International Journal of Scholarly Research in
Science and Technology
Journal homepage: https://srrjournals.com/ijsrst/
ISSN: 2961-3337 (Online)

(RESEARCH ARTICLE)

In silico characterization of three hypothetical proteins in the Monkeypox virus

Mundaganur Dastigir [1], Mundaganur Yasmeen [2] and Ashokan Kannarath [3,*]

[1] *Department of Zoology, Willingdon College, Sangli, India.*
[2] *Department of Zoology, Miraj College, Miraj, India.*
[3] *Department of zoology (research Guide) JJT University, Rajasthan, India.*

International Journal of Scholarly Research in Science and Technology, 2023, 03(01), 001–007

Publication history: Received on 28 May 2023; revised on 06 July 2023; accepted on 09 July 2023

Article DOI: https://doi.org/10.56781/ijsrst.2023.3.1.0034

Abstract

Monkeypox virus infection and death toll are alarmingly increasing the world over. An effective treatment and control protocol is essential to fight this viral infection. In this, in silico research article, we analyzed three hypothetical protein sequences from the Monkeypox virus. We studied various parameters of these proteins including amino acid composition, instability index, isoelectric point, extinction coefficient, aliphatic index, GRAVY, etc. We also performed MSA to get a clear idea of the mutation and its type and the virus's role over time. We find out various domains and their function that permit the virus during its infection processes. The amphipathicity is validated well. This study will help scientists to develop effective drugs against this virus.

Keywords: Viral infection; Monkeypox; MSA; Amphipathicity; GRAVY; EC; II

1 Introduction

The eradication of smallpox in 1980, and the cessation of smallpox vaccination leads to a suitable atmosphere for the emergence of monkeypox [1]. Monkeypox is a zoonotic orthopox virus that causes human disease much like smallpox. The only difference is the severity of monkeypox is lower than smallpox with low mortality. It is clinically very significant as it is endemic to western and central Africa. Outbreaks occur in the Western hemisphere due to the exotic pet trade and international migration. The monkeypox virus was first isolated and characterized in 1958[1]. The first confirmed human case was identified in 1970[2].

It belongs to the family *Poxiviridae*, subfamily *chordopoxyviriniae*, genus *orthopoxivirus,* and species Monkeypox virus. The virus is about 200-250 nanometres in size, brick-shaped with lipoprotein envelopes having double-stranded DNA[3,4]. It differs from other viruses as the genome contains all the genetic information in its genome necessary for replication, transcription, assembly, and egress. This virus depends on the host only for ribosomes necessary for mRNA transcription [3,5].

The present study is to determine the functional role and other parameters of three hypothetical protein sequences collected from NCBI. Except for three protein sequences deposited in the NCBI repository, three proteins' roles and other physicochemical parameters are unknown. Palmytilated EEV membrane protein, Ankyrin-like protein, EEV type-1 membrane glycoprotein, complement control protein, MPXV-COP-175, A32-5L, Viral membrane formation protein, Viron morphogenesis protein, Late protein H7, Crescent membrane and immature Viron formation protein, B17R, EEV membrane glycoprotein, DNA ligase, bifunctional hydroxysteroid dehydrogenase, DNA helicase, uracil-DNA glycosylase, Glutaredoxin, ribonucleotide reductase small subunit, Uracil-DNA glycosylase, etc are some of the proteins that are characterized form monkeypox virus. The structural details reveal that the monkeypox virus has distinctive surface tubules and dumbbell-shaped core components. The virus has a lipoprotein envelope and linear dsDNA.

* Corresponding author: Ashokan Kannarath

2 Material And method

The in-silico method is applied to characterize the two hypothetical protein sequences extracted from the NCBI data bank. The word protein of monkeypox virus prompted about 1070 protein sequences deposited in the repertories of the NCBI. Out of these 1070 sequences we extracted only three hypothetical protein sequences for characterization (Table1)

Table 1 Monkey pox virus hypothetical sequences extracted from NCBI

Accession number	No. of amino acids	Source	Annotation
YP_010415322.1	59	Monkey pox virus	Hypothetical
QGQ59750.1	43	Monkey pox virus	Hypothetical
QGQ59749.1	46	Monkey pox virus	Hypothetical

2.1 Sequence parameters

These sequences were analyzed for amino acid composition, instability index, GRAVY, negatively and positively charged amino acids, extinction coefficient, half-life aliphatic index, and isoelectric point by using Expassy [6] .

2.2 MSA

The sequences under study are subjected to multiple sequence alignment [7]. The data obtained from MSA were used for phylogenetic tree construction

2.3 Functional domain prediction

Protein domains often correspond to structural domains which are self-stabilizing and fold independently of the rest of the protein chain. They may occur independently or as part of complex multidomain protein architectures which evolve by domain accretion, domain loss, or domain recombination. We used the server PredictProtein to predict the functional domain of the target protein [8].

2.4 Emboss Pepwheel-to find out the amphipathicity of protein residues

Pepwheel draws a helical wheel diagram for a protein sequence. This displays the sequence in a helical representation as if looking down the axis of the helix. It is useful for highlighting amphipathicity and other properties of residues around a helix. By default, aliphatic residues are marked with squares, hydrophilic residues are marked with diamonds, and positively charged residues with octagons, although this can be changed. The predicted transmembrane helix was visualized and analyzed using a helical wheel plot generated by the program Pepwheel [9] included in the EMBOS 2.7 suit.

The presence of disulfide bridges (“SS” bonds) was analyzed by two methods. The first method involves the prediction of SS bonds using the primary structure (protein sequence data) by the tool CYS_REC (http://sunI.softberry.com/berry.html? trpic). CYS_REC identifies the position of cysteine, the total number of cysteine present, and the pattern, if present, of pairs in the protein sequence. The second method involves the visualization and identification of “SS” bonds using the three-dimensional structure of the protein (3D coordinate data).

The 3D structure of MAP (AIXF84) was generated using Expassy s server[10]. Similar 3D structures for MAP in the protein data bank (www.rscb.org) were identified by BLAST analysis (http://www.ncbin/mnil.gov.80/BLAST/). The modelled 3D structure was evaluated using the online server Rampage[11], ProQ[12], and CE (combinatorial extension) [13]. The tool Rasmol (http://openrasmol.org) is used to visualize the modelled 3D structures and to identify the cysteine and the presence of “SS” bonds. The presence of signal peptides and their localization was predicted by using SignalP 3.0 server [14]. The phosphorylation site in serine, threonine, and tyrosine of MAPs was predicted by the NetPhos program of Expasy.

3 Result and Discussion

The selected proteins sequences and their accession number and amino acid composition in percentage (table1), hypothetical protein NBT03_gp144, Monkeypox virus, Gly, Pyl and sec are least and is nil, the same trend is repeated in

hypothetical protein PDLMKLCO_00029 Monkeypox virus, but in hypothetical protein PDLMKLCO_00030, Monkeypox virus shows Pyl and Sex is nil with Gly 2,3 %. This abnormal occurrence of Gly may be due to geographical variation and survival in temperate regions. Ly and Ser were high in some and Gly in others and Val, Lys, and Tyr high in still other monkeypox viruses may indicate its polymorphic existence and survival in extreme environmental conditions by providing specific protein composition.

Table 2 Composition of various amino acids in the selected proteins sequences

Amino acids	YP_010415322.1	QGQ59750.1	QGQ59749.1
Ala	1.7	9.3	00
Arg	3.4	2.3	6.5
Asn	3.4	4.7	8.7
Asp	8.5	7.0	8.7
Cys	3.4	4.7	2.2
Gln	00	2.3	00
Glu	3.4	7.0	2.2
Gly	1.7	11.6	4.3
His	3.4	2.3	4.3
Ile	5.1	2.3	4.3
Leu	3.4	4.7	2.2
Lys	10.25	4.7	17.4
Met	5.1	2.3	4.3
Phe	5.1	4.7	2.2
Pro	1.7	4.7	00
Ser	18.6	4.7	00
Thr	10.2	2.3	4.3
Trp	1.7	00	00
Tyr	8.5	14	10.9
Val	1.7	4.7	17.4
Pyl	00	00	00
Sec	00	00	00

The other parameters studied (Table 2).

Table 3 Other parameter of the protein sequences studied**

Accession number	Mol.wt	(-) ve AA	(+) ve AA	EC	HL	II	AI	GRAVY	pI
YP_010415322.1	6933.74	7	8	13075	30 hrs.	41.01	39.66	-0.788	7.77
QGQ59750.1	4982.67	6	2	8940.0	29.73 hrs.	30	95.35	0.16	4.29
QGQ59749.1	5609.00	5	11	7450.0	30 hrs.	3.36	75.87	-0.778	9.67

** Mol.wt-Molecular weight, AA-Amino acids, EC-Extinction coefficient, HL-Half life, II-Instability index; AI-Aliphatic index, GRAVY-Grand Average of Hydropathy

The hypothetical protein NBT03_gp144 and QGQ59749.1 show more positively charged amino acids, but QGQ59750.1 shows more negatively charged amino acids. The extinction coefficient is highest in hypothetical protein YP_010415322.1 and is more or less equal in the other two sequences studied. The half-life period shows no variation among the selected hypothetical proteins. The instability index is least in QGQ59749.1 and light variation in the other two sequences studied. An aliphatic index is highest in QGQ59750.1 and least in YP_010415322.1. GRAVY is the least in QGQ59750.1 and has no considerable variation in the other two sequences. The isoelectric point is the least in QGQ59750.1 and nearly about the same in the other two sequences studied.

The various parameters studied show that the virus is polymorphic in different parts of the world and countries with different clads, it may be due to multiple mutations to survive in the prevailing climatic and other adverse conditions. The control measures should be adequately changed to achieve success in the control of the monkeypox virus in different parts of the world. The present studies show that there are two distinct genetic clades of the monkeypox virus; the Central African clade "(Congo Basin clade)", and "The West African Clade". The Central African clade is a more transmissible and severe disease impact[15]. The variation in the various parameters also suggests that the source of infection is also different in different parts of the world. The literature highlights that "African rodents, squirrels, mice, rats, dogs, and monkeys are the natural reservoir and cause infections in humans" [16].

The MSA (Fig.1) shows mutation occurred at different positions and that includes addition deletion and substitution. It is a clear indication of the virus's survival and natural selection by various mutations at different times as adaptability [17,18].

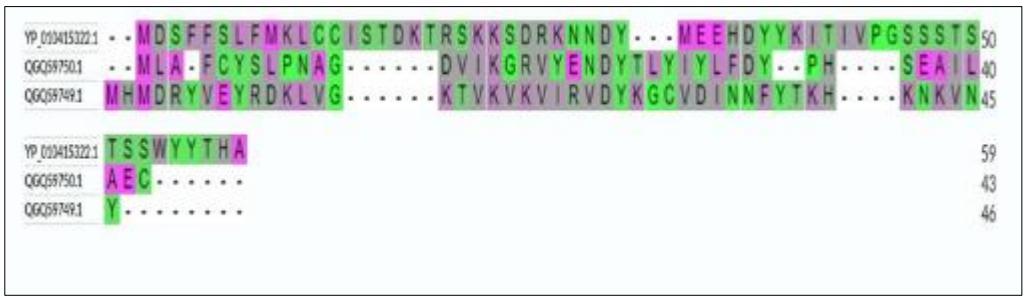

Figure 1 Multiple Sequence Alignment of Selected Hypothetical protein sequences

Presently Monkey viruses are quickly spread to different parts of the world.

The functional domain prediction (Fig-2-4) shows each hypothetical protein carries more than two domains an indication of different functions performed by the same protein for different purposes.

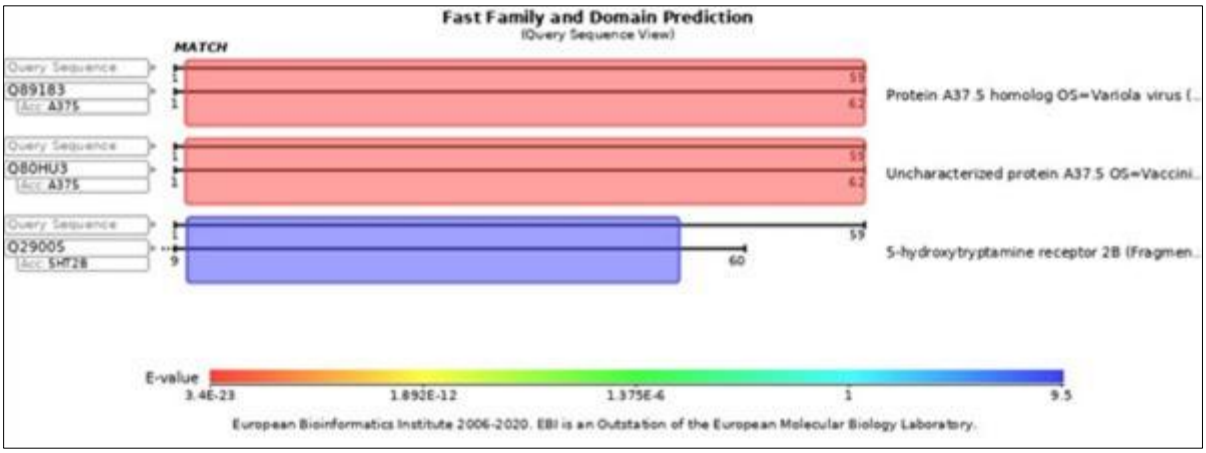

Figure 2 Functional domains for the sequence YP_010415322.1

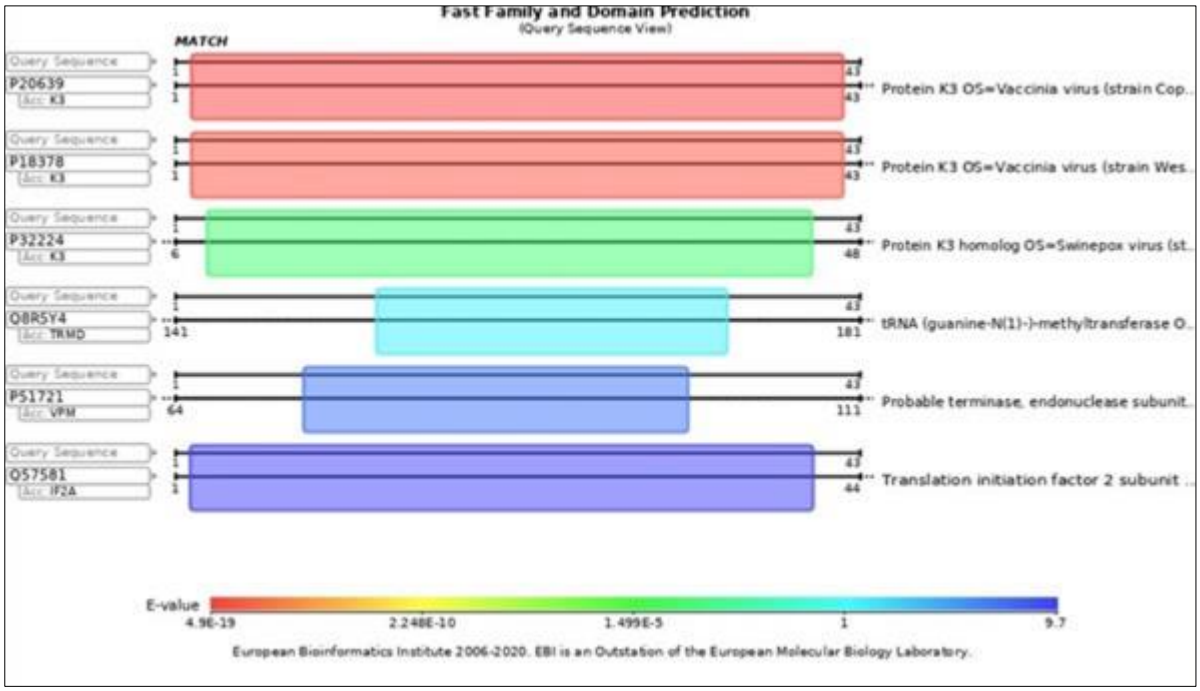

Figure 3 Functional domains for the sequence QGQ59750.1

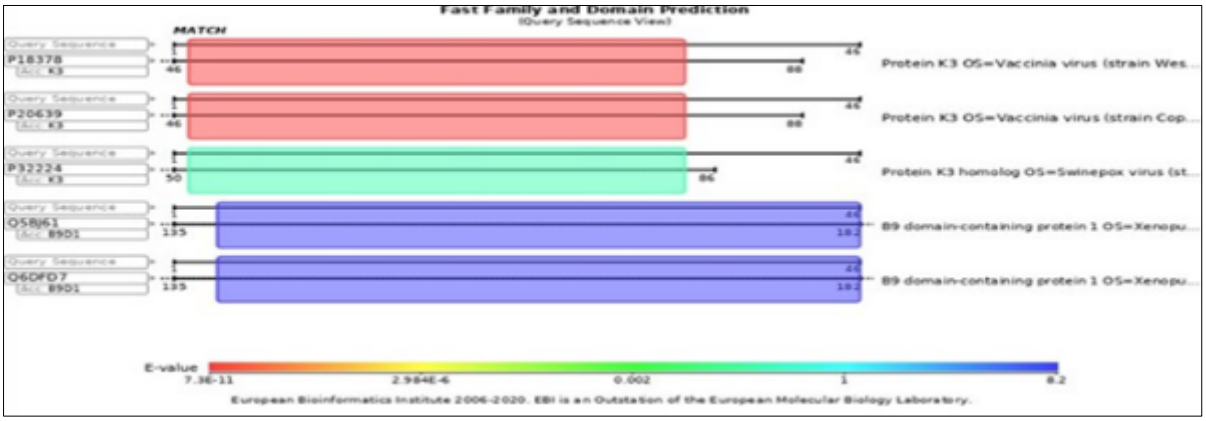

Figure 4 Functional domains for the sequence QGQ59749.1

The **amphipathicity** of the selected protein was predicted by Emboss-PepWheel (Fig5-7) and

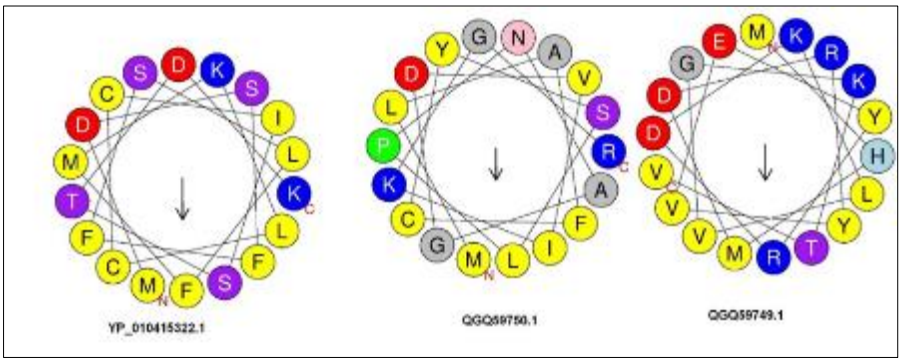

Figure 5 Amphipathicity predicted by PepWheel

Table 4 Amphipathicity Parameters of selected hypothetical protein sequences

Amphipathicity of Protein Residues Predicted by EMBOSS PepWheel			
Physico-chemical Properties	Protein sequences		
	YP_010415322.1	QGQ59750.1	QGQ59749.1
Hydrophobicity	0.707	0.548	0.219
Hydrophobic Moment	0.369	0.277	0.316
Net Charge	00	1.0	1.0
Polar residues+ Gly (n/%)	8/44.44	7/38.89	10/55.66
Uncharged Residues + Gly	Ser-3, Thr-1, Gly-0	Ser-1, Asn-1, Gly-2	His-1, Thr-1, Lys-1
Charged Residues	Ly-2, Asp-2	Lys-1, Arg-1, Asp-1	Lys-2, Arg-2, Glu=1, Asp-2
Hydrophobic phase	None	AFILMGC	None
No-polar Residues (n/%)	10/55.56	11/61.11	8/44.44
Aromatic Residues	Phe-3	Tyr-1, Phe-1	Tyr-2
Special Residues	Cys-2, Pro-0	Cys-1, Pro-1	Cys-0, Pro-0

PepWheel study shows that the sequence YP_010415322.1 and QGQ59750.1 is more stable due to the presence of cysteine. Polar residues are more or less equal in all three sequences but the M. virus with sequence QGQ59749.1 is more transient of lipophilic materials as it has more polar residues.

The overall study shows that a further study in correlation with various functional domains of the hypothetical protein sequences will help pharmacists to develop more efficient drug development against the monkeypox virus in an effective manner.

4 Conclusion

We conclude that all the three sequences are stable but QGQ59749.1 is less stable than other two due to the low cysteine content. The study also shows that all the proteins are positively charged and is substantiated by the other Physico-chemical parameters including instability index and half life of the proteins. This study may assist the scientists to develop some novel drugs to disrupt the stable protein envelop for the effective control of the Monkey pox virus and minimise the death toll due to this virus. A further study is needed to substantiate the findings to develop better control measures of the fatal virus.

Compliance with ethical standards

Acknowledgments

We immensely thanks to the principal Willingdon college and Miraj College for providing computer facilities to complete the works related to bioinformatics analysis using various tools and software. We also thanks to librarian of Willingdon college to complete necessary reference works to complete the post research writings.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Nguyen PY, Ajisegiri WS, Costantino V, Chughtai AA, MacIntyre CR. Reemergence of Human Monkeypox and Declining Population Immunity in the Context of Urbanization, Nigeria, 2017-2020. Emerg Infect Dis. 2021 Apr;27(4)

[2] Ladnyj ID, Ziegler P, Kima E. A human infection caused by the monkeypox virus in Basankusu Territory, Democratic Republic of the Congo. Bull World Health Organ. 1972; 46(5):593-7.

[3] Alakunle E, Moens U, Nchinda G, Okeke MI. Monkeypox Virus in Nigeria: Infection Biology, Epidemiology, and Evolution. Viruses. 2020 Nov 05; 12(11).

[4] Kugelman JR, Johnston SC, Mulembakani PM, Kisalu N, Lee MS, Koroleva G, McCarthy SE, Gestole MC, Wolfe ND, Fair JN, Schneider BS, Wright LL, Huggins J, Whitehouse CA, Wemakoy EO, Muyembe-Tamfum JJ, Hensley LE, Palacios GF, Rimoin AW. Genomic variability of monkeypox virus among humans, Democratic Republic of the Congo. Emerg Infect Dis. 2014 Feb; 20(2):232-9.

[5] Walsh D. Poxviruses: Slipping and sliding through transcription and translation. PLoS Pathog. 2017 Nov; 13(11):e1006634.

[6] Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005). pp. 571-607

[7] Sievers F., Wilm A., Dineen D., Gibson T.J., Karplus K., Li W., Lopez R., McWilliam H., Remmert M., Söding J., Thompson J.D., and Higgins D.G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7:539

[8] Dallago, C.; Karl, T.; Satagopam, V.; Heinzinger, M.; Littmann, M.; Olenyi, T.; Qiu, J.; Schütze, K.; Yachdav, G.; Ashkenazy, H.; Ben-Tal, N.; Bromberg, Y.; Goldberg, T.; Kajan, L.; O'Donoghue, S.; Sander, C.; Schafferhans, A.; Schlessinger, A.; Vriend, G.; Mirdita, M.; Gawron, P.; Gu, W.; Jarosz, Y.; Trefois, C.; Steinegger, M.; Schneider, R.; and Rost, B. Nucleic Acids Research, 49(W1): W535-W540. 05 2021.

[9] Ramachandran, G. N. and Sasisekharan, V. (1968) Conformation of polypeptides and proteins Advances in Protein Chemistry, 23. pp. 283-437. ISSN 0065-3233

[10] Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. SWISS-MODEL: homology modeling of protein structures and complexes. Nucleic Acids Res. 46(W1), W296-W303 (2018).

[11] Bienert, S., Waterhouse, A., de Beer, T.A.P., Tauriello, G., Studer, G., Bordoli, L., Schwede, T. The SWISS-MODEL Repository - new features and functionality. Nucleic Acids Res. 45, D313-D319 (2017).

[12] Gauex et al; N. Guex, M.C. Peitsch, T. Schwede Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective Electrophoresis, 30 (Suppl 1) (2009), pp. S162-S173

[13] Studer, G., Rempfer, C., Waterhouse, A.M., Gumienny, G., Haas, J., Schwede, T. QMEANDisCo - distance constraints applied on model quality estimation. Bioinformatics 36, 1765-1771 (2020).

[14] Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300:1005–1016

[15] World Health Organization (WHO). Multi-country monkeypox outbreak: situation update. Available at: https:// www.who.int/emergencies/disease-outbreak-news/ item/2022-DON390 Cited date June 29, 2022.

[16] Sklenovská N, VanRanst M. Emergence of Monkeypox as the Most Important Orthopoxvirus Infection in Humans. Front Public Health. 2018; 6:241. doi 10.3389/fpubh.2018.00241

[17] World Health Organization, Monkeypox Key Facts: Available at: https://www.who.int/news-room/fact-sheets/detail/ monkeypox. (Cited Date June 20, 2022).

[18] Centers for Disease Control and Prevention (CDC). Multistate outbreak of monkeypox--Illinois, Indiana, and Wisconsin, 2003. Morb Mortal Wkly Rep. 2003;52(23):537-540.