILEAMVNVISGPKVLMKQIPIWLPLGLVADQKTYSFSTTAAIMLAS 50
          *****
Sudan      TITHFGKANNPLVRVNRLLGQGI PDHPLRLLRMLGNQAFLLQEFVLPVLPVLPQYFTFDLTALK 180
Reston     TVTHFGKISNPLVRVNRLLGPGI PDHPLRLLRLLRGNQAFLLQEFVLPVLPVLPVLPQYFTFDLTALK 180
TaiForest  TITHFGKTSNPLVRINRLLGPGI PDHPLRLLRLLRGNQAFLLQEFVLPVLPVLPVLPQYFTFDLTALK 180
Bundibugyo TITHFGKTSNPLVRINRLLGPGI PDHPLRLLRLLRGNQAFLLQEFVLPVLPVLPVLPQYFTFDLTALK 180
Zaire     TITHFGKATNPLVRVNRLLGPGI PDHPLRLLRLLRGNQAFLLQEFVLPVLPVLPVLPQYFTFDLTALK 110
          *:*:*:*:*
Sudan      LVTQPLPAATWTDDETPSNLSGALRPGLSFHPKLRPVLLPGKTKGKKGHVSDLTAPDKIQTI 240
Reston     LITQPLPAATWTDDETPAGAVNALRPGLSLHPKLRPILLPGKTKGKKGHASDLTSPDKIQTI 240
TaiForest  LITQPLPAATWTDDETPAVSTGTLRPGISFHPKLRPILLPGRAGKKGNSDLTSPDKIQAI 240
Bundibugyo LITQPLPAATWTDDETPGTGTLRPGISFHPKLRPILLPGKTKGKKGSSDLTSPDKIQAI 240
Zaire     LITQPLPAATWTDDETPGSGALRPGISFHPKLRPILLPNKSGKKGNSADLTSPDKIQAI 170
          *:*:*:*:*
Sudan      VNLMDQDFKIVPIDPAKSIIGIEVEPELLVHKLTGKKMSQKNGQPIIPVLLPKYIGLDPISP 300
Reston     MNAIPDLKIVPIDPTKNIIVGIEVEPELLVQRLTGKKKQPKNGQPIIPVLLPKYVGLDPISP 300
TaiForest  MNFLQDLKIVPIDPTKNIMGIEVEPELLVHRLTGKKTITTKNGQPIIPILLPKYIGLDPISQ 300
Bundibugyo MNFLQDLKIVPIDPAKNIMGIEVEPELLVHRLTGKKTITTKNGQPIIPILLPKYIGMDPISQ 300
Zaire     MTSLQDFKIVPIDPTKNIIVGIEVEPELLVHKLTGKKVTSKNGQPIIPVLLPKYIGLDEVAS 230
          :. : *:*:*:*:*
Sudan      GDLTMVITPDYDDCHSPASCYSYLSEK----- 326
Reston     GDLTMVITQDCDSCHSPASHPYHMDKQNSYQ 331
TaiForest  GDLTMVITQDCDSCHSPASLPPVNEK----- 326
Bundibugyo GDLTMVITQDCDTCHSPASLPPVSEK----- 326
Zaire     GDLTMVITQDCDTCHSPASLPAVIEK----- 256
          *****
    
```

MEMBRANE ASSOCIATED PROTEIN

```

Sudan      MAKATGRYNLVTPKRELEQGVVFSDLNCFNLVTPVQGWKVVWAGLEFDVNQKGITLLNRL 60
Reston     MAKATGRYNLVPKKDMEKGVIFSDLCNFLITQTQGWKVVWAGIEFDVVSQKGMALLTRL 60
Zaire     MAKATGRYNLISPKKDLEKGVVLSDLNCFNLVSTIQGWKVVWAGIEFDVTHKGMALLHRL 60
TaiForest  MAKATGRYNLISPKKDLEKGLVNLNLDLCTLSVAQTVQGWKVTWAGIEFDVTQKGMALLHRL 60
Bundibugyo MAKATGRYNLVSPKKDLERGLVLSDDLCTFLVDQTIQGWRTVWVWAGIEFDIAQKGMALLHRL 60
          *****
Sudan      KVNDFAWAMTRNLFPHLFKQNPSEVQTP IAWALRVILAAGILDQLMDHSLIEPLSGALN 120
Reston     KTNDFAWAMTRNLFPHLFQNPNSVQSP IAWALRVILAAGLQDQLLDHSLVEPLTGALG 120
Zaire     KTNDFAPAWSMTRNLFPHLFQNPNSVQSP IAWALRVILAAGIQDQLIDQSLIEPLAGALG 120
TaiForest  KTSDFAPAWSMTRNLFPHLFQNPNSVQSP IAWALRVILAAGIQDQLIDQSLIEPLAGALG 120
Bundibugyo KTADFAPAWSMTRNLFPHLFQNSNST IESPLWALRVILAAGIQDQLIDQSLVEPLAGALS 120
          * . *****
Sudan      LIADWLLTSTNHFNMRTQRVKQDLSMRMLSLIRSNINFINFKLETLHVVNKGLLSSVE 180
Reston     LISDWLLTSTNHFNLRTSRVSKDQSLRMLSLIRSNILQFINKLDALHVVNYNGLLSSIE 180
Zaire     LISDWLLTSTNHFNMRTQRVKEQLSLKMLSLIRSNILKFINKLDALHVVNYNGLLSSIE 180
TaiForest  LIADWLLTGTNHFQMRTOQAKEQLSLKMLSLVRSNILKFINQLDALHVVNYNGLLSSIE 180
Bundibugyo LVSDWLLTSTNHFQMRTOQAKEQLSLKMLSLVRSNILKFIQLDALHVVNYNGLLSSIE 180
          *:*:*:*:*
Sudan      IGTPSYAIITRTNMGYLVEVQEPDKSAMD I RHPGPVKFSL LHES TLKPVATPKPSSITS 240
Reston     IGTSTHTIIITRTNMGFLVEVQEPDKSAMNSKRPGPVKFSLLHESAFKPFTRVPQSGMQS 240
Zaire     IGTQNHIIITRTNMGFLVELQEPDKSAMNRKPGPAKFSLLHES TLKFTQGSSTRMQS 240
TaiForest  IGTKSHIIITRTNMGFLVELQEPDKSAMNTRKPGPVKFSLLHES TLKTLAKKPAQTMQA 240
Bundibugyo IGTNRNHTIIITRTNMGFLVELQEPDKSAMNQKPGPVKFSLLHES TKALIKKPAQTMQA 240
          ***
Sudan      LIMEFNSSLAI 251
Reston     LIMEFNSSLAI 251
Zaire     LILEFNSSLAI 251
TaiForest  LILEFNSSLAI 251
Bundibugyo LILEFNSSLAI 251
          *:*:*:*
    
```



MINOR-NUCLEO PROTEIN

Reston MEHSREGRGRSSNMRHNSREPYENPSRSRSLSRDPNQVDRRQPRASQIRVFNLFHRKKT 60
 Sudan MERGREGRGRSRNSRADQQNSGTPQFRTRISRDKTTDYRSRSTSQVVRVTFVHKKGTG 60
 Zaire MEASYERGRPRAARQHSRDGHDHVRARSSSRENYRGEYRQSRASQVVRVTFVHKKRVE 60
 Taiforest MEVVHERGRSRI SRQNRDGP SHLVRRARSSSRASYRSEYHTPRASQIRVPTVFHRKKT 60
 Bundibugyo MDSFHERGRSRTIRQSARDGSPHQVTRSSSRDSSHREYHTPRSSSQVVRVTFVHRKRTD 60
 *: **** * : : * : * * : : * : * : * : * : * : *

Reston ALIVPPAPKIDICPTLKKGFCLDSFKCKKDHQLESLDRELLLLIARCTCGIIESNSQITS 120
 Sudan TLTVPPAPKIDVCPTRLRKGFLCDSNFCKKDHQLESLTDRELLLLIARCTCGSTDSLLNIAA 120
 Zaire PLTVPPAPKIDICPTLKKGFCLDSSFKCKKDHQLESLTDRELLLLIARCTCGSVEQQLNITA 120
 Taiforest LLTVPPAPKIDVCPTRLRKGFLCDSNFCKKDHQLESLTDRELLLLIARCTCGSTEQQLSIVA 120
 Bundibugyo SLTVPPAPKIDICPTLTKRKGFLCDSNFCKKDHQLESLTDRELLLLIARCTCGSLEQQLNITA 120
 * ***** : ***** : ***** : * : * : * : * : * : * : * : * : * : *

Reston PKDMRLANPTAEDFSQGNPKLTLAVLLQIAEHWATRDLRQIEDSKLRALLTLCVAVLTRK 180
 Sudan PKDLRLANPTADDFKQDGSFKLTLKLLVETAEFWANQINNEVDDAKLRALLTSAVLVRK 180
 Zaire PKDSRLANPTADDFQEEGPKITLLTLLIKTAEHWARQDIRTIEDSKLRALLTLCVAVMTRK 180
 Taiforest PKDSRLANPTAEDFQKQDGPVTLTSMLETAEYWSKQDIKNIDSRRLALLTLCVAVMTRK 180
 Bundibugyo PKDTRLANPTADDFQKQDGPKITLLTLETAEYWSKQDIKNIDSRRLALLTLCVAVMTRK 180
 *** ***** * : * : * : * : * : * : * : * : * : * : * : * : *

Reston FSKSQLGLLCEHLRHEGLGQDQADSLEVLVYQRLHSDKGGNFEAALWQQWDRQSLIMFIS 240
 Sudan FSKSQLSQCESHLLRRENLGQDQAESVLEVLVYQRLHSDKGGAFEAALWQQWDRQSLTMFIS 240
 Zaire FSKSQLSLLCEHLRREGLGQDQAEPLVYQRLHSDKGGNFEAALWQQWDRQSLIMFIT 240
 Taiforest FSKSQLSLLCEHLRREGLGQDQSESLEVLVYQRLHSDKGGNFEAALWQQWDRQSLIMFIT 240
 Bundibugyo FSKSQLSLLCEHLRREGLGQDQSESVLEVLVYQRLHSDKGGNFEAALWQQWDRQSLIMFIT 240
 ***** . * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *

Reston AFLNIALQIPCESSVSVVSGLATLYPAQDNSTPSEATNDTTWSSTVE -- 287
 Sudan AFLHVALQLSCESSVTVVLSGLRLLAPPVNEGLPPAPGEYTWSEDST -- 288
 Zaire AFLNIALQIPCESSAVVSVGLR TLVPSQDNEEASTNPGTCSWSDEGTP -- 288
 Taiforest AFLNIALQIPCESSVSVVLSGLRMLIPQSEATEVVTPTSETCTWSEGGSSH 289
 Bundibugyo AFLNIALQIPCESSVSVVLSGLRLLVPSQEDTETSTYTETRAWSEEGGPH 289
 * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *

NUCLEO-PROTEIN

Sudan MDKRVRGSWALGGQSEVDLDYHKILTAGLSVQQGIVRQRVIPVYVVS DLEGICQHIIQAF 60
 Reston MDRGTRRIWVSNQGD TDDLYHKILTAGLTVQQGIVRQKIISVYLVDNLEAMCQLVIQAF 60
 Zaire MDSRFQKIWMAPSLTESDMDYHKILTAGLSVQQGIVRQRVIPVYQVNNLEEICQLIIQAF 60
 Taiforest MESRAHKAWMHTASGFETDYHKILTAGLSVQQGIVRQRVIPVYQVNTNLEEICQLIIQAF 60
 Bundibugyo MDRPRIRTWMMHNTSEVEADYHKILTAGLSVQQGIVRQRIIPVYQISNLEEVCQLIIQAF 60
 *: * : * : * : * : * : * : * : * : * : * : * : * : * : * : *

Sudan EAGVDFQDNADSFLLMLCLHAYQGDRHLFLKSDAVQYLEGHGFRFEVREKENVHRLDEL 120
 Reston EAGIDFQENADSFLLMLCLHAYQGQDYKLFLESNAVQYLEGHGFKFELRKKDGVNRLDEL 120
 Zaire EAGVDFQESADSFLLMLCLHAYQGQDYKLFLESNAVQYLEGHGFRFEVVRKRDGVKRLDEL 120
 Taiforest EAGVDFQESADSFLLMLCLHAYQGQDYKLFLESNAVQYLEGHGFRFEVVRKKEGVKRLDEL 120
 Bundibugyo EAGVDFQDSADSFLLMLCLHAYQGQDYKLFLESNAVQYLEGHGFRFEMKKKEGVKRLDEL 120
 * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *

Sudan LPNVTGGKNLRRTLAAMPEEETTEANAGQFLSFASLFLPKLVVGEKACLEKQVQRQIQVHA 180
 Reston LPAATSGKNI RRTLAALPEEETTEANAGQFLSFASLFLPKLVVGEKACLEKQVQRQIQVHA 180
 Zaire LPVAVSSGKNIKRTLAAMPEEETTEANAGQFLSFASLFLPKLVVGEKACLEKQVQRQIQVHA 180
 Taiforest LPAASSGKNI RRTLAAMPEEETTEANAGQFLSFASLFLPKLVVGEKACLEKQVQRQIQVHS 180
 Bundibugyo LPAASSGKNIKRTLAAMPEEETTEANAGQFLSFASLFLPKLVVGEKACLEKQVQRQIQVHA 180
 ** . : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *

Sudan EQGLIQYPTSWQSVGHMMVIFRLMRTNFLIKFLLIHQGMHMVAGHDANDTVISNSVAQAR 240
 Reston EQGLIQYPTAWQSVGHMMVIFRLMRTNFLIKYLLIHQGMHMVAGHDANDAVIANSVAQAR 240
 Zaire EQGLIQYPTAWQSVGHMMVIFRLMRTNFLIKFLLIHQGMHMVAGHDANDAVIANSVAQAR 240
 Taiforest EQGLIQYPTAWQSVGHMMVIFRLMRTNFLIKFLLIHQGMHMVAGHDANDAVIANSVAQAR 240
 Bundibugyo EQGLIQYPTSWQSVGHMMVIFRLMRTNFLIKFLLIHQGMHMVAGHDANDAVIANSVAQAR 240
 * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *

Sudan FSGLLIVKTVLDHILQKTDLGVRLHPLARTAKVKNEVNSFKAALGSLAKHGEYAPFARLL 300
 Reston FSGLLIVKTVLDHILQKTDQVRLHPLARTAKVRNEVNAFKAALS SLAKHGEYAPFARLL 300
 Zaire FSGLLIVKTVLDHILQKTERGVRLHPLARTAKVKNEVNSFKAALS SLAKHGEYAPFARLL 300
 Taiforest FSGLLIVKTVLDHILQKTEHGVRLHPLARTAKVKNEVNSFKAALS SLAQHGEYAPFARLL 300
 Bundibugyo FSGLLIVKTVLDHILQKTEHGVRLHPLARTAKVKNEVNSFKAALS SLAQHGEYAPFARLL 300
 * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *

Sudan NLSGVNNLEHGLYPQLSAIALGVATAHGSTLAGVNVGEQYQQLREAA TEAEKQLQYQYAES 360
 Reston NLSGVNNLEHGLYPQLSAIALGVATAHGSTLAGVNVGEQYQQLREAA TEAEKQLQYQYAES 360
 Zaire NLSGVNNLEHGLFPQLSAIALGVATAHGSTLAGVNVGEQYQQLREAA TEAEKQLQYQYAES 360
 Taiforest NLSGVNNLEHGLFPQLSAIALGVATAHGSTLAGVNVGEQYQQLREAA TEAEKQLQYQYAES 360
 Bundibugyo NLSGVNNLEHGLFPQLSAIALGVATAHGSTLAGVNVGEQYQQLREAA TEAEKQLQYQYAES 360
 * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *



Zaire ST---ASDTPSATTA-----GPPKAEN---TNTSKSTDFL---DPA----- 447
 Taiforest KT---TSQPTNSTEST-----TLNPTSE-----FSSRGTFSSPTVENTESHAELEGKT 453
 Bundibugyo TL---ANNPPDNTFES-----TPQDGE-----RTSSHTTSPSRPVPSTIHPHTRETHI 453
 Reston NTASIEDSPPSASNETIYHSEMDFIQSSNNSAQSPQTKTTPAFTTSPMTQDPQETANSSK 466
 Sudan TAPSPFA-----QTPPTH-----TSGP-----SVMATEE 440

Zaire TTTSPQNHSETAG-----NNNTHHQDTGEESASSGKLGITNTIAGVAGLITGRRR 500
 Taiforest PTTLPEQHTAASA-----IPRAVHPDELSPGFLTNTIRGVNLLTGSRRR 500
 Bundibugyo PTTMTTSHDTSN-----RPNPIDISESTEPGLTNTTRGAANLLTGSRRR 500
 Reston PGTSPGSAAGPS-----QPGLTINTVSKVADSLSPTRKQK 501
 Sudan PTTFPGSSPGPTTEAPTLTTPENITTAVKTVLPQESTSNGLITSTVTGILGSLGLRKR 500

Zaire REAIVNAQPKCNPNLHYWTTQDEGAAIGLAWIPYFPGPAEAGIYIEGLMHNQDGLICGLRQ 560
 Taiforest RDVTPNTQPKCNPNLHYWTALDEGAAIGLAWIPYFPGPAEAGIYIEGLMHNQDGLICGLRQ 560
 Bundibugyo REITLRTQAKCNPNLHYWTTQDEGAAIGLAWIPYFPGPAEAGIYIEGLMHNQDGLICGLRQ 560
 Reston RSVRQNTANKCNPDLIYWTAVDEGAAIGLAWIPYFPGPAEAGIYIEGLMHNQDGLICGLRQ 561
 Sudan RQNTKATGKCNPNLHYWTAQEQHNAAGIAWIPYFPGPAEAGIYIEGLMHNQNALVCGLRQ 560

Zaire LANETTQALQFLRATTELRTFSILNRKAIIDFLLQRWGGTCHILGPDCCIEPHDWTKNIT 620
 Taiforest LANETTQALQFLRATTELRTFSILNRKAIIDFLLQRWGGTCHILGPDCCIEPHDWTKNIT 620
 Bundibugyo LANETTQALQFLRATTELRTFSILNRKAIIDFLLQRWGGTCHILGPDCCIEPHDWTKNIT 620
 Reston LANETTQALQFLRATTELRTYSLLNRKAIIDFLLQRWGGTCHILGPDCCIEPHDWTKNIT 621
 Sudan LANETTQALQFLRATTELRTYTILNRKAIIDFLLRRWGGTCHILGPDCCIEPHDWTKNIT 620

Zaire DKIDQIIHDFVDKTLFDQGDNDNWWTGWQRWIPAGIGITGVI IIAVIALFCICKFVF 676
 Taiforest DKIDQIIHDFVDNPNLNDNNDNWWTGWQRWIPAGIGITGVI IIAI IALLCICKFLL 676
 Bundibugyo DKIDQIIHDFIDKPLFDQTDNDNWWTGWQRWIPAGIGITGVI IIAVIALFCICKFLL 676
 Reston DEINQIKHDFIDNPLFDHGDVLDNLTGWQRWIPAGIGITGVI IIAI IALLCICKILC 677
 Sudan DKINQIIHDFIDNPLFNQDNDNWWTGWQRWIPAGIGITGVI IIAI IALLCICKLLC 676

RNA DEPENDENT RNA POLYMERASE

Zaire -MATQHTQYDPDARLSSPIVLDQCDLVTRACGLYSSYSLNPQLRNCKLPKHIYRLKYDVTV 59
 Tai -MATQHTQYDPDARLSSPIVLDQCDLVTRACGLYSSYSLNPQLKNCRLPKHIYRLKYDTTV 59
 Bundibugyo -MATQHTQYDPDARLSSPIVLDQCDLVTRACGLYSSYSLNPQLKNCRLPKHIYRLKFDATV 59
 Sudan MMATQHTQYDPDARLSSPIVLDQCDLVTRACGLYSEYSLNPKLKTCRLPKHIYRLKYDTIV 60
 Reston -MATQHTQYDPDARLSSPIVLDQCDLVTRACGLYSSYSLNPQLRQCKLPKHIYRLKFDTIV 59

Zaire TKFLSDVPVATLPIIDFIVPVLKALSNGNFCVPEPRCQFLDEI IKYTMQDALFLKYLYL 119
 Tai TEFLSDVPVATLPADFLVPTFLRSLGNGSCPIDPKCSQFLDEIVNYTLQDIRFLNYLYN 119
 Bundibugyo TKFLSDVPVATLPIIDYLPDLLRSLGGLCPVEPKCSQFLDEIVSYVLQDAREFLRYFR 119
 Sudan LRFISDVPVATIPIDYIAPMLINVLADSKNVPLEPPCLSFLDEIVNYTVQDAFLNYLYM 120
 Reston SKFLSDTPVATLPIIDYLPILLRSLTGHDRPLTPTCNQFLDEIINYTLHDAFLDYLYL 119

Zaire NVGAQEDCVDEHFQEKILSSIQNEFLHQMFWDYDLA ILTRRGRRLNRGNSRSTWFVHDDL 179
 Tai RAGVHNDHVDRDFGQKIRNLICDNEVLHQMFHWYDLA I LARRGRRLNRGNSRSTWFASDNL 179
 Bundibugyo HVGVHDDNVGKNFEPKIKALIDNEFLHQMFHWYDLA I LTRRGRRLNRGNSRSTWFANDL 179
 Sudan QIKTQEGVITDQLKQIRRV I HKNRYLSALFVHDLA I LTRRGRMNRGNSRSTWFVTNEV 180
 Reston ATGAQDHLTN IATREKLNK I LNNDYVHQLFFVHDL I LARRGRRLNRGNSRSTWFVHDF 179

Zaire IDILGYGDYVFWKIPISMLPLNTQGI PHAAMDWYQASVFKAEVQGHGTHI VSVSTADVLIM 239
 Tai VDILGYGDYVFWKIPISMLPLVDTQGL PHAAMDWYHESVFKAEVQGHGTHI VSVSTADVLIM 239
 Bundibugyo IDILGYGDYVFWKIPISMLPLNTEGI PHAAMDWYHESVFKAEVQGHGTHI VSVSTADVLIM 239
 Sudan VDILGYGDYVFWKIPIALPLMNTANVPHASTDWYQPN I FKEVQGHGTHI VSVSTADVLIM 240
 Reston IDILGYGDYVFWKIPISMLPLVPTIDGVPHAATDWYQPTL FKSILGHSQILSVSTAEILIM 239

Zaire CKDLITCRFNTLLISKIAEIEDPVCS DYPNFKIVSMLYQSGDYLLSILGSDGYKI IKFLE 299
 Tai CKDIITCRFNTLLIAAVANLEDSVHSDYPLPETVSDLYKAGDYLLSILGSEGYKVIKLE 299
 Bundibugyo CKDIITCRFNTLLIAALANLEDSICSDYPPPETISNLYKAGDYLLSILGSEGYKVIKLE 299
 Sudan CKDLVTSRFNTLLIAELARLEDPVSADYPLVDNIQSLYNAGDYLLSILGSEGYKI IKYLE 300
 Reston CKDIITCRFNTSLIASIAKLEDDVSDYPPDSDILKIYNAGDYV I SILGSEGYKI IKYLE 299

Zaire PLCLAKIQLCSKYTERKGRFLTQMH LAVNHTLEETEMRALKPSQAQKIREFHRTLIRLE 359
 Tai PLCLAKIQLCSNYTERKGRFLTQMH LAVNHTLEELTGSRELRPQQ IRKRVREHFHQLINLK 359
 Bundibugyo PLCLAKIQLCSNYTERKGRFLTQMH LAVNHTLEELIEGRGLKSQQDWMKREFHRLVNLK 359
 Sudan PLCLAKIQLCSQYTERKGRFLTQMH LAVIQTLELRLNRLGLKKSQLSKIREFHQLLLRLR 360
 Reston PLCLAKIQLCSKFTTERKGRFLTQMH LSVINDLRELSNRRLKDYQEQKIRDFHKLILLQLQ 359

Zaire MTPQQLCELFSIQKHGHPVLHSETAIQKVKKHATVLKALRP I VIFETYCVFKYSIAKHY 419
 Tai ATPQQLCELFSVQKHGHPVLHSEKAIQKVKKHATVIKALRP I IIFETYCVFKYSIAKHY 419
 Bundibugyo STPQQLCELFSVQKHGHPVLHSEKAIQKVKKHATVIKALRP I IIFETYCVFKYSIAKHY 419
 Sudan STPQQLCELFSIQKHGHPVLHSEKAIQKVKKHATVLKALRP I IIFETYCVFKYSIAKHY 420
 Reston LSPQQLCELFSVQKHGHPILHSEKAIQKVKRHATILKALRP NVI FETYCVFKYNIKHY 419

Zaire FDSQGSWYSVTSDRNLTPLGNSYIKRNQFPPLPMIKELLWEFYHLDHPPLFSTKI ISDLS 479
 Tai FDSQGTWYSVTSDRCLTPGLSSYIKRNQFPPLPMIKELLWEFYHLDHPPLFSTKVISDLS 479
 Bundibugyo FDSQGSWYSVSDKHLTPGLHSYIKRNQFPPLPMIKDLLWEFYHLDHPPLFSTKI ISDLS 479
 Sudan FDSQGTWYSVTSDRCLTPGLNSYIRRNQFPPLPMIKDLLWEFYHLDHPPLFSTKI ISDLS 480
 Reston FDSQGTWYSVTSDRNLTPLGNSYIKRNHFPPLPMIKDLLWEFYHLDHPPLFSTKVISDLS 479



Zaire IFIKDRATAVERTCWDVAFEPNVLGYNPPHKFSTKRVPPEQFLEQENFSIENVLSYAQKLE 539
 Tai IFIKDRATAVEKTCWDVAFEPNVLGYNPPNKFATKRVPPEQFLEQENFSIESVLHYAQRLE 539
 Bundibugyo IFIKDRATAVEKTCWDVAFEPNVLGYSPPNKFSTKRVPPEQFLEQENFSIDSVLTYAQRLE 539
 Sudan IFIKDRATAVEQTCWDVAFEPNVLGYSPPYRFNTKRVPPEQFLEQEDFSIESVLTYAQRLE 540
 Reston IFIKDRATAVEQTCWDVAFEPNVLGYNPPNKFSTKRVPPEQFLEQEDFSIESVNLNYAQELH 539
 *****:*****. * : * *****:*****. * * * *

Zaire YLLPQYRNFSFSLKEKELNVGRFTFGKLPYPTRNVQTLCEALLADGLAKAFPSNMMVVTER 599
 Tai YLLPEYRNFSFSLKEKELNIGRAFGLKLPYPTRNVQTLCEALLADGLAKAFPSNMMVVTER 599
 Bundibugyo YLLPQYRNFSFSLKEKELNVGRFTFGKLPYPTRNVQTLCEALLADGLAKAFPSNMMVVTER 599
 Sudan YLLPQNRNFSFSLKEKELNVGRFTFGKLPYLTRNVQTLCEALLADGLAKAFPSNMMVVTER 600
 Reston YLLPQNRNFSFSLKEKELNIGRTFGKLPYLTRNVQTLCEALLADGLAKAFPSNMMVVTER 599
 *****:*****:***** ***** *****

Zaire EQKESLLHQASWHHTSDDFGEHATVRGSSFVTDLEKYNLAFRYEFTAPFIEYCNRCYGVK 659
 Tai EQKESLLHQASWHHTSDDFGENATVRGSSFVTDLEKYNLAFRYEFTAPFIEYCNRCYGVK 659
 Bundibugyo EQKESLLHQASWHHTSDDFGENATVRGSSFVTDLEKYNLAFRYEFTAPFIEYCNRCYGVK 659
 Sudan EQKESLLHQASWHHTSDDFGEHATVRGSSFVTDLEKYNLAFRYEFTAPFIEYCNRCYGVK 660
 Reston EQKESLLHQASWHHTSDDFGENATVRGSSFVTDLEKYNLAFRYEFTAPFIEYCNRCYGVK 659
 *****:*****:***** ***** *****

Zaire NVFNWMHYTI PQCYMHVSDYNNPPHNLTLNDRNPPEGPSSYRGHMGIEGLQQLWTSI 719
 Tai NLFNWMHYTI PQCYIHVSDYNNPPHNSLENRENPEGPSSYRGHMGIEGLQQLWTSI 719
 Bundibugyo NLFNWMHYTI PQCYIHVSDYNNPPHNSLENREDPEGPSSYRGHMGIEGLQQLWTSI 719
 Sudan NVFDWMHFLI PQCYMHVSDYNNPPHNVTLNREYPPPEGPSAYRGHMGIEGLQQLWTSI 720
 Reston NVFNWMHYLI PQCYMHVSDYNNPPHNVNLSNREYPPPEGPSSYRGHMGIEGLQQLWTSI 719
 *:*****:*****:*****. :. * * * : *****:*****:*****

Zaire SCAQISLVEIKTGFKLRSVAVMGDNQCITVLSVFPLETDADEQEQSAEDNAARVAASLAKV 779
 Tai SCAQISLVEIKTGFKLRSVAVMGDNQCITVLSVFPLETESSEQELSSEDNAARVAASLAKV 779
 Bundibugyo SCAQISLVEIKTGFKLRSVAVMGDNQCITVLSVFPLETDSNEQEHSSEDNAARVAASLAKV 779
 Sudan SCAQISLVEIKTGFKLRSVAVMGDNQCITVLSVFPLESPNEQERCAEDNAARVAASLAKV 780
 Reston SCAQISLVEIKTGFKLRSVAVMGDNQCITVLSVFPPLKTDPEEQEQSAEDNAARVAASLAKV 779
 *****:*****:*****:*****:*****:*****:*****

Zaire TSACGIFLKPDETFVHSGFIYFGKKQYLNQVQLPQSLKTATRMAPLSDAIFDDLQGTLAS 839
 Tai TSACGIFLKPDETFVHSGFIYFGKKQYLNQVQLPQSLKTATRIAPLSDAIFDDLQGTLAS 839
 Bundibugyo TSACGIFLKPDETFVHSGFIYFGKKQYLNQVQLPQSLKTATRIAPLSDAIFDDLQGTLAS 839
 Sudan TSACGIFLKPDETFVHSGFIYFGKKQYLNQVQLPQSLKTAARMAPLSDAIFDDLQGTLAS 840
 Reston TSACGIFLKPDETFVHSGFIYFGKKQYLNQVQLPQSLKTAARMAPLSDAIFDDLQGTLAS 839
 *****:*****:*****:*****:*****:*****:*****

Zaire IGTAHERSISETRHIFPCRIATAAFHTFFSVRILQYHHLGFNKGFDLQGLTLGKPLDFGTI 899
 Tai IGTAHERSISETRHVPCRVAAAFHTFFSVRILQYHHLGFNKGTDLQGLSLSKPLDFGTI 899
 Bundibugyo IGTAHERSISETRHVPCRVAAAFHTFFSVRILQYHHLGFNKGTDLQGLSLSKPLDFGTI 899
 Sudan IGTAHERSISETRHILPCRVAAAFHTYFSVRILQHHLGFNKGSDLQGLLAINKPLDFGTI 900
 Reston IGTAHERAISETRHILPCRIATAAFHTYFAVRILQYHHLGFNKGIDLQGLSLSKPLDYGTI 899
 *****:*****:*****:*****:*****:*****:*****:*****:*****

Zaire SLALAVPQVLGGLSFLNPEKCFYRNLDGDPVTSGLFQQLKTYLRMIEMDDLFLPLIAKNPGN 959
 Tai TLALAVPQVLGGLSFLNPEKCFYRNLDGDPVTSGLFQQLKTYLQMIHMDLFLPLIAKNPGN 959
 Bundibugyo TLALAVPQVLGGLSFLNPEKCFYRNLDGDPVTSGLFQQLRQTYLQMINMDLFLPLIAKNPGN 959
 Sudan ALSLAVPQVLGGLSFLNPEKCLYRNLDGDPVTSGLFQQLKHYLSMVGMSDI FHALIAKSPGN 960
 Reston TLTALAVPQVLGGLSFLNPEKCFYRNLDGDPVTSGLFQQLRVYLEMVNMKDLFCPLISKPNPGN 959
 *:*****:*****:*****:*****:*****:*****:*****:*****:*****

Zaire CTAIDFVLNPSGLNVPGSQDLTSFLRQIVRRTITLSAKNKLINTLFHASADLEDEMVCWK 1019
 Tai CSAIDFVLNPSGLNVPGSQDLTSFLRQIVRRTITLSAKNKLINTLFHASADLEDEMVCWK 1019
 Bundibugyo CSAIDFVLNPSGLNVPGSQDLTSFLRQIVRRTITLSAKNKLINTLFHASADLEDEMVCWK 1019
 Sudan CSAIDFVLNPSGLNVPGSQDLTSFLRQIVRRTITLSAKNKLINTLFHASADLEDEMVCWK 1020
 Reston CSAIDFVLNPSGLNVPGSQDLTSFLRQIVRRTITLTARNKLINTLFHASADLEDEMVCWK 1019
 *:*****:*****:*****:*****:*****:*****:*****:*****:*****

Zaire LLSSTPVMSRFAADIFSRTPSGKRLQILGYLEGTRTLLASKI INNNSETPVLDRLRKITL 1079
 Tai LLSSTPVMSRFAADIFSRTPSGKRLQILGYLEGTRTLLASKI INNNSETPVLDRLRKITL 1079
 Bundibugyo LLSSTPVMSRFAADIFSRTPSGKRLQILGYLEGTRTLLASKVINNNAETPILDRLRKITL 1079
 Sudan LLSSTPVMSRFAADIFSRTPSGKRLQILGYLEGTRTLLASKMISNNAETPILDRLRKITL 1080
 Reston LLSSNPVMSRFAADIFSRTPSGKRLQILGYLEGTRTLLASKI INNNSETPVLDRLRKITL 1079
 *****:*****:*****:*****:*****:*****:*****:*****:*****

Zaire QRWSLWFSYLDHCDNLAEALTOITCTVDLAQILREYSWAHILEGRPLIGATLPCMI EQF 1139
 Tai QRWSLWFSYLDHCDQVLADALTOITCTVDLAQILREYTWAHILEGRQLIGATLPCILEQL 1139
 Bundibugyo QRWSLWFSYLDHCDQVLADALIKVSCVDLAQILREYTWAHILEGRQLIGATLPCMI EQF 1139
 Sudan QRWNLWFSYLDHCDPALMEAIQPKCTVDLAQILREYSWAHILEGRPLIGATLPCMI EQF 1140
 Reston QRWNLWFSYLDHCDQLLADALQKISCTVDLAQILREYTWASHILEGRSLIGATLPCMVEQF 1139
 * * * * * : * * * * * : * * * * * : * * * * * : * * * * * : * * * * * : * * * * *

Zaire KVEWLPKPYEQCPQCSNAKQPGGKPEFVSVAVKHHIVSAWPNASRISWTIGDGPYIGSRTE 1199
 Tai NVIWLPKPYEQCPKCAKSNPKGPEFVSVIAIKKHVSAWPNQSRISWTIGDGPYIGSRTE 1199
 Bundibugyo NVEWLPKPYEQCPKCAKSNPKGPEFVSVIAIKKHVSAWPNQSRISWTIGDGPYIGSRTE 1199
 Sudan QVTWLPKPYEQCPVCSSTN--NSSPYVSVALKRNVVSAWPNASRISWTIGDGPYIGSRTE 1198
 Reston KVKWLGQYEPCEPELNKK--GSNAYVSAVKDQVSAWPNASRISWTIGDGPYIGSRTE 1197
 :. * * * * * : * * * * * : * * * * * : * * * * * : * * * * * : * * * * *

Zaire DKIGQPAIKPKCPSAALREAIELASRLTWVTVQGSNSDLLKPFLEARVNLVQBEILQMT 1259
 Tai DKIGQPAIKPKCPSAALREAIELASRLTWVTVQGSNSDLLKPFLEARVNLVQBEILQMT 1259
 Bundibugyo DKIGQPAIKPKCPSAALREAIELASRLTWVTVQGSNSDLLKPFLEARVNLVQBEILQMT 1259
 Sudan DKIGQPAIKPRCPSAALREAIELASRLTWVTVQGSNSDLLKPFLEARVNLVQBEILQMT 1258
 Reston DKIGQPAIKPRCPSAALKEAIELASRLTWVTVQGSNSDQLIRPFLEARVNLVSEVLQMT 1257
 *****:*****:*****:*****:*****:*****:*****:*****:*****



```

Zaire      LIVQALKHNGTWQAEFKKLPPELISVCNRFYHIRDCNCEERFLVQTLYLHRMQDSEVKLIE 2195
Tai       LIIQALKHNCTWQEELRALPDLISVCTRFRYHTRNCSCENRFLVQTLYLHRMQDSEIKLID 2195
Bundibugyo LVVQALKHNCLWQEELRTPDLINVCNRFYHIRDCSCEDRFLIQTLYLTRMQDSEAKLME 2195
Sudan     LVIRALKNNSTWHHELKLLPELIGVCHRFNHRNCTCSEFLVQTLYLHRMSDAEIKLMD 2192
Reston    LIVQALKNNSSWYTELKKLPEVINVCNRFYHHTNCECQEKFFVQTLYLQRLRDAEIKLIE 2193
          *:::***:* * *: **::*.** ** * :.* *:::*** ** * :.* *:::
    
```

```

Zaire      RLTGLLSLFPDGLYRFD--          2212
Tai       RLTGLLSLCPNGFFR----          2210
Bundibugyo RLTGFLGLYPNGINA----          2210
Sudan     RLTSLVNMFPEGFRSSV-          2210
Reston    RLTGLMRFYPEGLIYSNHT          2212
          ***.: : *:*:
    
```

MODULE-3

Nucleoprotein:

	Tai Forest Ebolavirus	Bundibugyo Ebolavirus	Sudan Ebolavirus	Zaire Ebolavirus	Reston Ebolavirus
Gene number; Exon number and Exon type	Gene Number = 01 Exon Number = 01 Type of Exon = Sngl	Gene Number = 01 Exon Number = 01 Type of Exon = Sngl	Gene Number = 01 Exon Number = 01 Type of Exon = Sngl	Gene Number = 01 Exon Number = 01 Type of Exon = Sngl	Gene Number = 01 Exon Number = 01 Type of Exon = Sngl
	Gene Number = 01 Exon Number = 02 Type of Exon = PlyA		Gene Number = 01 Exon Number = 02 Type of Exon = PlyA	Gene Number = 01 Exon Number = 02 Type of Exon = PlyA	Gene Number = 01 Exon Number = 02 Type of Exon = PlyA
Type of DNA Strand + = Input Strand - = Output Strand	E-1(=Input strand) E-2(=Input Strand)	E-1(= Input Strand)	E-1(=Input Strand) E-2(=Input Strand)	E-1(=Input Strand) E-2(=Input strand)	E-1(=Input Strand) E-2(=Input Strand)
Beginning of Exon/Signal	E-1=409 E-2=2677	E-1=403	E-1 = 403 E-2 = 2734	E-1= 415 E-2= 2743	E-1= 409 E-2= 2947
Ending of Exon/signal	E-1=2628 E-2=2682	E-1=2622	E-1= 2619 E-2 = 2739	E-1= 2634 E-2= 2748	E-1= 2628 E-2= 2952
Length of Exon/signal	E-1=2220 E-2=06	E-1=2220	E-1 = 2217 E-2 = 06	E-1= 2220 E-2= 06	E-1= 2220 E-2= 06
Reading Frame	E-1=0 E-2=()	E-1=0	E-1 = 0 E-2 = ()	E-1= 0 E-2 = ()	E-1 = 0 E-2 = ()
Net-Phase of Exon/Signal	E-1=0 E-2=()	E-1=0	E-1 = 0 E-2 = ()	E-1 = 0 E-2 = ()	E-1 = 0 E-2 = ()
Initiation signal/3'-Splice site score	E-1=99 E-2=()	E-1=75	E-1 = 81 E-2 = ()	E-1= 55 E-2= ()	E-1 = 71 E-2 = ()
Termination signal/5'-Splice site score	E-1=28 E-2=()	E-1=42	E-1 = 55 E-2 = ()	E-1 = 42 E-2 = ()	E-1 = 41 E-2 = ()
Coding Region score	E-1=1622 E-2=()	E-1=1481	E-1 = 1662 E-2 = ()	E-1 = 1366 E-2 = ()	E-1 = 1420 E-2 = ()
Probability of Exon	E-1=0.776 E-2=()	E-1=0.478	E-1 = 0.998 E-2 = ()	E-1 = 0.578 E-2 = ()	E-1 = 0.921 E-2 = ()
Exon score	E-1=151.05 E-2=-3.64	E-1=135.75	E-1 = 155.98 E-2 = 1.05	E-1 = 122.45 E-2 = 1.05	E-1 = 129.35 E-2 = -1.75

Polymerase complex protein

	Tai Forest Ebolavirus	Bundibugyo Ebolavirus	Sudan Ebolavirus	Zaire Ebolavirus	Reston Ebolavirus
Gene number; Exon number and Exon type	Gene number = 01 Exon number = 01 Type of Exon = Init	Gene number = 01 Exon number = 01 Type of Exon = Sngl	Gene number = 01 Exon number = 01 Type of Exon = Sngl	Gene number = 01 Exon number = 01 Type of Exon = Sngl	Gene number = 01 Exon number = 01 Type of Exon = Sngl
		Gene number = 01 Exon number = 02 Type of Exon = PlyA	Gene number = 01 Exon number = 02 Type of Exon = PlyA	Gene number = 01 Exon number = 02 Type of Exon = PlyA	Gene number = 01 Exon number = 02 Type of Exon = PlyA
Type of DNA Strand + = Input Strand - = Output Strand	E-1(=Input Strand)	E-1(= Input strand) E-2(=Input Strand)	E-1(=Input strand) E-2(=Input Strand)	E-1(=Input strand) E-2(=Input strand)	E-1(=Input strand) E-2(=Input strand)
Beginning of Exon/Signal	E-1 = 89	E-1 = 89 E-2 = 1303	E-1 = 126 E-2 = 1343	E-1= 98 E-2= 1238	E-1 = 137 E-2 = 1218
Ending of Exon/signal	E-1 = 1004	E-1 = 1114 E-2 = 1308	E-1 = 1115 E-2 = 1348	E-1 = 1120 E-2 = 1243	E-1 = 1126 E-2 = 1223
Length of Exon/signal	E-1 = 916	E-1 = 1026 E-2 = 06	E-1 = 990 E-2 = 06	E-1 = 1023 E-2 = 06	E-1 = 990 E-2 = 06
Reading Frame	E-1 = 01	E-1 = 01 E-2 = ()	E-1 = 02 E-2 = ()	E-1 = 01 E-2 = ()	E-1 = 01 E-2 = ()
Net-Phase of Exon/Signal	E-1 = 01	E-1 = 00 E-2 = ()	E-1 = 00 E-2 = ()	E-1 = 00 E-2 = ()	E-1 = 00 E-2 = ()



Initiation signal/3'-Splice site score	E-1 =42	E-1 = 75 E-2 = ()	E-1 =71 E-2 = ()	E-1 = 66 E-2 = ()	E-1 = 69 E-2 = ()
Termination signal/5'-Splice site score	E-1=3	E-1 = 45 E-2=()	E-1 = 43 E-2 = ()	E-1 = 52 E-2 = ()	E-1 = 48 E-2 = ()
Coding Region score	E-1 =568	E-1 = 296 E-2=()	E-1 = 819 E-2 = ()	E-1 = 843 E-2 = ()	E-1 = 813 E-2 = ()
Probability of Exon	E-1 = 0.455	E-1 = 0.283 E-2 = ()	E-1 = 0.534 E-2 = ()	E-1 = 0.847 E-2 = ()	E-1 = 0.846 E-1 = ()
Exon score	E-1=37.93	E-1 = 21.09 E-2 = 1.05	E-1 = 72.31 E-2 = -1.75	E-1 = 75.31 E-2 = -1.75	E-1 = 72.01 E-2 = -1.75

Matrix Protein:

	Tai Forest Ebolavirus	Bundibugyo Ebolavirus	Sudan Ebolavirus	Zaire Ebolavirus	Reston Ebolavirus
Gene number;Exon number and Exon type	Gene number = 01 Exon number = 01 Type of Exon = Init	Gene number = 01 Exon number = 01 Type of Exon = Init	Gene number = 01 Exon number = 01 Type of Exon = Snagl	Gene number = 01 Exon number = 01 Type of Exon = Snagl	Gene number = 01 Exon number = 01 Type of Exon = Snagl
	Gene number = 01 Exon number = 02 Type of Exon = Term	Gene number = 01 Exon number = 02 Type of Exon = Intra	Gene number = 01 Exon number = 02 Type of Exon = PlyA	Gene number = 01 Exon number = 02 Type of Exon = PlyA	Gene number = 01 Exon number = 02 Type of Exon = PlyA
Type of DNA Strand + = Input Strand - = Output Strand	E-1(+ = Input strand) E-2(+ = Input strand)	E-1(+ = Input strand) E-2(+ = Input strand)	E-1(+ = Input strand) E-2(+ = Input strand)	E-1(+ = Input strand) E-2(+ = Input strand)	E-1(+ = Input Strand) E-2(+ = Input Strand)
Beginning of Exon/Signal	E-1 = 90 E-2 = 1098	E-1 = 90 E-2 = 1046	E-1 = 90 E-2 = 1293	E-1 = 90 E-2 = 1435	E-1 = 90 E-2 = 1274
Ending of Exon/signal	E-1 = 990 E-2 = 1498	E-1 = 1004 E-1 = 1225	E-1 = 1070 E-2 = 1298	E-1 = 1070 E-2 = 1440	E-1 = 1085 E-2 = 1279
Length of Exon/signal	E-1 = 901 E-2 = 401	E-1 = 915 E-2 = 180	E-1 = 981 E-2 = 06	E-1 = 981 E-2 = 06	E-1 = 996 E-2 = 06
Reading Frame	E-1 = 02 E-2 = 01	E-1 = 02 E-2 = 01	E-1 = 02 E-2 = ()	E-1 = 02 E-2 = ()	E-1 = 02 E-2 = ()
Net-Phase of Exon/Signal	E-1 = 01 E-2 = 02	E-1 = 00 E-2 = 00	E-1 = 00 E-2 = ()	E-1 = 00 E-2 = ()	E-1 = 00 E-2 = ()
Initiation signal/3'-Splice site score	E-1 = 81 E-2 = 20	E-1 = 101 E-2 = 72	E-1 = 64 E-2 = ()	E-1 = 48 E-2 = ()	E-1 = 60 E-2 = ()
Termination signal/5'-Splice site score	E-1 = 53 E-2 = 48	E-1 = -6 E-2 = 69	E-1 = 37 E-2 = ()	E-1 = 32 E-2 = ()	E-1 = 28 E-2 = ()
Coding Region score	E-1 = 349 E-2 = 287	E-1 = 427 E-2 = 239	E-1 = 279 E-2 = ()	E-1 = 837 E-2 = ()	E-1 = 692 E-2 = ()
Probability of Exon	E-1 = 0.752 E-2 = 0.989	E-1 = 0.769 E-2 = 0.823	E-1 = 0.632 E-2 = ()	E-1 = 0.840 E-2 = ()	E-1 = 0.728 E-2 = ()
Exon score	E-1 = 25.13 E-2 = 13.18	E-1 = 28.67 E-2 = 20.46	E-1 = 17.22 E-2 = -1.75	E-1 = 70.65 E-2 = -0.45	E-1 = 57.05 E-2 = -1.75

Small Secreted Glycoprotein:

	Tai Forest Ebolavirus	Bundibugyo Ebolavirus	Sudan Ebolavirus	Zaire Ebolavirus	Reston Ebolavirus
Gene number;Exon number and Exon type	Gene number = 01 Exon number = 01 Type of Exon= Init	Gene number = 01 Exon number = 01 Type of Exon = Init	Gene number = 01 Exon number = 01 Type of Exon = Init	Gene number = 01 Exon number = 01 Type of Exon = Intra	Gene number = 01 Exon Number = 01 Type of Exon = Init
	Gene number = 01	Gene number = 01	Gene number = 01	Gene number = 01	Gene number = 01



	Exon number = 02 Type of Exon = Term Gene number = 01 Exon number = 03 Type of Exon = PlyA	Exon number = 02 Type of Exon = Term Gene number = 01 Exon number = 03 Type of Exon = PlyA	Exon number = 02 Type of Exon = Term	Exon number = 02 Type of Exon = Intr	Exon number = 02 Type of Exon = Intr
Type of DNA Strand + = Input Strand - = Output Strand	E-1(+ = Input strand) E-2(+ = Input strand) E-3(+ = Input strand)	E-1(+ = Input Strand) E-2(+ = Input Strand) E-3(+ = Input Strand)	E-1(+ = Input Strand) E-2(+ = Input Strand)	E-1(+ = Input Strand) E-2(+ = Input Strand)	E-1(+ = Input strand) E-2(+ = Input Strand)
Beginning of Exon/Signal	E-1 = 140 E-2 = 1076 E-3 = 2356	E-1 = 140 E-2 = 1052 E-3 = 2373	E-1 = 116 E-2 = 1043	E-1 = 425 E-2 = 1035	E-1 = 142 E-2 = 1102
Ending of Exon/signal	E-1 = 995 E-2 = 2169 E-3 = 2361	E-1 = 995 E-2 = 2169 E-3 = 2378	E-1 = 971 E-2 = 2145	E-1 = 999 E-2 = 2101	E-1 = 1000 E-2 = 1938
Length of Exon/signal	E-1 = 856 E-2 = 1094 E-3 = 06	E-1 = 856 E-2 = 1118 E-3 = 06	E-1 = 856 E-2 = 1103	E-1 = 575 E-2 = 1067	E-1 = 859 E-2 = 837
Reading Frame	E-1 = 01 E-2 = 00 E-3 = ()	E-1 = 01 E-2 = 00 E-3 = ()	E-1 = 01 E-2 = 00	E-1 = 01 E-2 = 00	E-1 = 00 E-2 = 02
Net-Phase of Exon/Signal	E-1 = 01 E-2 = 02 E-3 = ()	E-1 = 01 E-2 = 02 E-3 = ()	E-1 = 01 E-2 = 02	E-1 = 02 E-2 = 02	E-1 = 01 E-2 = 00
Initiation signal/3'-Splice site score	E-1 = 84 E-2 = 35 E-3 = ()	E-1 = 87 E-2 = -37 E-3 = ()	E-1 = 82 E-2 = -50	E-1 = 39 E-2 = -12	E-1 = 87 E-2 = 34
Termination signal/5'-Splice site score	E-1 = 92 E-2 = 43 E-3 = ()	E-1 = 36 E-2 = 33 E-3 = ()	E-1 = -4 E-2 = 42	E-1 = -12 E-2 = 72	E-1 = 97 E-2 = 67
Coding Region score	E-1 = 285 E-2 = 497 E-3 = ()	E-1 = 496 E-2 = 430 E-3 = ()	E-1 = 305 E-2 = 578	E-1 = 423 E-2 = 611	E-1 = 415 E-2 = 425
Probability of Exon	E-1 = 0.896 E-2 = 0.979 E-3 = ()	E-1 = 0.908 E-2 = 0.134 E-3 = ()	E-1 = 0.237 E-2 = 0.309	E-1 = 0.365 E-2 = 0.734	E-1 = 0.835 E-2 = 0.362
Exon score	E-1 = 23.30 E-2 = 32.16 E-3 = 1.05	E-1 = 39.10 E-2 = 17.20 E-3 = 1.05	E-1 = 15.51 E-2 = 31.64	E-1 = 18.94 E-2 = 39.43	E-1 = 37.09 E-2 = 26.36

Second Secreted Glycoprotein:

	Tai Forest Ebolavirus	Bundibugyo Ebolavirus	Sudan Ebolavirus	Zaire Ebolavirus	Reston Ebolavirus
Gene number; Exon number and Exon type	Gene number = 01 Exon number = 01 Type of Exon = init	Gene number = 01 Exon number = 01 Type of Exon = Init	Gene number = 01 Exon number = 01 Type of Exon = Init	Gene number = 01 Exon number = 01 Type of Exon = Intr	Gene number = 01 Exon number = 01 Type of Exon = Init
	Gene number = 01 Exon number = 02 Type of Exon = Term	Gene number = 01 Exon number = 02 Type of Exon = Term	Gene number = 01 Exon number = 02 Type of exon = Term	Gene number = 01 Exon number = 02 Type of exon = Intr	Gene number = 01 Exon number = 02 Type of exon = Intr
	Gene number = 01 Exon number = 03 Type of Exon = PlyA	Gene number = 01 Exon number = 03 Type of Exon = PlyA			
Type of DNA Strand + = Input Strand - = Output Strand	E-1(+ = Input Strand) E-2(+ = Input strand) E-3(+ = Input strand)	E-1(+ = Input Strand) E-2(+ = Input Strand) E-3(+ = Input Strand)	E-1(+ = Input Strand) E-2(+ = Input Strand)	E-1(+ = Input strand) E-2(+ = Input Strand)	E-1(+ = Input strand) E-2(+ = Input strand)
Beginning of Exon/Signal	E-1 = 140 E-2 = 1076 E-3 = 2356	E-1 = 140 E-2 = 1052 E-3 = 2373	E-1 = 116 E-2 = 1043	E-1 = 425 E-2 = 1035	E-1 = 142 E-2 = 1102
Ending of Exon/signal	E-1 = 995 E-2 = 2169 E-3 = 2361	E-1 = 995 E-2 = 2169 E-3 = 2378	E-1 = 971 E-2 = 2145	E-1 = 999 E-2 = 2101	E-1 = 1000 E-2 = 1938
Length of Exon/signal	E-1 = 856 E-2 = 1094 E-3 = 06	E-1 = 856 E-2 = 1118 E-3 = 06	E-1 = 856 E-2 = 1103	E-1 = 575 E-2 = 1067	E-1 = 859 E-2 = 837
Reading Frame	E-1 = 01 E-2 = 00 E-3 = ()	E-1 = 01 E-2 = 00 E-3 = ()	E-1 = 01 E-2 = 00	E-1 = 01 E-2 = 00	E-1 = 00 E-2 = 02



Net-Phase of Exon/Signal	E-1 = 01 E-2 = 02 E-3 = ()	E-1 = 01 E-2 = 02 E-3 = ()	E-1 = 01 E-2 = 02	E-1 = 02 E-2 = 02	E-1 = 01 E-2 = 00
Initiation signal/3'-Splice site score	E-1 = 84 E-2 = 35 E-3 = ()	E-1 = 87 E-2 = -37 E-3 = ()	E-1 = 82 E-2 = -50	E-1 = 39 E-2 = -12	E-1 = 87 E-2 = 34
Termination signal/5'-Splice site score	E-1 = 92 E-2 = 43 E-3 = ()	E-1 = 36 E-2 = 33 E-3 = ()	E-1 = -4 E-2 = 42	E-1 = -12 E-2 = 72	E-1 = 97 E-2 = 67
Coding Region score	E-1 = 285 E-2 = 497 E-3 = ()	E-1 = 496 E-2 = 430 E-3 = ()	E-1 = 305 E-2 = 578	E-1 = 423 E-2 = 611	E-1 = 415 E-2 = 425
Probability of Exon	E-1 = 0.896 E-2 = 0.979 E-3 = ()	E-1 = 0.908 E-2 = 0.134 E-3 = ()	E-1 = 0.237 E-2 = 0.309	E-1 = 0.365 E-2 = 0.734	E-1 = 0.835 E-2 = 0.362
Exon score	E-1 = 23.30 E-2 = 32.16 E-3 = 1.05	E-1 = 39.10 E-2 = 17.20 E-3 = 1.05	E-1 = 15.51 E-2 = 31.64	E-1 = 18.94 E-2 = 39.43	E-1 = 37.09 E-2 = 26.36

Spike Glycoprotein:

	Tai Forest Ebolavirus	Bundibugyo Ebolavirus	Sudan Ebolavirus	Zaire Ebolavirus	Reston Ebolavirus
Gene number; Exon number and Exon type	Gene number = 01 Exon number = 01 Type of Exon = Init	Gene number = 01 Exon number = 01 Type of Exon = Init	Gene number = 01 Exon number = 01 Type of Exon = Init	Gene number = 01 Exon number = 01 Type of Exon = Intr	Gene number = 01 Exon number = 01 Type of Exon = Init
	Gene number = 01 Exon number = 02 Type of Exon = Term	Gene number = 01 Exon number = 02 Type of Exon = Term	Gene number = 01 Exon number = 02 Type of Exon = Term	Gene number = 01 Exon number = 02 Type of Exon = Intr	Gene number = 01 Exon number = 02 Type of Exon = Intr
	Gene number = 01 Exon number = 02 Type of Exon = PlyA	Gene number = 01 Exon number = 03 Type of Exon = PlyA			
Type of DNA Strand + = Input Strand - = Output Strand	E-1(+ = Input Strand) E-2(+ = Input Strand) E-3(+ = Input Strand)	E-1(+ = Input Strand) E-2(+ = Input Strand) E-3(+ = Input Strand)	E-1(+ = Input Strand) E-2(+ = Input Strand)	E-1(+ = Input Strand) E-2(+ = Input Strand)	E-1(+ = Input Strand) E-2(+ = Input Strand)
Beginning of Exon/Signal	E-1 = 140 E-2 = 1076 E-3 = 2356	E-1 = 140 E-2 = 1052 E-3 = 2373	E-1 = 116 E-2 = 1043	E-1 = 425 E-2 = 1035	E-1 = 142 E-2 = 1102
Ending of Exon/signal	E-1 = 995 E-2 = 2169 E-3 = 2361	E-1 = 995 E-2 = 2169 E-3 = 2378	E-1 = 971 E-2 = 2145	E-1 = 999 E-2 = 2101	E-1 = 1000 E-2 = 1938
Length of Exon/signal	E-1 = 856 E-2 = 1094 E-3 = 06	E-1 = 856 E-2 = 1118 E-3 = 06	E-1 = 856 E-2 = 1103	E-1 = 575 E-2 = 1067	E-1 = 859 E-2 = 837
Reading Frame	E-1 = 01 E-2 = 00 E-3 = ()	E-1 = 01 E-2 = 00 E-3 = ()	E-1 = 01 E-2 = 00	E-1 = 01 E-2 = 00	E-1 = 00 E-2 = 02
Net-Phase of Exon/Signal	E-1 = 01 E-2 = 02 E-3 = ()	E-1 = 01 E-2 = 02 E-3 = ()	E-1 = 01 E-2 = 02	E-1 = 02 E-2 = 02	E-1 = 01 E-2 = 00
Initiation signal/3'-Splice site score	E-1 = 84 E-2 = 35 E-3 = ()	E-1 = 87 E-2 = -37 E-3 = ()	E-1 = 82 E-2 = -50	E-1 = 39 E-2 = -12	E-1 = 87 E-2 = 34
Termination signal/5'-Splice site score	E-1 = 92 E-2 = 43 E-3 = ()	E-1 = 36 E-2 = 33 E-3 = ()	E-1 = -4 E-2 = 42	E-1 = -12 E-2 = 72	E-1 = 97 E-2 = 67



Coding Region score	E-1 = 285 E-2 = 497 E-3 = ()	E-1 = 496 E-2 = 430 E-3 = ()	E-1 = 305 E-2 = 578	E-1 = 423 E-2 = 611	E-1 = 415 E-2 = 425
Probability of Exon	E-1 = 0.896 E-2 = 0.979 E-3 = ()	E-1 = 0.908 E-2 = 0.134 E-3 = ()	E-1 = 0.237 E-2 = 0.309	E-1 = 0.365 E-2 = 0.734	E-1 = 0.835 E-2 = 0.362
Exon score	E-1 = 23.30 E-2 = 32.16 E-3 = 1.05	E-1 = 39.10 E-2 = 17.20 E-3 = 1.05	E-1 = 15.51 E-2 = 31.64	E-1 = 18.94 E-2 = 39.43	E-1 = 37.09 E-2 = 26.36

RNA Dependent RNA Polymerase:

	Tai Forest Ebolavirus	Bundibugyo Ebolavirus	Sudan Ebolavirus	Zaire Ebolavirus	Reston Ebolavirus
Gene number; Exon number and Exon type	Gene number = 01 Exon number = 01 Type of Exon = Sngl	Gene number = 01 Exon number = 01 Type of Exon = Sngl	Gene number = 01 Exon number = 01 Type of Exon = Sngl	Gene number = 01 Exon number = 01 Type of Exon = Sngl	Gene number = 01 Exon number = 01 Type of Exon = Sngl
	Gene number = 01 Exon number = 02 Type of Exon = PlyA	Gene number = 01 Exon number = 02 Type of Exon = PlyA	Gene number = 01 Exon number = 02 Type of Exon = PlyA	Gene number = 01 Exon number = 02 Type of Exon = PlyA	Gene number = 01 Exon number = 02 Type of Exon = PlyA
Type of DNA Strand + = Input Strand - = Output Strand	E-1(+ = Input Strand) E-2(+ = Input Strand)	E-1(+ = Input Strand) E-2(+ = Input Strand)	E-1(+ = Input Strand) E-2(+ = Input Strand)	E-1(+ = Input Strand) E-2(+ = Input Strand)	E-1(+ = Input Strand) E-2(+ = Input Strand)
Beginning of Exon/Signal	E-1 = 81 E-2 = 6959	E-1 = 81 E-2 = 6803	E-1 = 82 E-2 = 6986	E-1 = 81 E-2 = 6775	E-1 = 87 E-2 = 7202
Ending of Exon/signal	E-1 = 6713 E-2 = 6964	E-1 = 6713 E-2 = 6808	E-1 = 6711 E-2 = 6991	E-1 = 6719 E-2 = 6780	E-1 = 6725 E-2 = 7207
Length of Exon/signal	E-1 = 6633 E-2 = 06	E-1 = 6633 E-2 = 06	E-1 = 6630 E-2 = 06	E-1 = 6639 E-2 = 06	E-1 = 6639 E-2 = 06
Reading Frame	E-1 = 02 E-2 = ()	E-1 = 02 E-2 = ()	E-1 = 00 E-2 = ()	E-1 = 02 E-2 = 00	E-1 = 02 E-2 = ()
Net-Phase of Exon/Signal	E-1 = 00 E-2 = ()	E-1 = 00 E-2 = ()	E-1 = 00 E-2 = ()	E-1 = 00 E-2 = 00	E-1 = 00 E-2 = ()
Initiation signal/3'-Splice site score	E-1 = 87 E-2 = ()	E-1 = 36 E-2 = ()	E-1 = 52 E-2 = ()	E-1 = 15 E-2 = ()	E-1 = 40 E-2 = ()
Termination signal/5'-Splice site score	E-1 = 49 E-2 = ()	E-1 = 44 E-2 = ()	E-1 = 31 E-2 = ()	E-1 = 38 E-2 = ()	E-1 = 43 E-2 = ()
Coding Region score	E-1 = 2393 E-2 = ()	E-1 = 2459 E-2 = ()	E-1 = 2403 E-2 = ()	E-1 = 2927 E-2 = ()	E-1 = 2884 E-2 = ()
Probability of Exon	E-1 = 0.884 E-2 = ()	E-1 = 0.996 E-2 = ()	E-1 = 0.967 E-2 = ()	E-1 = 0.972 E-2 = ()	E-1 = 0.975 E-2 = ()
Exon score	E-1 = 227.53 E-2 = -3.44	E-1 = 228.53 E-2 = -0.45	E-1 = 223.23 E-2 = -1.75	E-1 = 272.63 E-2 = -3.44	E-1 = 271.33 E-2 = 1.05

Membrane Associated Protein:

	Tai Forest Ebolavirus	Bundibugyo Ebolavirus	Sudan Ebolavirus	Zaire Ebolavirus	Reston Ebolavirus
Gene number; Exon number and Exon type	Gene number = 01 Exon number = 01 Type of Exon = Term	Gene number = 01 Exon number = 01 Type of Exon = Sngl	Gene number = 01 Exon number = 01 Type of Exon = Sngl	Gene number = 01 Exon number = 01 Type of Exon = Sngl	Gene number = 01 Exon number = 01 Type of Exon = Sngl
	Gene number = 01 Exon number = 02 Type of Exon = PlyA	Gene number = 01 Exon number = 02 Type of Exon = PlyA	Gene number = 01 Exon number = 02 Type of Exon = PlyA	Gene number = 01 Exon number = 02 Type of Exon = PlyA	Gene number = 01 Exon number = 02 Type of Exon = PlyA
Type of DNA Strand	E-1(+ = Input Strand)	E-1(+ = Input Strand)	E-1(+ = Input Strand)	E-1(+ = Input Strand)	E-1(+ = Input Strand)



+ = Input Strand - = Output Strand	E-2(+ = Input Strand)	E-2(+ = Input Strand)	E-2(+ = Input Strand)	E-2(+ = Input Strand)	E-2(+ = Input Strand)
Beginning of Exon/Signal	E-1 = 327 E-2 = 1439	E-1 = 467 E-2 = 1348	E-1 = 474 E-2 = 1237	E-1 = 461 E-2 = 1310	E-1 = 472 E-2 = 1241
Ending of Exon/signal	E-1 = 1222 E-2 = 1444	E-1 = 1222 E-2 = 1353	E-1 = 1229 E-2 = 1242	E-1 = 1216 E-2 = 1315	E-1 = 1227 E-2 = 1246
Length of Exon/signal	E-1 = 896 E-2 = 06	E-1 = 756 E-2 = 06	E-1 = 756 E-2 = 06	E-1 = 756 E-2 = 06	E-1 = 756 E-2 = 06
Reading Frame	E-1 = 01 E-2 = ()	E-1 = 01 E-2 = ()	E-1 = 02 E-2 = ()	E-1 = 01 E-2 = ()	E-1 = 00 E-2 = ()
Net-Phase of Exon/Signal	E-1 = 02 E-2 = ()	E-1 = 00 E-2 = ()	E-1 = 00 E-2 = ()	E-1 = 00 E-2 = ()	E-1 = 00 E-2 = ()
Initiation signal/3'-Splice site score	E-1 = 45 E-2 = ()	E-1 = 80 E-2 = ()	E-1 = 80 E-2 = ()	E-1 = 88 E-2 = ()	E-1 = 83 E-2 = ()
Termination signal/5'-Splice site score	E-1 = 35 E-2 = ()	E-1 = 37 E-2 = ()	E-1 = 32 E-2 = ()	E-1 = 36 E-2 = ()	E-1 = 42 E-2 = ()
Coding Region score	E-1 = 535 E-2 = ()	E-1 = 450 E-2 = ()	E-1 = 504 E-2 = ()	E-1 = 622 E-2 = ()	E-1 = 299 E-2 = ()
Probability of Exon	E-1 = 0.795 E-2 = ()	E-1 = 0.776 E-2 = ()	E-1 = 0.992 E-2 = ()	E-1 = 0.999 E-2 = ()	E-1 = 0.922 E-2 = ()
Exon score	E-1 = 35.47 E-2 = -1.75	E-1 = 34.99 E-2 = 1.05	E-1 = 39.89 E-2 = 1.05	E-1 = 52.89 E-2 = 1.05	E-1 = 20.69 E-2 = -3.24

Minor Nucleoprotein:

	Tai Forest Ebolavirus	Bundibugyo Ebolavirus	Sudan Ebolavirus	Zaire Ebolavirus	Reston Ebolavirus
Gene number; Exon number and Exon type	Gene number = 01 Exon number = 01 Type of Exon = Sngl	Gene number = 01 Exon number = 01 Type of Exon = Sngl	Gene number = 01 Exon number = 01 Type of Exon = Sngl	Gene number = 01 Exon number = 01 Type of Exon = Sngl	Gene number = 01 Exon number = 01 Type of Exon = Sngl
	Gene number = 01 Exon number = 02 Type of Exon = PlyA	Gene number = 01 Exon number = 02 Type of Exon = PlyA	Gene number = 01 Exon number = 02 Type of Exon = PlyA	Gene number = 01 Exon number = 02 Type of Exon = PlyA	Gene number = 01 Exon number = 02 Type of Exon = PlyA
Type of DNA Strand + = Input Strand - = Output Strand	E-1(+ = Input Strand) E-2(+ = Input Strand)	E-1(+ = Input Strand) E-2(+ = Input Strand)	E-1(+ = Input Strand) E-2(+ = Input Strand)	E-1(+ = Input Strand) E-2(+ = Input Strand)	E-1(+ = Input Strand) E-2(+ = Input Strand)
Beginning of Exon/Signal	E-1 = 228 E-2 = 1174	E-1 = 228 E-2 = 1174	E-1 = 218 E-2 = 1161	E-1 = 222 E-2 = 1401	E-1 = 229 E-2 = 1132
Ending of Exon/signal	E-1 = 1097 E-2 = 1179	E-1 = 1097 E-2 = 1179	E-1 = 1084 E-2 = 1166	E-1 = 1088 E-2 = 1406	E-1 = 1092 E-2 = 1137
Length of Exon/signal	E-1 = 870 E-1 = 06	E-1 = 870 E-2 = 06	E-1 = 867 E-2 = 06	E-1 = 867 E-2 = 06	E-1 = 864 E-2 = 06
Reading Frame	E-1 = 02 E-2 = ()	E-1 = 02 E-2 = ()	E-1 = 01 E-2 = ()	E-1 = 02 E-2 = ()	E-1 = 00 E-2 = ()
Net-Phase of Exon/Signal	E-1 = 00 E-2 = ()	E-1 = 00 E-2 = ()	E-1 = 00 E-2 = ()	E-1 = 00 E-2 = ()	E-1 = 00 E-2 = ()
Initiation signal/3'-Splice site score	E-1 = 78 E-2 = ()	E-1 = 52 E-2 = ()	E-1 = 73 E-2 = ()	E-1 = 06 E-2 = ()	E-1 = 77 E-2 = ()
Termination signal/5'-Splice site score	E-1 = 48 E-2 = ()	E-1 = 38 E-2 = ()	E-1 = 38 E-2 = ()	E-1 = 38 E-2 = ()	E-1 = 43 E-2 = ()
Coding Region score	E-1 = 595 E-2 = ()	E-1 = 540 E-2 = ()	E-1 = 513 E-2 = ()	E-1 = 414 E-2 = ()	E-1 = 566 E-2 = ()
Probability of Exon	E-1 = 0.990 E-2 = ()	E-1 = 0.865 E-2 = ()	E-1 = 0.971 E-2 = ()	E-1 = 0.891 E-2 = ()	E-1 = 0.693 E-2 = ()
Exon score	E-1 = 50.17 E-2 = 1.05	E-1 = 41.07 E-2 = -1.75	E-1 = 40.44 E-2 = 1.05	E-1 = 23.84 E-2 = -0.45	E-1 = 46.62 E-2 = 1.05



Interpretation Of Data- The above data provides Gene structural information of the genes encoded by Ebola virus proteins. The data represented in the above table shows different terminologies involving gene structure i.e. Exon Number, Type of Exon [Init = Initial exon (ATG to 5' splice site); Intr = Internal exon (3' splice site to 5' splice site); Term = Terminal exon (3' splice site to stop codon); Sngl = Single exon gene (ATG to stop); Prom = Promoter (TATA box/ initiation site); Ply A = poly A signal (consensus: AATAAA)]. We also study DNA strand i.e. (+) = input strand and (-) = negative strand. We also studied beginning of the exon / signal, end of the exon/signal, Length of exon, Reading frame, Net phase of exon, Initial signal/3' splice site score, Termination signal/5' splice site score, Coding region score, probability of exon along with exon score.

X. CONCLUSION

This three module study help us to know different angles i.e; the Extinction coefficient estimations of proteins help us to distinguish the adjustment in light collecting proficiency and surface inclusion esteem with submersion solvent, Immersion time and drenching focus in the protein. It is likewise used to ponder the effect of dissolvable to control the adsorption kinetics. This is utilized to characterize the scope of wavelength where the light has its greatest profundity of infiltration in tissue. This is the inherent property of various species so it is utilized to separate between the molecules. Instability record clarifies the steady property of protein. Aliphatic file estimates the dissolvability of focused proteins. We even assessed whether the protein is hydrophobic/hydrophilic dependent on the amazing normal of hydrophaticity values. we can anticipate the protein structure, Function and developmental history of groupings and its utilization structure superposition programs and phylogenetic examination programmes. By the quality basic data we can discover infection seriousness and foresee quality structure to explore function, expression level, disease, mutation. By this quality basic investigation we can avoid the sickness by postponement of occurrence of ailment.

XI. REFERENCES

1. Multiple sequence alignment: a methodology for protein identification 2015 (Srinivasa Rao V.*1, Das S. K.2, Nageswara Rao K.3 and Kusuma Kumari E.4)
2. Protein Identification and Analysis Tools on the ExPASy Server 1998 (Elisabeth Gasteiger, Christine Hoogland, Alexandre Gattiker, Séverine Duvaud, Marc R. Wilkins, Ron D. Appel, and Amos Bairoch)
3. Ebola virus disease by CDC/Dr F. A. Murphy 2003 (World health organisation-western pacific region)
4. Genome sequence analysis of Ebola virus in clinical samples from three british healthcare workers, August 2014 to March 2015 (A Bell, K Lewandowski, R Myers, D Woolridge, E Aarons, A simpson, R Vipond, M Jacobs, S Gharbia, M Zambon.

5. IN SILICO approach of structure prediction and funtional characterization of Zaire Ebola and identification of binding site for drug development 2016 (Md. Jahirul Islam, Kaniz Fatima, Pipasha Biswas-2008 Department of Biochemistry and Biotechnology; University Of Science And Technology, Chittagong)

6. Advance protein alingments based on sequence, structure and hydropathy profiles 2018 The paradigm of viral polymerase enzyme (Alexandros Armaos , Dimitrios Vlachakis, Sophia Kossida)

7. Extinction coefficient- (2007)A guide to understanding extinction coefficient, with emphasis spectrophotometric determination of protein concentration

8. Ebola virus comparative genomics (1999)(Se-Ran Jun; Michael R.Leuze; and David W. Ussery)

9. Ebolaviruses New Roles For Old Proteins (2013)(Diegocantoni, Jeremy S. Rossman-School of Biosciences, University Of Kent, Canterbury,UK) 10. Ebola Virus: (2012) Identification of Virion Structural Proteins By Michael P. Kiley; Russell L. Ragnery and Karl M. Johnson (Special pathogens branch, Virology Divison , Bureau Of Laboratories, Centre for disease control , Public Health Service , US Health , Education and Welfare, Atlanta, Georgia 30333) 1. 11. Bermejo M, Rodriguez-Teijeiro JD, Illera G, Barroso A, Vilà C, Walsh PD. Ebola outbreak killed 5000 gorillas. Science 2006;314:1564. 2.

12. Chepurinov AA, Bakulina LF, Dadaeva AA, Ustinova EN, Chepurnova TS, Baker JR Jr. Inactivation of Ebola virus

IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

ABOUT IJEAST

International Journal of Engineering Applied Science and Technology (IJEAST) is a peer-reviewed, open access journal that publishes high-quality research papers in the field of Engineering, Applied Science and Technology.

IJEAST aims to provide a platform for researchers, academicians, and professionals to share their innovative ideas, research findings, and practical experiences with the global scientific community.

FOCUS AREAS

- Engineering
- Applied Science
- Technology
- Innovation & Development
- Interdisciplinary Studies



PEER REVIEWED

All submissions are rigorously peer reviewed to ensure quality.



OPEN ACCESS

Free and unrestricted access to research for all.



GLOBAL REACH

Connecting researchers and professionals worldwide.



TIMELY PUBLICATION

We ensure a swift and efficient publication process.



For more information, visit our website

www.ijeast.com



INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

✉ editor@ijeast.com

🌐 www.ijeast.com

📍 India



2455-2143