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 casualty 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 hypothesis, 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 inactive 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 cysteine. Formula for calculating Extinction coefficient is given below.

\[ E(Prot) = \text{Numb(Tyr)} \times \text{Ext(Tyr)} + \text{Numb(Trp)} \times \text{Ext(Trp)} + \text{Numb(Cystine)} \times \text{Ext(Cystine)} \]

With:
- \( E(Prot) \) = Extinction coefficient of protein
- \( \text{Numb(Tyr)} \) = Number of tyrosine
- \( \text{Ext(Tyr)} \) = Extinction coefficient of tyrosine
- \( \text{Numb(Trp)} \) = Number of tryptophan
- \( \text{Ext(Trp)} \) = Extinction coefficient of tryptophan
- \( \text{Numb(Cystine)} \) = Number of cysteine
- \( \text{Ext(Cystine)} \) = Extinction coefficient of cysteine
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.

\[
\text{Aliphatic index} = X(\text{Ala}) + a \times X(\text{Val}) + b \times (X(\text{Ile}) + X(\text{Leu}))
\]

Where,
- \(X(\text{Ala})\) = Mole percent of Alanine
- \(X(\text{Val})\) = Mole percent of valine
- \(X(\text{Ile})\) = Mole percent of Isoleucine
- \(X(\text{Leu})\) = Mole percent of leucine

V. GRAND AVERAGE OF HYDROPHATICITY

The Grand Average of hydropathy (GRAVY) esteem for a peptide or protein is determined as the entirety of hydropathy 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

```r
# @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:
QWGRCCGWGPGRYRCVRWC
# # The instability index (II) is computed to be 83.68
#
# instaIndex(seq "QWGRCCGWGPGRYRCVRWC")
# [1] 83.68
```
MS = 44.94, FG = 1,
MG = 1, FA = 1,
MA = 13.34, FL = 1,
ML = 1, QW = 1,
HW = -1.88, QC = -6.54,
HC = 1, QM = 1,
HM = 1, QH = 1,
HH = 1, QY = -6.54,
HY = 44.94, QF = -6.54,
HF = -9.37, QQ = 20.26,
HQ = 1, QN = 1,
HN = 24.68, QI = 1,
HI = 44.94, QR = 1,
HR = 1, QD = 20.26,
HD = 1, QT = 1,
HP = -1.88, QK = 1,
HT = -6.54, QE = 20.26,
HK = 24.68, QV = -6.54,
HE = 1, QS = 44.94,
HV = 1, QG = 1,
HS = 1, QA = 1,
HG = -9.37, QL = 1,
HA = 1, NW = -9.37,
HL = 1, NC = -1.88,
YW = -9.37, NM = 1,
YC = 1, NH = 1,
YM = 44.94, NY = 1,
YH = 13.34, NF = -14.03,
YY = 13.34, NQ = -6.54,
YF = 1, NN = 1,
YQ = 1, NI = 44.94,
YN = 1, NR = 1,
YI = 1, ND = 1,
YR = -15.91, NP = -1.88,
YD = 24.68, NT = -7.49,
YP = 13.34, NK = 24.68,
YT = -7.49, NE = 1,
YK = 1, NV = 1,
YE = -6.54, NS = 1,
YV = 1, NG = -14.03,
YS = 1, NL = 1,
YG = -7.49, IW = 1,
YA = 24.68, IC = 1,
YL = 1, IM = 1,
FW = 1, IH = 13.34,
FC = 1, JY = 1,
FM = 1, IF = 1,
FH = 1, IQ = 1,
FY = 33.6, IN = 1,
FF = 1, II = 1,
FQ = 1, IR = 1,
FN = 1, ID = 1,
FI = 1, IP = -1.88,
FR = 1, IT = 1,
FD = 13.34, IK = -7.49,
FP = 20.26, IE = -44.94,
FT = 1, IV = -7.49,
FK = -14.03, IS = 1,
FE = 1, IG = 1,
FV = 1, IA = 1,
FS = 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,  
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,
International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 4, Issue 9, ISSN No. 2455-2143, Pages 241-271
Published Online January 2020 in IJEAST (http://www.ijeast.com)

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,
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 builds ups 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.

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
furan 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 builds ups 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. Not withstanding 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 nuclease,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 properties of Ebolavirus along with Gene structural
information and sequence homology to interpret significant aspects on
E bola 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

Published Online January 2020 in IJEAST (http://www.ijeast.com)
differs from other viral hemorrhagic fevers in its 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.
| Number and composition of amino acids | Bundibugyo Ebola virus | Sudan Ebola virus | Reston Ebola virus | Zaire Ebola virus | Tai forest ebola virus |
|--------------------------------------|------------------------|------------------|-------------------|------------------|-----------------------|
| 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%)         | 30(4.3%)          | 33(4.5%)         | 29(3.9%)              |
| Asparagine (Asn)                     | 46(6.3%)               | 44(5.9%)         | 40(5.4%)          | 35(4.8%)         | 43(6.1%)              |
| Aspartic acid (Asp)                  | 57(7.7%)               | 59(8.0%)         | 59(8.0%)          | 59(8.0%)         | 48(6.5%)              |
| Cysteine (Cys)                       | 50(6.9%)               | 50(6.9%)         | 50(6.9%)          | 50(6.9%)         | 50(6.9%)              |
| Glutamine (Gln)                      | 51(7.0%)               | 70(9.3%)         | 62(8.7%)          | 68(9.6%)         | 49(6.9%)              |
| Glutamic acid (Glu)                  | 56(7.6%)               | 58(7.8%)         | 59(8.0%)          | 59(8.0%)         | 62(8.5%)              |
| Glycine (Gly)                        | 37(5.0%)               | 53(7.2%)         | 42(5.7%)          | 41(5.9%)         | 37(5.0%)              |
| Histidine (His)                      | 25(3.4%)               | 26(3.4%)         | 28(3.8%)          | 30(4.1%)         | 36(4.1%)              |
| Isoleucine (Ile)                     | 29(3.9%)               | 31(4.3%)         | 31(4.3%)          | 29(3.9%)         | 31(4.3%)              |
| Leucine (Leu)                        | 62(8.4%)               | 57(10.9%)        | 54(10.9%)         | 67(13.1%)        | 64(7.9%)              |
| Lysine (Lys)                         | 37(5.0%)               | 56(8.4%)         | 31(5.2%)          | 29(4.8%)         | 41(5.9%)              |
| Methionine (Met)                     | 20(2.8%)               | 13(1.8%)         | 15(2.0%)          | 20(2.5%)         | 17(2.3%)              |
| Phenylalanine (Phe)                  | 25(3.4%)               | 24(3.5%)         | 26(3.4%)          | 26(3.5%)         | 26(3.5%)              |
| Proline (Pro)                        | 40(5.5%)               | 40(5.5%)         | 40(5.5%)          | 42(5.7%)         | 37(5.0%)              |
| Serine (Ser)                         | 47(6.4%)               | 49(6.6%)         | 41(6.0%)          | 43(5.8%)         | 51(6.0%)              |
| Threonine (Thr)                      | 41(5.5%)               | 39(5.4%)         | 32(4.9%)          | 38(5.1%)         | 49(6.9%)              |
| Tryptophan (Trp)                     | 0.09(0.5%)             | 0.09(0.9%)       | 0.09(0.9%)        | 0.09(0.9%)       | 0.09(0.9%)            |
| Tyrosine (Tyr)                       | 22(3.0%)               | 22(3.0%)         | 22(3.0%)          | 21(2.8%)         | 21(2.8%)              |
| Value (Val)                          | 37(5.0%)               | 45(6.1%)         | 34(4.8%)          | 42(5.7%)         | 36(4.9%)              |

Molecular weight: 82956.34, 81804.90, 81482.62, 83286.68, 82905.24, 81804.90, 83286.68, 83308.50
Theoretical pI: 5.90, 5.73, 5.73, 5.90, 5.33, 5.33
Atomic composition: 11487, 11365, 11587, 11587, 11587

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 same number and composition of the aminocids on the whole but varies when compared with the individual aminocids. When we calculated average isotope mass of the aminocids in the protein and one watermolecule 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 Ebolavirus with the smaller molecular weight.

Compute pI/Mw algorithmis mainly used to enhance a region in a 2-D gel to which an 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 aminocid 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 with out assuming cysteine residues with different absorbance. Thenucleo-protein of sudan ebola virus is more stable when compared with the other forms of the virus species with value 38.60 as the value is less than 40 and Zaire Ebola virus is more unstable as it exceeds 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 Ebolavirus) > 77.93(Reston Ebolavirus) > 75.82 (Bundibugyo Ebolavirus) > 74.32 (Zaire Ebolavirus) > 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 Ebolavirus(-0.565) is greater and for the Tai forest Ebolavirus(-0.714)is less.
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 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. 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) > 85.50 (Reston Ebolavirus) > 86.39 (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.
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 > 35452.31 > 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 aminoacid 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.
Table 4: The second secreted glycol-protein shows approximation in the total number and the composition of the aminoacid and difference when compared individual aminoacids. 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 > 3339187. 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 an 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 aminoacid contributions are studied without 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.
Table-5: The Small Secreted Glyco-Protein although shows approximation in total number and composition of aminocids, but varies when compared to individual aminocids. In the protein and one water molecule we calculated the average isotope mass, the total molecular weight of the protein the order is obtained in this way: 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 all 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 aminocids are studied without considering the secondary and tertiary structure, we observed different values of the extinction coefficient along with their absorbance values both with assuming and without 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: 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.
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 glyco-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 and one water molecule the total molecular weight of the protein is estimated to be in this order: 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 allow 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 without considering the secondary and tertiary structures, we observed different values of extinction coefficient with and without 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.

| Number and composition of amino acids | Bundibugyo Ebolavirus | Sudan Ebolavirus | Reston Ebolavirus | Zaire Ebolavirus | Tai forest Ebolavirus |
|--------------------------------------|-----------------------|-----------------|-------------------|-----------------|----------------------|
| Alanine (Ala)                        | 4(6.0%)               | 4(6.0%)         | 4(6.0%)           | 4(6.0%)         | 4(6.0%)              |
| Arginine (Arg)                       | 35(4.2%)              | 35(4.2%)        | 35(4.2%)          | 35(4.2%)        | 35(4.2%)             |
| Aspartic acid (Asp)                  | 45(6.4%)              | 45(6.4%)        | 45(6.4%)          | 45(6.4%)        | 45(6.4%)             |
| Cysteine (Cys)                       | 12(1.6%)              | 12(1.6%)        | 12(1.6%)          | 12(1.6%)        | 12(1.6%)             |
| Glutamine (Gln)                      | 26(3.5%)              | 26(3.5%)        | 26(3.5%)          | 26(3.5%)        | 26(3.5%)             |
| Glutamic acid (Glu)                  | 40(5.5%)              | 40(5.5%)        | 40(5.5%)          | 40(5.5%)        | 40(5.5%)             |
| Histidine (His)                      | 15(2.0%)              | 15(2.0%)        | 15(2.0%)          | 15(2.0%)        | 15(2.0%)             |
| Ile-Leucine (Ile)                    | 29(3.9%)              | 29(3.9%)        | 29(3.9%)          | 29(3.9%)        | 29(3.9%)             |
| Lysine (Lys)                         | 26(3.5%)              | 26(3.5%)        | 26(3.5%)          | 26(3.5%)        | 26(3.5%)             |
| Methionine (Met)                     | 60(8.0%)              | 60(8.0%)        | 60(8.0%)          | 60(8.0%)        | 60(8.0%)             |
| Phenylalanine (Phe)                  | 24(3.3%)              | 24(3.3%)        | 24(3.3%)          | 24(3.3%)        | 24(3.3%)             |
| Proline (Pro)                        | 50(6.7%)              | 50(6.7%)        | 50(6.7%)          | 50(6.7%)        | 50(6.7%)             |
| Serine (Ser)                         | 56(7.5%)              | 56(7.5%)        | 56(7.5%)          | 56(7.5%)        | 56(7.5%)             |
| Threonine (Thr)                      | 11(1.5%)              | 11(1.5%)        | 11(1.5%)          | 11(1.5%)        | 11(1.5%)             |
| Tryptophan (Trp)                     | 40(5.4%)              | 40(5.4%)        | 40(5.4%)          | 40(5.4%)        | 40(5.4%)             |
| Tyrosine (Tyr)                       | 42(5.6%)              | 42(5.6%)        | 42(5.6%)          | 42(5.6%)        | 42(5.6%)             |
| Value (Val)                          | 48(6.7%)              | 48(6.7%)        | 48(6.7%)          | 48(6.7%)        | 48(6.7%)             |
| Molecular weight                     | 75689.18              | 7467.43         | 74594.18          | 74464.46        | 74416.73             |
| Theoretical pI                       | 6.41                  | 5.97            | 5.99              | 5.6             | 5.10                 |
| Atomic composition                   | 10165                 | 11434           | 10365             | 10375           | 10365                |
| Total number of positively charged residues | 81                  | 60              | 56                | 65              | 39                   |
| Total number of Negatively charged residues | 72              | 69              | 65                | 71              | 87                   |
| Absorbance assuming cysteine residues | 1.253                | 1.253           | 1.253             | 1.253           | 1.253                |
| Extinction coefficient assuming cysteine residues | 93250M cm^-1         | 93250M cm^-1   | 93250M cm^-1     | 93250M cm^-1   | 93250M cm^-1         |
| Extinction coefficient with out cysteine residues | 91000M cm^-1         | 91000M cm^-1   | 91020M cm^-1     | 91000M cm^-1   | 94800M cm^-1         |
| Absorbance with out cysteine residues | 1.335                | 1.335           | 1.335             | 1.335           | 1.335                |
| Incubation Index                     | 38.53                 | 42.10           | 44.28             | 40.38           | 37.21                |
| Aliphatic Index                      | 54.69                 | 50.68           | 55.92             | 55.77           | 57.46                |
| Grand Average of Hydrophobicity      | 0.466                 | 0.352           | 0.404             | 0.330           | 0.320                |
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 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 without considering the secondary and tertiary structures, we observed similar values of extinction coefficient and absorbance in all species of Ebola virus with and without 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) > 10.50 (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.
Table 1: The above table describes different physico-chemical properties associated with the minor nucleo-protein of Ebola virus species, in which all forms of minor nucleo 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; 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 algorithms mainly used to enhance a region in a 2-D gel to which an 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 without considering the secondary and tertiary structures, we observed different values of extinction coefficient and absorbance in all species of Ebola virus with and without assuming cysteine residues. The Minor Nucleo-protein 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

| Amino Acid | Bundibugyo Ebolavirus | Sudan Ebolavirus | Reston Ebolavirus | Zaire Ebolavirus | Taiforest Ebolavirus |
|------------|-----------------------|-----------------|------------------|----------------|---------------------|
| Alanine    | 14(4.9%)              | 16(6.9%)        | 16(6.8%)         | 16(6.9%)       | 16(5.9%)           |
| Arginine   | 46(12%)               | 46(12%)         | 46(12%)          | 46(12%)        | 46(12%)            |
| Aspartic acid | 46(12%)            | 46(12%)         | 46(12%)          | 46(12%)        | 46(12%)            |
| Cysteine   | 46(12%)               | 46(12%)         | 46(12%)          | 46(12%)        | 46(12%)            |
| Glutamic acid | 46(12%)            | 46(12%)         | 46(12%)          | 46(12%)        | 46(12%)            |
| Glycine    | 46(12%)               | 46(12%)         | 46(12%)          | 46(12%)        | 46(12%)            |
| Histidine  | 46(12%)               | 46(12%)         | 46(12%)          | 46(12%)        | 46(12%)            |
| Isoleucine | 46(12%)               | 46(12%)         | 46(12%)          | 46(12%)        | 46(12%)            |
| Leucine    | 46(12%)               | 46(12%)         | 46(12%)          | 46(12%)        | 46(12%)            |
| Lysine     | 46(12%)               | 46(12%)         | 46(12%)          | 46(12%)        | 46(12%)            |
| Methionine | 46(12%)               | 46(12%)         | 46(12%)          | 46(12%)        | 46(12%)            |
| Phenylalanine | 46(12%)        | 46(12%)         | 46(12%)          | 46(12%)        | 46(12%)            |
| Proline    | 46(12%)               | 46(12%)         | 46(12%)          | 46(12%)        | 46(12%)            |
| Serine     | 46(12%)               | 46(12%)         | 46(12%)          | 46(12%)        | 46(12%)            |
| Threonine  | 46(12%)               | 46(12%)         | 46(12%)          | 46(12%)        | 46(12%)            |
| Tryptophan| 46(12%)               | 46(12%)         | 46(12%)          | 46(12%)        | 46(12%)            |
| Tyrosine   | 46(12%)               | 46(12%)         | 46(12%)          | 46(12%)        | 46(12%)            |
| Valine     | 46(12%)               | 46(12%)         | 46(12%)          | 46(12%)        | 46(12%)            |

Molecular weight

| Molecular Weight | Bundibugyo Ebolavirus | Sudan Ebolavirus | Reston Ebolavirus | Zaire Ebolavirus | Taiforest Ebolavirus |
|-----------------|-----------------------|-----------------|------------------|----------------|---------------------|
| Theoretical pI | 5.34                  | 5.39            | 5.40             | 5.40           | 5.40                |
| Aromatic content | 4600                 | 4649            | 4683             | 4683           | 4683                |
| Total number of positively charged residues | 40 | 39 | 39 | 39 | 39 |
| Total number of negatively charged residues | 27 | 27 | 27 | 27 | 27 |
| Extinction coefficient assuming cysteine residues | 5.435 M⁻¹ cm⁻¹ | 5.435 M⁻¹ cm⁻¹ | 5.435 M⁻¹ cm⁻¹ | 5.435 M⁻¹ cm⁻¹ | 5.435 M⁻¹ cm⁻¹ |
| Absorbance assuming cysteine residues | 0.865 | 0.865 | 0.865 | 0.865 | 0.865 |
| Extinction coefficient with out cysteine residues | 2.900 M⁻¹ cm⁻¹ | 2.900 M⁻¹ cm⁻¹ | 2.900 M⁻¹ cm⁻¹ | 2.900 M⁻¹ cm⁻¹ | 2.900 M⁻¹ cm⁻¹ |
| Absorbance with out cysteine residues | 0.614 | 0.614 | 0.614 | 0.614 | 0.614 |
| Instability Index | 56.23               | 56.23           | 56.23            | 56.23          | 56.23               |
| Aliphatic Index | 79.31                | 79.31           | 79.31            | 79.31          | 79.31               |
| Grand Average of Hydrophobicity | 0.624 | 0.624 | 0.624 | 0.624 | 0.624 |
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 Ebolavirus 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 without 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.

| Number and composition of amino acids | Bundibugyo Ebolavirus | Sudan Ebolavirus | Reston Ebolavirus | Zaire Ebolavirus | Taiforest Ebolavirus |
|--------------------------------------|-----------------------|-----------------|------------------|-----------------|---------------------|
| Alanine (Ala)                        | 121 (5.5%)            | 122 (5.5%)      | 115 (5.2%)       | 119 (5.4%)      | 127 (5.7%)          |
| Arginine (Arg)                       | 113 (5.3%)            | 113 (5.3%)      | 113 (5.3%)       | 116 (5.5%)      | 116 (5.5%)          |
| Asparagine (Asn)                     | 109 (4.9%)            | 110 (4.9%)      | 126 (5.6%)       | 106 (4.8%)      | 113 (5.1%)          |
| Aspartic acid (Asp)                  | 104 (4.7%)            | 99 (4.4%)       | 107 (4.9%)       | 101 (4.7%)      | 99 (4.4%)           |
| Cysteine (Cys)                       | 102 (4.4%)            | 95 (4.4%)       | 103 (5.6%)       | 107 (4.3%)      | 106 (4.3%)          |
| Glutamic acid (Glu)                  | 107 (4.5%)            | 104 (4.7%)      | 106 (4.5%)       | 109 (4.9%)      | 104 (4.7%)          |
| Glycine (Gly)                        | 102 (4.4%)            | 107 (4.8%)      | 109 (4.9%)       | 104 (4.7%)      | 104 (4.7%)          |
| Histidine (His)                      | 76 (3.2%)             | 72 (3.2%)       | 70 (3.4%)        | 72 (3.4%)       | 72 (3.4%)           |
| Isoleucine (Ile)                     | 106 (4.7%)            | 106 (4.7%)      | 110 (5.0%)       | 114 (5.0%)      | 110 (5.0%)          |
| Leucine (Leu)                        | 126 (5.5%)            | 126 (5.5%)      | 126 (5.5%)       | 126 (5.5%)      | 126 (5.5%)          |
| Lysine (Lys)                         | 133 (5.5%)            | 130 (4.9%)      | 116 (5.0%)       | 116 (5.0%)      | 116 (5.0%)          |
| Methionine (Met)                     | 36 (1.4%)             | 36 (1.4%)       | 34 (1.5%)        | 38 (1.7%)       | 35 (1.7%)           |
| Phenylalanine (Phe)                  | 106 (4.3%)            | 97 (4.0%)       | 102 (4.6%)       | 116 (5.2%)      | 106 (4.6%)          |
| Proline (Pro)                        | 105 (4.3%)            | 110 (5.0%)      | 98 (4.4%)        | 102 (4.6%)      | 107 (4.6%)          |
| Serine (Ser)                         | 103 (4.4%)            | 105 (4.5%)      | 104 (4.5%)       | 104 (4.5%)      | 104 (4.5%)          |
| Threonine (Thr)                      | 106 (4.5%)            | 106 (4.5%)      | 106 (4.5%)       | 106 (4.5%)      | 106 (4.5%)          |
| Tyrosine (Tyr)                       | 96 (4.0%)             | 90 (4.2%)       | 97 (4.0%)        | 77 (3.9%)       | 87 (3.9%)           |
| Value (Val)                          | 118 (5.3%)            | 116 (5.0%)      | 111 (5.0%)       | 116 (5.0%)      | 115 (5.2%)          |

| Molecular weight                     | 252724.47             | 252549.04       | 251649.28        | 251294.24       | 250746.25           |
| Theoretical pI                       | 6.4                   | 5.77            | 6.4             | 6.4             | 5.9                |
| Aromatic composition                 | 3586                  | 3586            | 3586            | 3586            | 3586               |
| Total number of negatively charged residues | 214                   | 232             | 229             | 231             | 226                |
| Extinction coefficient assuming cysteine residues | 206200M<sup>2</sup>cm<sup>-1</sup>M<sup>-1</sup>cm<sup>-1</sup>k<sup>-1</sup> | 206200M<sup>2</sup>cm<sup>-1</sup>M<sup>-1</sup>cm<sup>-1</sup>k<sup>-1</sup> | 206200M<sup>2</sup>cm<sup>-1</sup>M<sup>-1</sup>cm<sup>-1</sup>k<sup>-1</sup> | 206200M<sup>2</sup>cm<sup>-1</sup>M<sup>-1</sup>cm<sup>-1</sup>k<sup>-1</sup> | 206200M<sup>2</sup>cm<sup>-1</sup>M<sup>-1</sup>cm<sup>-1</sup>k<sup>-1</sup> |
| Extinction coefficient with out cysteine residues | 206200M<sup>2</sup>cm<sup>-1</sup>M<sup>-1</sup>cm<sup>-1</sup>k<sup>-1</sup> | 206200M<sup>2</sup>cm<sup>-1</sup>M<sup>-1</sup>cm<sup>-1</sup>k<sup>-1</sup> | 206200M<sup>2</sup>cm<sup>-1</sup>M<sup>-1</sup>cm<sup>-1</sup>k<sup>-1</sup> | 206200M<sup>2</sup>cm<sup>-1</sup>M<sup>-1</sup>cm<sup>-1</sup>k<sup>-1</sup> | 206200M<sup>2</sup>cm<sup>-1</sup>M<sup>-1</sup>cm<sup>-1</sup>k<sup>-1</sup> |
| Absorbance assuming cysteine residues | 1.256                 | 1.102           | 1.258           | 1.344           | 1.344              |
| Absorbance with out cysteine residues | 1.256                 | 1.102           | 1.258           | 1.344           | 1.344              |
| Instability Index                    | 89.74                 | 42.87           | 41.69           | 41.54           | 40.33              |
| Aliphatic Index                      | 92.48                 | 92.48           | 92.48           | 92.48           | 89.80              |
| Grand Average of Hydrophobicity      | 0.218                 | -0.206          | -0.252          | -0.252          | -0.191             |
MODULE-2
MULTIPLE SEQUENCE ALIGNMENT
MATRIX PROTEIN

**MEMBRANE ASSOCIATED PROTEIN**

| Sudan       | MRRVTIFTAFAFAFAYDIYFMNMRMLMKRNSAVGSGTGGEVEFEPDMTFPSENEMFRVAFDIVID 60 |
| Reston      | MRRVTIFTAFAFAFAYDIYFMNMRMLMKRNSAVGSGTGGEVEFEPDMTFPSENEMFRVAFDIVID 60 |
| Taiforest   | MRRVTIFTAFAFAFAYDIYFMNMRMLMKRNSAVGSGTGGEVEFEPDMTFPSENEMFRVAFDIVID 60 |
| Bundibuyo   | MRRVTIFTAFAFAFAYDIYFMNMRMLMKRNSAVGSGTGGEVEFEPDMTFPSENEMFRVAFDIVID 60 |

| Sudan       | MTSHTENGVASAFILATKYNVSGPKVLMKQPIFIPWPLGPDADQKTYSFDSTTTAIAIMLSY 120 |
| Reston      | MTSHTENGVASAFILATKYNVSGPKVLMKQPIFIPWPLGPDADQKTYSFDSTTTAIAIMLSY 120 |
| Taiforest   | MTSHTENGVASAFILATKYNVSGPKVLMKQPIFIPWPLGPDADQKTYSFDSTTTAIAIMLSY 120 |
| Bundibuyo   | MTSHTENGVASAFILATKYNVSGPKVLMKQPIFIPWPLGPDADQKTYSFDSTTTAIAIMLSY 120 |

| Sudan       | TTITFHKNNPLFRKMLVLIRGQHDPFLRLLRALRMAQFAQFLEGFSPVVSFVRQFQFQFDLTDL 180 |
| Reston      | TTITFHKNNPLFRKMLVLIRGQHDPFLRLLRALRMAQFAQFLEGFSPVVSFVRQFQFQFDLTDL 180 |
| Taiforest   | TTITFHKNNPLFRKMLVLIRGQHDPFLRLLRALRMAQFAQFLEGFSPVVSFVRQFQFQFDLTDL 180 |
| Bundibuyo   | TTITFHKNNPLFRKMLVLIRGQHDPFLRLLRALRMAQFAQFLEGFSPVVSFVRQFQFQFDLTDL 180 |

| Sudan       | LVTDPLAATPDPWDTPLPSLSLGRLPSFLSHPFLPRFPLVPLPCPGKEKGKHGDSLEPTSPKFK 240 |
| Reston      | LVTDPLAATPDPWDTPLPSLSLGRLPSFLSHPFLPRFPLVPLPCPGKEKGKHGDSLEPTSPKFK 240 |
| Taiforest   | LVTDPLAATPDPWDTPLPSLSLGRLPSFLSHPFLPRFPLVPLPCPGKEKGKHGDSLEPTSPKFK 240 |
| Bundibuyo   | LVTDPLAATPDPWDTPLPSLSLGRLPSFLSHPFLPRFPLVPLPCPGKEKGKHGDSLEPTSPKFK 240 |

| Sudan       | VNLMDQFVRDIDFVDFAEKICIGIEIEVEFVLEVKLQGKMGQKPGQFIPFVLLPFIYGLIDSPF 300 |
| Reston      | VNLMDQFVRDIDFVDFAEKICIGIEIEVEFVLEVKLQGKMGQKPGQFIPFVLLPFIYGLIDSPF 300 |
| Taiforest   | VNLMDQFVRDIDFVDFAEKICIGIEIEVEFVLEVKLQGKMGQKPGQFIPFVLLPFIYGLIDSPF 300 |
| Bundibuyo   | VNLMDQFVRDIDFVDFAEKICIGIEIEVEFVLEVKLQGKMGQKPGQFIPFVLLPFIYGLIDSPF 300 |

| Sudan       | GDLTVMTYPCDCSDCHSSCADSVLSEK -- 326 |
| Reston      | GDLTVMTYPCDCSDCHSSCADSVLSEK -- 326 |
| Taiforest   | GDLTVMTYPCDCSDCHSSCADSVLSEK -- 326 |
| Bundibuyo   | GDLTVMTYPCDCSDCHSSCADSVLSEK -- 326 |

| Sudan       | MAKATGSryNLYTFKRELEQGVSDFSLCNFLTPTVQKXWYVAGGFVQGKXGTTLNRL 60 |
| Reston      | MAKATGSryNLYTFKRELEQGVSDFSLCNFLTPTVQKXWYVAGGFVQGKXGTTLNRL 60 |
| Taiforest   | MAKATGSryNLYTFKRELEQGVSDFSLCNFLTPTVQKXWYVAGGFVQGKXGTTLNRL 60 |
| Bundibuyo   | MAKATGSryNLYTFKRELEQGVSDFSLCNFLTPTVQKXWYVAGGFVQGKXGTTLNRL 60 |

| Sudan       | KVDPAFAWAMTRNLFFHLFLNFQKVEFOQFQFQPIFPLAWRLVIAAGIQQLDMLHSDLBFLSGLN 120 |
| Reston      | KVDPAFAWAMTRNLFFHLFLNFQKVEFOQFQFQPIFPLAWRLVIAAGIQQLDMLHSDLBFLSGLN 120 |
| Taiforest   | KVDPAFAWAMTRNLFFHLFLNFQKVEFOQFQFQPIFPLAWRLVIAAGIQQLDMLHSDLBFLSGLN 120 |
| Bundibuyo   | KVDPAFAWAMTRNLFFHLFLNFQKVEFOQFQFQPIFPLAWRLVIAAGIQQLDMLHSDLBFLSGLN 120 |

| Sudan       | LDWDLTTSTTSHNFMRTQFQKDQSMGSLMLSLIRSNITNPNKFNKLFTPFLHVUNYVGILSSVE 180 |
| Reston      | LDWDLTTSTTSHNFMRTQFQKDQSMGSLMLSLIRSNITNPNKFNKLFTPFLHVUNYVGILSSVE 180 |
| Taiforest   | LDWDLTTSTTSHNFMRTQFQKDQSMGSLMLSLIRSNITNPNKFNKLFTPFLHVUNYVGILSSVE 180 |
| Bundibuyo   | LDWDLTTSTTSHNFMRTQFQKDQSMGSLMLSLIRSNITNPNKFNKLFTPFLHVUNYVGILSSVE 180 |

| Sudan       | IGTFAYIALITRTMNSLGVEQEPKQSMIRSHFVFKSLHGEMLTFLFAVPIFPEPSJL 240 |
| Reston      | IGTFAYIALITRTMNSLGVEQEPKQSMIRSHFVFKSLHGEMLTFLFAVPIFPEPSJL 240 |
| Taiforest   | IGTFAYIALITRTMNSLGVEQEPKQSMIRSHFVFKSLHGEMLTFLFAVPIFPEPSJL 240 |
| Bundibuyo   | IGTFAYIALITRTMNSLGVEQEPKQSMIRSHFVFKSLHGEMLTFLFAVPIFPEPSJL 240 |

| Sudan       | LIFENSSSLAI 251 |
| Reston      | LIFENSSSLAI 251 |
| Taiforest   | LIFENSSSLAI 251 |
| Bundibuyo   | LIFENSSSLAI 251 |

**MEMBRANE ASSOCIATED PROTEIN**

Sudan:

Reston:

Taiforest:

Bundibuyo:
POLYMERASE COMPLEX PROTEIN
### SECOND SECRETED GLYCO PROTEIN

| Country   | Accession Number | Description |
|-----------|------------------|-------------|
| Taiforest | MGASGLIQPLRFRKTSFFVWVILFHKWVFSLPGYVHNNTLCVSIDKEKVCIRKLSS |             |
| Bundibugyo | MGVSTIGGLQPLRFRKTSFFVWVILFHKWVFSLPGYVHNNTLCVSIDKEKVCIRKLSS |             |
| Zaire    | MGSTIGGLQPLRFRKTSFFVWVILFHKWVFSLPGYVHNNTLCVSIDKEKVCIRKLSS |             |
| Reston   | MGGTIGGLQPLRFRKTSFFVWVILFHKWVFSLPGYVHNNTLCVSIDKEKVCIRKLSS |             |
| Sudan    | MGGLIGGLQPLRFRKTSFFVWVILFHKWVFSLPGYVHNNTLCVSIDKEKVCIRKLSS |             |

### SMALL SECRETED GLYCO PROTEIN

| Country   | Accession Number | Description |
|-----------|------------------|-------------|
| Taiforest | CVQVLEARTTFPGFVLLNNDTITYDNQNKNTTQGKMLINFTDVTMSMEAWEMKKHLN |             |
| Bundibugyo | CVQVLEARTTFPGFVLLNNDTITYDNQNKNTTQGKMLINFTDVTMSMEAWEMKKHLN |             |
| Zaire    | CVQVLEARTTFPGFVLLNNDTITYDNQNKNTTQGKMLINFTDVTMSMEAWEMKKHLN |             |
| Reston   | CVQVLEARTTFPGFVLLNNDTITYDNQNKNTTQGKMLINFTDVTMSMEAWEMKKHLN |             |
| Sudan    | CVQVLEARTTFPGFVLLNNDTITYDNQNKNTTQGKMLINFTDVTMSMEAWEMKKHLN |             |

| Country   | Accession Number | Description |
|-----------|------------------|-------------|
| Taiforest | FFK---------- | 302         |
| Bundibugyo | FFK---------- | 302         |
| Zaire    | FFK---------- | 297         |
| Reston   | FFK---------- | 331         |
| Sudan    | FFK---------- | 318         |
SPIKE GLYCO-PROTEIN

Zaire
- MSGVTILQLPLRDKFRTSFLLFLVMVLIFQRTTSIPFILGVHINSLILTVQVSİLDKVCRCRKLSS59
Zaire
- MCGSGLQGFAPFKRKSFLKSFYWYVIIHFKVYFLIGVHNNTLQVSİLDKVCRCRKLSS59
Zaire
- MVTSIGLQPLRDKFRTSFLLFLVMVLIFQRTTSIPFILGVHINSLILTVQVSİLDKVCRCRKLSS59
Zaire
- MSGSYQLQLPLRDKFRTSFLLFLVMVLIFQRTTSIPFILGVHINSLILTVQVSİLDKVCRCRKLSS59
Zaire
- MSGSLQGFAPFKRKSFLKSFYWYVIIHFKVYFLIGVHNNTLQVSİLDKVCRCRKLSS59

Taiforest
- TQÇQDFQ---------- 365
Taiforest
- TQÇRHPQOTQŚQQL 373
Taiforest
- TÇQCFKQ---------- 364
Taiforest
- RCKRČQK---------- 367
Taiforest
- QHČRĮRQKVEE- 372

Sudan
- TQÇQDFQ---------- 365
Sudan
- TQÇRHPQOTQŚQQL 373
Sudan
- TÇQCFKQ---------- 364
Sudan
- RCKRČQK---------- 367
Sudan
- QHČRĮRQKVEE- 372

Bundibugyo
- MTUVŚQGQPLRDKFRTSFLLFLVMVLIFQRTTSIPFILGVHINSLILTVQVSİLDKVCRCRKLSS59
Bundibugyo
- TQÇQDFQ---------- 365
Bundibugyo
- TQÇRHPQOTQŚQQL 373
Bundibugyo
- TÇQCFKQ---------- 364
Bundibugyo
- RCKRČQK---------- 367
Bundibugyo
- QHČRĮRQKVEE- 372

Reston
- MÇGŚGLQGFAPFKRKSFLKSFYWYVIIHFKVYFLIGVHNNTLQVSİLDKVCRCRKLSS59
Reston
- MVTSIGLQPLRDKFRTSFLLFLVMVLIFQRTTSIPFILGVHINSLILTVQVSİLDKVCRCRKLSS59
Reston
- MSGSYQLQLPLRDKFRTSFLLFLVMVLIFQRTTSIPFILGVHINSLILTVQVSİLDKVCRCRKLSS59
Reston
- MSGSLQGFAPFKRKSFLKSFYWYVIIHFKVYFLIGVHNNTLQVSİLDKVCRCRKLSS59
Reston
- MSGVTILQLPLRDKFRTSFLLFLVMVLIFQRTTSIPFILGVHINSLILTVQVSİLDKVCRCRKLSS59

Published Online January 2020 in IJEAST (http://www.ijeast.com)
RNA DEPENDENT RNA POLYMERASE

RNA DEPENDENT RNA POLYMERASE

RNA DEPENDENT RNA POLYMERASE

RNA DEPENDENT RNA POLYMERASE

RNA DEPENDENT RNA POLYMERASE

RNA DEPENDENT RNA POLYMERASE

RNA DEPENDENT RNA POLYMERASE

RNA DEPENDENT RNA POLYMERASE

RNA DEPENDENT RNA POLYMERASE

RNA DEPENDENT RNA POLYMERASE

RNA DEPENDENT RNA POLYMERASE

RNA DEPENDENT RNA POLYMERASE

RNA DEPENDENT RNA POLYMERASE

RNA DEPENDENT RNA POLYMERASE

RNA DEPENDENT RNA POLYMERASE
## Nucleoprotein:

| Gene number; Exon number and Exon type | Tai Forest Ebola virus | Bundibugyo Ebola virus | Sudan Ebola virus | Zaïre Ebola virus | Reston Ebola virus |
|----------------------------------------|------------------------|------------------------|-------------------|-------------------|-------------------|
| Exon Number = 01; Type of Exon = Nsp1 | Gene Number = 01       | Gene Number = 01       | Gene Number = 01  | Gene Number = 01  | Gene Number = 01  |
| Exon Number = 04; Type of Exon = Nsp1 |                        |                        |                    |                   |                   |

| 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-2 (Input Strand) | E-1 (Input Strand) |

| Beginning of Exon/Signal | E-1 = 403 |
|--------------------------|----------|
| E-2 = 2877               |          |

| Ending of Exon/Signal | E-1 = 2619 |
|-----------------------|-----------|
| E-2 = 2739            |          |

| Length of Exon/Signal | E-1 = 2320 |
|-----------------------|-----------|
| E-2 = 06              |          |

| Reading Frame | E-1 = 0 |
|---------------|---------|
| E-2 = ( )    |         |

| Net Phase of Exon/Signal | E-1 = 0 |
|--------------------------|---------|
| E-2 = ( )                |         |

| Initiation signal 3'-Splice site score | E-1 = 55 |
|----------------------------------------|---------|
| E-2 = ( )                              |         |

| Termination signal 5'-Splice site score | E-1 = 41 |
|----------------------------------------|---------|
| E-2 = ( )                              |         |

| Coding Region score | E-1 = 1420 |
|---------------------|------------|
| E-2 = ( )           |           |

| Probability of Exon | E-1 = 0.921 |
|---------------------|--------------|
| E-2 = ( )           |              |

| Exon score | E-1 = 132.45 |
|------------|--------------|
| E-2 = 1.05  |              |
| E-1 = 129.55 |             |
| E-2 = 1.75  |              |

## Polymerase complex protein

| Gene number; Exon number and Exon type | Tai Forest Ebola virus | Bundibugyo Ebola virus | Sudan Ebola virus | Zaïre Ebola virus | Reston Ebola virus |
|----------------------------------------|------------------------|------------------------|-------------------|-------------------|-------------------|
| Exon Number = 01; Type of Exon = Nsp1 | Gene Number = 01       | Gene Number = 01       | Gene Number = 01  | Gene Number = 01  | Gene Number = 01  |
| Exon Number = 04; Type of Exon = Nsp1 |                        |                        |                    |                   |                   |

| 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-2 (Input Strand) | E-1 (Input Strand) |

| Beginning of Exon/Signal | E-1 = 89 |
|--------------------------|---------|
| E-2 = 1303               |          |

| Ending of Exon/Signal | E-1 = 1114 |
|-----------------------|------------|
| E-2 = 1308            |           |

| Length of Exon/Signal | E-1 = 1115 |
|-----------------------|------------|
| E-2 = 1348            |           |

| Reading Frame | E-1 = 903 |
|---------------|----------|
| E-2 = ( )    |         |

| Net Phase of Exon/Signal | E-1 = 01 |
|--------------------------|---------|
| E-2 = ( )                |         |
| Initiation signal/3'-Splice site score | E-1 =42 | E-1 = 73 | E-1 =71 | E-1 = 60 | E-1 = 69 | E-2 = ( ) | E-2 = ( ) | E-2 = ( ) | E-2 = ( ) |
|--------------------------------------|---------|---------|---------|---------|---------|-----------|-----------|-----------|-----------|
| Termination signal/5'-Splice site score | E-1=3 | E-1 = 45 | E-1 = 43 | E-1 = 52 | E-1 = 48 | E-2 = ( ) | E-2 = ( ) | E-2 = ( ) | E-2 = ( ) |
| Coding Region score | E-1 =568 | E-1 = 296 | E-1 = 819 | E-1 = 843 | E-1 = 833 | E-2 = ( ) | E-2 = ( ) | E-2 = ( ) | E-2 = ( ) |
| Probability of Exon | E-1 = 0.455 | E-1 = 0.283 | E-1 = 0.534 | E-1 = 0.847 | E-1 = 0.846 | E-2 = ( ) | E-2 = ( ) | E-2 = ( ) | E-2 = ( ) |
| Exon score | E-1=37.91 | E-1 = 31.09 | E-1 = 72.31 | E-1 = 75.31 | E-1 = 72.01 | E-2 = 1.05 | E-2 = -1.75 | E-2 = -1.75 | E-2 = -1.75 |

Matrix Protein:

| Tai Forest Ebola virus | Bundibugyo Ebola virus | Sudan Ebola virus | Zaire Ebola virus | Reston Ebola virus |
|------------------------|------------------------|------------------|------------------|-------------------|
| Gene number:Exon number and Exon type | Gene number:Exon number and Exon type | Gene number:Exon number and Exon type | Gene number:Exon number and Exon type | Gene number:Exon number and Exon type |
| Type of DNA Strand + = Input Strand - = Output Strand | Type of DNA Strand + = Input Strand - = Output Strand | Type of DNA Strand + = Input Strand - = Output Strand | Type of DNA Strand + = Input Strand - = Output Strand | Type of DNA Strand + = Input Strand - = Output 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 = 1145 | E-1 = 90 | E-2 = 1274 |
| Ending of Exon/Signal | E-1 = 900 | E-2 = 1498 | E-1 = 1004 | E-2 = 1220 | 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 = 88 | E-1 = 981 | E-2 = 86 | E-1 = 996 | E-2 = 86 |
| Reading Frames | E-1 = 07 | 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 = 24.67 | 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 Ebola virus | Bundibugyo Ebola virus | Sudan Ebola virus | Zaire Ebola virus | Reston Ebola virus |
|------------------------|------------------------|------------------|------------------|-------------------|
| Gene number:Exon number and Exon type | Gene number:Exon number and Exon type | Gene number:Exon number and Exon type | Gene number:Exon number and Exon type | Gene number:Exon number and Exon type |
| Type of DNA Strand + = Input Strand - = Output Strand | Type of DNA Strand + = Input Strand - = Output Strand | Type of DNA Strand + = Input Strand - = Output Strand | Type of DNA Strand + = Input Strand - = Output Strand | Type of DNA Strand + = Input Strand - = Output 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 = 1145 | E-1 = 90 | E-2 = 1274 |
| Ending of Exon/Signal | E-1 = 900 | E-2 = 1498 | E-1 = 1004 | E-2 = 1220 | 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 = 88 | E-1 = 981 | E-2 = 86 | E-1 = 996 | E-2 = 86 |
| Reading Frames | E-1 = 07 | 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 = 24.67 | 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 |
Second Secreted Glycoprotein:

| Tai Forest Ebolavirus | Bundibugyo Ebolavirus | Sudan Ebolavirus | Zaire Ebolavirus | Reston Ebolavirus |
|------------------------|------------------------|------------------|------------------|-------------------|
| Gene number: Exon number and Exon type | Gene number: Exon number and Exon type | Gene number: Exon number and Exon type | Gene number: Exon number and Exon type | Gene number: Exon number and Exon type |
| Gene number: Exon number | Gene number: Exon number | Gene number: Exon number | Gene number: Exon number | Gene number: Exon number |
| Type of DNA Strand | Type of DNA Strand | Type of DNA Strand | Type of DNA Strand | Type of DNA Strand |
| + = Input Strand | - = Output Strand | + = Input Strand | - = Output Strand | + = Input Strand |
| Type of Exon = Term | Type of Exon = Term | Type of Exon = Term | Type of Exon = Term | Type of Exon = Term |
| Gene number = 01 | Gene number = 01 | Gene number = 01 | Gene number = 01 | Gene number = 01 |
| Exon number = 01 | Exon number = 01 | Exon number = 01 | Exon number = 01 | Exon number = 01 |
| Type of Exon = Infr | Type of Exon = Infr | Type of Exon = Infr | Type of Exon = Infr | Type of Exon = Infr |
| Gene number = 01 | Gene number = 01 | Gene number = 01 | Gene number = 01 | Gene number = 01 |
| Exon number = 03 | Exon number = 03 | Exon number = 03 | Exon number = 03 | Exon number = 03 |
| Type of Exon = P/A | Type of Exon = P/A | Type of Exon = P/A | Type of Exon = P/A | Type of Exon = P/A |
| Type of DNA Strand | Type of DNA Strand | Type of DNA Strand | Type of DNA Strand | Type of DNA Strand |
| + = Input Strand | - = Output Strand | + = Input Strand | - = Output Strand | + = Input Strand |
| Type of Exon = Term | Type of Exon = Term | Type of Exon = Term | Type of Exon = Term | Type of Exon = Term |
| Gene number = 01 | Gene number = 01 | Gene number = 01 | Gene number = 01 | Gene number = 01 |
| Exon number = 02 | Exon number = 02 | Exon number = 02 | Exon number = 02 | Exon number = 02 |
| Type of Exon = Term | Type of Exon = Term | Type of Exon = Term | Type of Exon = Term | Type of Exon = Term |
| Gene number = 01 | Gene number = 01 | Gene number = 01 | Gene number = 01 | Gene number = 01 |
| Exon number = 03 | Exon number = 03 | Exon number = 03 | Exon number = 03 | Exon number = 03 |
| Type of Exon = P/A | Type of Exon = P/A | Type of Exon = P/A | Type of Exon = P/A | Type of Exon = P/A |
| Type of DNA Strand | Type of DNA Strand | Type of DNA Strand | Type of DNA Strand | Type of DNA Strand |
| + = Input Strand | - = Output Strand | + = Input Strand | - = Output Strand | + = Input Strand |
| Type of Exon = Term | Type of Exon = Term | Type of Exon = Term | Type of Exon = Term | Type of Exon = Term |
| Gene number = 01 | Gene number = 01 | Gene number = 01 | Gene number = 01 | Gene number = 01 |
| Exon number = 02 | Exon number = 02 | Exon number = 02 | Exon number = 02 | Exon number = 02 |
| Type of Exon = Term | Type of Exon = Term | Type of Exon = Term | Type of Exon = Term | Type of Exon = Term |
| Gene number = 01 | Gene number = 01 | Gene number = 01 | Gene number = 01 | Gene number = 01 |
| Exon number = 03 | Exon number = 03 | Exon number = 03 | Exon number = 03 | Exon number = 03 |
| Type of Exon = P/A | Type of Exon = P/A | Type of Exon = P/A | Type of Exon = P/A | Type of Exon = P/A |

267
### Spike Glycoprotein

| TaT Forest Ebolavirus | Bundibugyo Ebolavirus | Sudan Ebolavirus | Zaire Ebolavirus | Reston Ebolavirus |
|-----------------------|-----------------------|-----------------|------------------|------------------|
| **Gene number, Exon number and Exon type** | **Gene number = 01** | **Gene number = 01** | **Gene number = 01** | **Gene number = 01** |
| **Type of DNA Strand** | **Exon number = 01** | **Exon number = 01** | **Exon number = 01** | **Exon number = 01** |
| **+ = Input Strand** | **Type of Exon = Term** | **Type of Exon = Term** | **Type of Exon = Term** | **Type of Exon = Term** |
| **- = Output Strand** | **Type of Exon = Term** | **Type of Exon = Term** | **Type of Exon = Term** | **Type of Exon = Term** |
| **Beginning of Exon/Signal** | **E1 = 140** | **E1 = 140** | **E1 = 140** | **E1 = 140** |
| **E2 = 1076** | **E2 = 1076** | **E2 = 1076** | **E2 = 1076** | **E2 = 1076** |
| **E3 = 2356** | **E3 = 2356** | **E3 = 2356** | **E3 = 2356** | **E3 = 2356** |
| **Ending of Exon/Signal** | **E1 = 995** | **E1 = 995** | **E1 = 995** | **E1 = 995** |
| **E2 = 2169** | **E2 = 2169** | **E2 = 2169** | **E2 = 2169** | **E2 = 2169** |
| **E3 = 2361** | **E3 = 2361** | **E3 = 2361** | **E3 = 2361** | **E3 = 2361** |
| **Length of Exon/Signal** | **E1 = 856** | **E1 = 856** | **E1 = 856** | **E1 = 856** |
| **E2 = 1094** | **E2 = 1094** | **E2 = 1094** | **E2 = 1094** | **E2 = 1094** |
| **E3 = 868** | **E3 = 868** | **E3 = 868** | **E3 = 868** | **E3 = 868** |
| **Reading Frame** | **E1 = 01** | **E1 = 01** | **E1 = 01** | **E1 = 01** |
| **E2 = 00** | **E2 = 00** | **E2 = 00** | **E2 = 00** | **E2 = 00** |
| **E3 = ( )** | **E3 = ( )** | **E3 = ( )** | **E3 = ( )** | **E3 = ( )** |
| **Net-Phase of Exon/Signal** | **E1 = 01** | **E1 = 01** | **E1 = 01** | **E1 = 01** |
| **E2 = 02** | **E2 = 02** | **E2 = 02** | **E2 = 02** | **E2 = 02** |
| **E3 = ( )** | **E3 = ( )** | **E3 = ( )** | **E3 = ( )** | **E3 = ( )** |
| **Initiation signal 5’-Splice site score** | **E1 = 84** | **E1 = 84** | **E1 = 84** | **E1 = 84** |
| **E2 = 35** | **E2 = 35** | **E2 = 35** | **E2 = 35** | **E2 = 35** |
| **E3 = ( )** | **E3 = ( )** | **E3 = ( )** | **E3 = ( )** | **E3 = ( )** |
| **Termination signal 5’-Splice site score** | **E1 = 92** | **E1 = 92** | **E1 = 92** | **E1 = 92** |
| **E2 = 43** | **E2 = 43** | **E2 = 43** | **E2 = 43** | **E2 = 43** |
| **E3 = ( )** | **E3 = ( )** | **E3 = ( )** | **E3 = ( )** | **E3 = ( )** |
**RNA Dependent RNA Polymerase:**

| Gene number:Exon number and Exon type | Tat Forest Ebolavirus | Bundibugyo Ebolavirus | Sudan Ebolavirus | Zaire Ebolavirus | Reston Ebolavirus |
|--------------------------------------|-----------------------|-----------------------|------------------|------------------|------------------|
| Gene number = 01 Exon number = 01 Type of Exon = Sagl | Gene number = 01 Exon number = 01 Type of exon = Sagl | Gene number = 01 Exon number = 01 Type of Exon = Sagl | Gene number = 01 Exon number = 01 Type of Exon = Sagl | Gene number = 01 Exon number = 01 Type of Exon = Sagl | Gene number = 01 Exon number = 01 Type of Exon = Sagl |
| Gene number = 01 Exon number = 02 Type of Exon = PyA | Gene number = 01 Exon number = 02 Type of Exon = PyA | Gene number = 01 Exon number = 02 Type of Exon = PyA | Gene number = 01 Exon number = 02 Type of Exon = PyA | Gene number = 01 Exon number = 02 Type of Exon = PyA | Gene number = 01 Exon number = 02 Type of Exon = PyA |

**Type of DNA Strand**
- = Input Strand
- = Output Strand

| Type of DNA Strand | Tat Forest Ebolavirus | Bundibugyo Ebolavirus | Sudan Ebolavirus | Zaire Ebolavirus | Reston Ebolavirus |
|--------------------|-----------------------|-----------------------|------------------|------------------|------------------|
| E-1 (+ = Input Strand) E-1 (- = Output Strand) | E-1 (+ = Input Strand) E-1 (- = Output Strand) | E-1 (+ = Input Strand) E-1 (- = Output Strand) | E-1 (+ = Input Strand) E-1 (- = Output Strand) | E-1 (+ = Input Strand) E-1 (- = Output Strand) | E-1 (+ = Input Strand) E-1 (- = Output Strand) |

**Coding Region score**
E-1 = 282
E-2 = 497
E-3 = ( )

**Probability of Exon**
E-1 = 0.896
E-2 = 0.979
E-3 = ( )

| Exon score | Tat Forest Ebolavirus | Bundibugyo Ebolavirus | Sudan Ebolavirus | Zaire Ebolavirus | Reston Ebolavirus |
|------------|-----------------------|-----------------------|------------------|------------------|------------------|
| E-1 = 23.30 E-2 = 32.16 E-3 = 1.05 | E-1 = 19.10 E-2 = 17.20 E-3 = 1.03 | E-1 = 15.51 E-2 = 31.64 | E-1 = 18.84 E-2 = 39.43 | E-1 = 17.09 E-2 = 26.36 |

**Membrane Associated Protein:**

| Gene number:Exon number and Exon type | Tat Forest Ebolavirus | Bundibugyo Ebolavirus | Sudan Ebolavirus | Zaire Ebolavirus | Reston Ebolavirus |
|--------------------------------------|-----------------------|-----------------------|------------------|------------------|------------------|
| Gene number = 01 Exon number = 01 Type of Exon = Tmem | Gene number = 01 Exon number = 01 Type of Exon = Sagl | Gene number = 01 Exon number = 01 Type of Exon = Sagl | Gene number = 01 Exon number = 01 Type of Exon = Sagl | Gene number = 01 Exon number = 01 Type of Exon = Sagl | Gene number = 01 Exon number = 01 Type of Exon = Sagl |
| Gene number = 01 Exon number = 02 Type of Exon = PyA | Gene number = 01 Exon number = 02 Type of Exon = PyA | Gene number = 01 Exon number = 02 Type of Exon = PyA | Gene number = 01 Exon number = 02 Type of Exon = PyA | Gene number = 01 Exon number = 02 Type of Exon = PyA | Gene number = 01 Exon number = 02 Type of Exon = PyA |

**Type of DNA Strand**
E-1 (+ = Input Strand) E-1 (- = Input Strand) E-1 (+ = Input Strand) E-1 (- = Input Strand)

269
| Gene number/Exon number and Exon type | Tai Forest Ebola Virus | Bundibugyo Ebola Virus | Sudan Ebola Virus | Zaire Ebola Virus | Reston Ebola Virus |
|-------------------------------------|-----------------------|------------------------|------------------|------------------|------------------|
| Type of DNA Strand                 | E-1= Input Strand     | E-1= Input Strand      | E-1= Input Strand | E-1= Input Strand | E-1= Input Strand |
| E-2= Output Strand                | E-2= Output Strand    | E-2= Output Strand     | E-2= Output Strand | E-2= Output Strand | E-2= Output Strand |

**Minor Nucleoprotein:**

| Probability of Exon | E-1=0.990 |
|---------------------|-----------|
| E-2=           |   |

| Exon score         | E-1=50.17 |
|--------------------|-----------|
| E-2=           |   |

| Probability of Exon | E-1=0.965 |
|---------------------|-----------|
| E-2=           |   |

| Exon score         | E-1=41.07 |
|--------------------|-----------|
| E-2=           |   |

| Probability of Exon | E-1=0.971 |
|---------------------|-----------|
| E-2=           |   |

| Exon score         | E-1=40.44 |
|--------------------|-----------|
| E-2=           |   |

| Probability of Exon | E-1=0.891 |
|---------------------|-----------|
| E-2=           |   |

| Exon score         | E-1=23.84 |
|--------------------|-----------|
| E-2=           |   |

| Probability of Exon | E-1=0.693 |
|---------------------|-----------|
| E-2=           |   |

| Exon score         | E-1=46.62 |
|--------------------|-----------|
| E-2=           |   |

| Probability of Exon | E-1=1.85 |
|---------------------|-----------|
| E-2=           |   |
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. Ebolavirus 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 functional 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 alignments 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-Teljeiro JD, Illera G, Barroso A, Vilà C, Walsh PD. Ebola outbreak killed 5000 gorillas. Science 2006;314:1564. 2.
12Chepurnov AA, Bakulina LF, Dadaeva AA, Ustinova EN, Chepurnova TS, Baker JR Jr. Inactivation of Ebola virus