



# Microbes and Infectious Diseases

Journal homepage: <https://mid.journals.ekb.eg/>

## Original article

# In silico approach of exploring the structural and functional characteristics of a hypothetical protein from Monkeypox virus: A potential insight for antiviral therapeutics

Sadia Hossain Jeba<sup>1</sup>, Mohammad Arfatul Hasan<sup>1</sup>, A.S.M. Faisal<sup>1</sup>, Most. Ummay Salma Khatun<sup>2</sup>, Abrar Hossain Chowdhury, Nur Islam and Md. Abunasar Miah<sup>1\*</sup>

1- Department of Biotechnology and Genetic Engineering, Noakhali Science and Technology University, Noakhali-3814, Bangladesh.

2- Bangladesh Jute Research Institute (BJRI), Regional Station, Chandina, Comilla-3503, Bangladesh.

## ARTICLE INFO

### Article history:

Received 22 February 2024

Received in revised form 20 March 2024

Accepted 23 March 2024

### Keywords:

Annotation  
Hypothetical protein  
Kinase activity  
Monkeypox virus  
3D structure etc.

## ABSTRACT

**Background:** Monkeypox disease caused by the monkeypox virus is a major primary public concern worldwide, especially in developing countries. This virus contains several proteins are thought to be responsible for the onset of monkeypox disease. Among them, Many of these proteins are hypothetical, and their whose functions have not yet to be been discovered. A hypothetical protein [Monkeypox virus] (accession number URK21226.1) was selected in this study for extensive structural and functional analysis. Our study will predict the this protein's homological structure and biological function of this protein to acquire more knowledge about the protein. **Methods:** The physicochemical properties of the protein were determined using Protparam. Functional annotation tools such as NCBI-CD Search, Pfam, and InterProScan have predicted that the target protein has kinase activity. Multiple alignments and phylogenetic trees were built using CLC Sequence Viewer version 8.0. Then, the secondary structure was predicted by using PSIPRED. At last, the protein energy minimization and active site of the protein was were detected by respectively YASARA and (CASTp), respectively. **Results:** The multiple alignment of the homologous sequence was consistent with previous findings. Moreover, helices have been found to dominate in secondary structure. The protein's 3D structure of the protein was obtained using a homology modeling technique with a SWISS-MODEL server using a template protein with 98.53% sequence identity. The 3D structure was more stable after YESARA energy minimization and was validated by Procheck, Qmean, Verify3D, and Erratt. **Conclusions:** Protein has plays an essential role in nucleotide metabolism and It is a well-established therapeutic target against both diseases. Eventually, The research results will eventually provide an advantageous platform for future antiviral treatments.

## Introduction

Human monkeypox is a an *Orthopoxvirus* with a presentation similar to smallpox, which is a zoonotic disease. It has significant a major

similarity with identical to camelpox, cowpox, vaccinia, and variola viruses. It came out at the end of the eradication of the smallpox virus [1]. It was 1959 when the monkeypox (mpox) virus was first

DOI: 10.21608/MID.2024.272174.1818

\* Corresponding author: Md.Abunasar Miah

E-mail address: [nasar.bge@nstu.edu.bd](mailto:nasar.bge@nstu.edu.bd)

© 2020 The author (s). Published by Zagazig University. This is an open access article under the CC BY 4.0 license <https://creativecommons.org/licenses/by/4.0/>.

isolated and identified, and monkeys were shipped from Singapore to Denmark [2]. This is the reason behind naming this virus. However, compared to smallpox infection, case fatality rates and person-to-person dissemination beyond initial close contacts are significantly reduced in monkeypox.

In 1970, in the Democratic Republic of the Congo, a 9-month-old boy who was the only member of his family without a smallpox vaccine was diagnosed with the first instance of monkeypox infection in humans.<sup>3</sup> Monkeypox can be prevented to an estimated 85% of a degree by having had a smallpox vaccination in the past, while its long-term effectiveness is unknown. Following the discovery of the Rare outbreaks were documented after discovering the first human case, primarily mostly affecting young children in remote, forested parts of West and Central Africa were documented [3]. Monkeypox virus isolates were divided into two clades based on clinical manifestation and the results of genomic sequencing results. Several outbreaks of the monkeypox virus clade 1, which had significant mortality rates (1–12%), occurred in the Democratic Republic of the Congo between 1981 and 2017 [4]. Mathematical modeling in the setting of declining herd immunity to orthopoxviruses reveals an increased hazard of disease transfer between humans, even if the human-to-human transmission has historically been controlled. Transmission can occur by in various ways, like skin lesions, body fluids, and respiratory droplets. The Centers for Disease Control (CDC) and prevention (CDC) has recommended to isolate in a negative pressure room, contact and droplet precautions in the healthcare setting [2].

However, there are no approved therapies, or information on viral kinetics or the length of viral shedding. Two oral medications, brincidofovir and tecovirimat, have received approval for the treatment of smallpox and have shown effectiveness in animal tests against monkeypox [5]. Recently, WHO arranged a consultation from all over the world to manage the current outbreak of monkeypox. The main goal decided to betterment the future research, diagnostic tools, engagement with local communities, and also research on vaccines, and therapeutic agents [6].

Hypothetical proteins, whose functions are still unknown, pose a challenge not just to functional genomics but also to and general biology. For many

conserved proteins, the computational analysis provides only a general prediction of biochemical function; their exact biological functions have to be established through direct experimentation. Additionally, *in silico* analysis of this protein is important to for developing its vaccine or medication. It is also essential to understanding the genome structure and also the virus functional activity of the virus.

Thus, the main objective of this paper is to assess the potential biological function and homological structure prediction of the Hypothetical Protein from monkeypox virus. The FASTA sequence was analyzed using the updated bioinformatics software and tools for homology structure against functionally characterized proteins, determination of domain, physicochemical properties, and determination of active site. This interpretation will be helpful in order to strengthen our knowledge of the functional and molecular activity of the hypothetical protein (accession no UWO20788.1) found in the monkeypox virus. Further, we may also be able to find the potential pharmacological targets also.

## Materials and Methods

### Sequence retrieval using NCBI

Hypothetical protein MPXV-SI-2022V502223\_00170 [Monkeypox virus] with accession number of URK21226.1 containing 204 amino acid residues was selected for this study. The sequence was retrieved using the NCBI website (<http://www.ncbi.nlm.nih.gov/>). Subsequent analysis was done by retrieving the fasta format of the protein [7]. The fasta sequence is:

```
MSRGALIVFEGLDKSGKTTQCMNIM
ESIPANTIKYLNFPQRSTVTGKMIDDYLTRKK
TYNDHIVNLLFCANRWEFASFIQEQLQGITLI
VDRYAFSGVAYATAKGASMTLSKSYESGLP
KPDLVIFLESGSKEINRNIGEEIYEDVEFQQKV
LQEYKKMIEEGDIHWQIISSEFEEDVKKELIK
NIVIEAIHTVTGPVGQLWM
```

### Analysis of physicochemical properties:

We used The ExPASy ProtParam (<https://web.expasy.org/protparam/>) tool was used to characterize HPs for their HPs' physiological features. It includes molecular weight, aliphatic index, extinction coefficients, amino acid composition, grand average of hydrophobicity, isoelectric point, and estimated half-life.

**Function Prediction by analysis of domain and motif:** For domain analysis, NCBI

Conserved Domain Search Service (CD Search) (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) along with some other important essential tools like Pfam (<https://pfam.xfam.org/>) [8] and Interproscan [9] were used. A CD search identifies the conserved domains present in the protein sequence. By performing RPS-BLAST (Reverse Position-Specific BLAST), it compares the query sequence with the position-specific score matrix resulting from the alignment of the Conserved Domains that resides in the Conserved Domain Database (CDD). Pfam is a protein family database containing annotations and multiple sequence alignments generated using the Hidden Markov Model (HMM). Protein sequence motifs were analyzed using the MOTIF server (<http://www.genome.jp/tools/motif/>). CD-Search is NCBI's interface to searching the Conserved Domain Database with protein or nucleotide query sequences. It uses RPS-BLAST, a variant of PSI-BLAST, to quickly scan a set of pre-calculated position-specific scoring matrices (PSSMs) with a protein query. The results of CD-Search results are presented as an annotation of protein domains on the user query sequence [10] and can be visualized as domain multiple sequence alignments with embedded user queries.

### Multiple Sequence Alignment

A Blastp search (<https://blast.ncbi.nlm.nih.gov/Blast>) which was performed on the basis of based on a non-redundant database using default parameters to find protein homologues. Multiple alignments and phylogenetic trees were built using CLC Sequence Viewer version 8.0. Multiple alignment (MSA) is a tool used to identify evolutionary relationships and shared patterns between genes. More specifically, it refers to a sequence alignment of three or more biological sequences, usually DNA, RNA, or protein. Alignments are generated and analyzed using computational algorithms. Most MSA algorithms use a dynamic and heuristic approach. A seed tree is a phylogenetic tree that models the evolutionary history of a set of species (or populations). The gene tree is a phylogenetic tree that models the genealogy of genes. Due to various factors, the gene trees of different genes obtained from many species may not match each other, or the species tree may not match.

### Secondary Structure Determination

The Secondary structure of the monkeypox virus's selected hypothetical protein of the monkeypox virus was determined by using

PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>). PSIPRED Protein Analysis Workbench is a world-renowned web service that offering a diverse suite of protein prediction and annotation tools suite primarily focused on protein structural annotation.

### Homology Modelling

The SWISS-MODEL (<https://swissmodel.expasy.org/>) server was used to determine the 3D structure of the target protein based on homology modeling. The server automatically performs a BLASTp search to find potential templates for each protein sequence. Template protein 2v54.1. A was chosen for homology modeling from the search result with 98.53% sequence identity, which was a reliable score to start modeling; this is an X-ray diffraction model of the C-terminal domain of the vaccinia virus

### Quality check

The quality of the model structure and Ramachandran plot was evaluated using SWISS-MODEL workspace (<https://swissmodel.expasy.org/>) and ExPASy server programme procheck (<https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/>), QMEAN and Errat. The Ramachandran plot can be utilized for to assess the exactness of the anticipated protein structure. A few apparatuses and servers, such as the PDBsum server, MolProbity, STAN Server, and others, are utilized to produce the Ramachandran plots. The Spares v6.0 could be a comprehensive toolkit and has and with five instruments (ERRAT, VERIFY3D, Demonstrate, PROCHECK, and WHAT CHECK), which foresee distinctive sorts of stereochemical parameters of the protein structure.

The PROCHECK programs are useful for assessing the quality not only of protein structures in the process of that are being solved, but also existing structures and of those being modelled on known structures. Verify3D profiles are useful in the evaluating of undetermined protein models based on low-resolution electron-density maps, on NMR spectra with inadequate distance constraints, or on computational procedures. QMEAN evaluates several structural features of proteins. The absolute quality estimate of a model is expressed in terms of how well the model score agrees with the expected values from a representative set of high-resolution experimental structures. The resulting QMEAN Z-score is a measure of the 'degree of nativeness' of a given protein structure [11].

### Active site determination

(CASTp) (<http://sts.bioe.uic.edu/castp/>) was used to determine the active site of the protein. The topographical features of proteins are comprehensively and quantitatively recorded in detail by CASTp [12]. Active pockets located on the protein surface and within 3D structures can be accurately placed and measured. Therefore, it has become an essential platform for predicting regions of proteins that interact with ligands and critical residues. CASTp results were also visualized by using PyMOL software.

### Result

#### Analysis and determination of physicochemical properties of selected protein

Analysis of the physicochemical properties of the selected strain of the monkeypox virus hypothetical protein MPXV-SI-2022V502225\_00170 [Monkeypox virus] were estimated by the ProtParam tool shown in table 1. The protein was predicted to contain 204 amino acids, possess a molecular weight of 23290.69, a theoretical pI of 5.11, and a grand average of hydropathicity (GRAVY) of  $-0.243$ . The instability index (II) of the target protein was predicted to be 52.04, classifying the protein as unstable. The formula is Formula: C1046H1644N266O316S9. The amino acids are alanine (4.9%), arginine (2.9%), asparagine (4.4%), aspartic acid (4.4%), cystine (1.0%), glutamine (4.9%), glutamic acid (9.8%), histidine (1.5%), isoleucine (10.3%), leucine (7.4%), lysine (7.8%), methionine (3.4%), phenylalanine (4.4%), proline (2.5%), serine (6.9%), threonine (5.9%), tryptophan (3.9%), valine (5.9%).

#### Functional prediction by domain and motif analysis

The target protein was predicted by the NCBI-CD Search, Pfam, and InterProScan predicated the target protein to have the Chordopox\_A20R superfamily domain name Thymidylate kinase classification as a Chordopox 9-181 amino acid content. The target protein was predicted by The NCBI-CD Search, Pfam, and InterProScan predicated the target protein to have the Chordopox\_A20R superfamily domain with a  $2.30e-60$  classification as a Chordopox\_A20R domain-containing protein. It is a nucleotide kinase (NK). The multiprotein DNA replication complex's formation or stability may be influenced by A20R  $2.30e-60$ . The result of the conserved CD search is

shown in the table 2. Moreover, the interproScan result is arranged in the table 3.

#### Multiple sequence alignment and phylogeny analysis

Sequence similarity with other known DNA polymerase processivity factor protein families of the monkeypox virus was detected in the non-redundant database, with up to 99.51% similarity. In order To examine the conserved and dissimilar residue among the homologs, Multiple Sequence Alignments (MSA) of the top ten (10) chosen proteins were performed on the BLASTp findings to examine the conserved and dissimilar residue among the homologs,. The Blastp result is incorporated in the table 4. Multiple sequence alignment was created by The CLC sequence viewer created Multiple sequence alignments, and we got the phylogenetic tree from Cluster Omega.

#### Secondary structure prediction

PSIPRED denoted the protein's secondary structure. The alpha helix was identified as being the most prevalent, followed by extended strand, random coil, and beta-turn, according to the SOPMA estimate. PSIPRED also produced comparable findings (alpha helix, random coil, and extended strand). Fig. depicts the protein's secondary structure as predicted by PSIPRED.

#### Tertiary structure prediction and energy minimization

The target protein's tertiary structure was extracted from the SWISS-MODEL service using the 2v54.1.A, which shares 98.53% of its sequence with the source protein. Employing the YASARA force field minimizer, the energy of the protein's three-dimensional structure was reduced. Energy minimization resulted in a reduction in energy to  $-264427.2$  kJ/g from  $-205229.5$  KJ/g. The original value was, but following the minimizing procedure, the final value was, indicating a stable structure.

#### Quality assessment

PROCHECK, verify 3D, QMEAN, and ERRAT programs were used to evaluate the quality of the modeled 3D structure. The PROCHECK result showed that 94% of amino acid residues fell into the "Ramachandran plot's" most liked area, indicating an excellent stereochemical quality of the protein model. The result is incorporated in the table 5. The model structure was approved by the Verify

3D server, with an averaged 3D-1D score  $\geq 0.1$  for 74.94% of the residues. The model was placed extremely near to the dark gray zone by the QMEAN tool, with a projected QMEAN4 value of 0.50, which is regarded as reasonably excellent. ERRATA predicted the protein structure to be excellent, with a quality factor 97.6864. It also predicted the protein structure to be of good quality with a quality factor of 98.19%. the z score is Z score is -6.54.

The quality of the protein is good enough for deep study.

#### Active site analysis

The generated 3D structure's active site was evaluated utilizing the CASTp server. The surface of solvent-accessible molecules (Richard surface) and molecular surfaces (Connolly surface) are employed to define a pocket and volume spectrum or vacuum. Here, the most active site was

determined in one of the largest pockets, having a 1069.032A<sup>2</sup> Solvent Accessible (SA) surface area and a total volume of 630.865 A<sup>3</sup> amino acids (Fig. 7). The key active residues predicted from pockets are ASP13, PHE<sup>38</sup>, PRO<sup>39</sup>, ARG<sup>41</sup>, ILE<sup>49</sup>, ASP<sup>50</sup>, LEU<sup>53</sup>, THR<sup>54</sup>, and HIS62.

#### Energy minimizations

Energy minimization of protein was conducted by YESERA software [13]. Energy minimization is a key crucial step in finding a protein movement search space to figure out a configuration that is stable and minimized locally configuration. In order To being dynamics, energy minimization ensures that the sustem system has no steric clashes or inappropriate structure. During this process, the geometry is relaxed. The Energy Minimization Server minimized the energy of the model protein structure from -205229.5 to -264427.2 kJ/g.

**Table 1.** Physicochemical properties of the HP estimated by ProtParam tool.

Description	Value
No. of amino acid	204
Molecular weight (Da)	23290.69
Theoretical PI	5.11
No. of positively charged residue	22
No. of negatively charged residue	29
No. of atoms	3281
Instability Index	52.04
Aliphatic Index	90.78
Grand average of hydropathicity	-0.243

**Table 2.** Results from NCBI Conserved CD Searches.

Name	Description	Interval	E-value
NK super family	Nucleoside/nucleotide kinase (NK) is a protein superfamily consisting of multiple families of enzymes that share structural similarity and are functionally related to the catalysis of the reversible phosphate group. Members of this family play a wide variety of essential roles in nucleotide metabolism as well as the metabolism of sugar and sulfate.	2-203	2.30e-60

**Table 3:** Result From Interproscan.

Function	Description
Dtdp Biosynthetic Process	The Chemical Reactions And Pathways Resulting In The Formation Of Dtdp, Deoxyribosylthymine Diphosphate (2'-Deoxyribosylthymine5'-Diphosphate). <sup>1</sup>
Dttp Biosynthetic Process	The Chemical Reactions And Pathways Resulting In The Formation Of Dttp, Deoxyribosylthymine Triphosphate. <sup>2</sup>
ATP Binding	Binding To ATP, Adenosine 5'-Triphosphate, A Universally Important Coenzyme And Enzyme Regulator. <sup>3</sup>
Thymidylate Kinase Activity	Catalysis Of The Reaction: ATP + Thymidine 5'-Phosphate = ADP + Thymidine 5'-Diphosphate. <sup>4</sup>
Nucleoside Diphosphate Kinase Activity	Catalysis Of The Reaction: ATP + Nucleoside Diphosphate = ADP + Nucleoside Triphosphate. <sup>5</sup>

**Table 4.** Blastp results showing similarity between proteins.

Protein name	Score	Percent Identity	E Value
MPXVgp159 [Monkeypox virus]	423	100%	2e-149
A49R [Monkeypox virus Zaire-96-I-16]	422	99.51%	3e-149
thymidylate kinase [Cowpox virus]	419	100%	1e-147
thymidylate kinase [Cowpox virus]	419	100%	1e-147
Putative A48 Rprotein [Orthopox virus Abatino]	419	100%	2e-147
thymidylate kinase [Vaccinia virus]	419	100%	2e-147
Cowpox virus [Brighton Red]	419	100%	3e-147
thymidylate kinase [Cowpox virus]	417	100%	3e-147
CPXV186 protein [Cowpox virus]	417	100%	3e-147
Rabbitpox virus Utrecht	418	100%	3e-147

**Table 5:** Ramachandran Plot result.

Statistics	Number of A.A residues	Percentage (%)
Residues in the most favored regions [ A, B, L]	345	94.0%
Residues in the additional allowed regions [a,b, I,p]	20	5.4%
Residues in the generously allowed regions [ ~a, ~b, ~I, ~P]	0	0.0%
Residues in disallowed regions	2	0.5%
		Total: 100%
Number of nonglycine and nonproline residues	367	
Number of end-residues (excl. Gly and Pro)	6	
Number of end-residues (excl. Gly and Pro)	26	
Number of proline residues	10	
Total number of residues	409	

Figure 1. Multiple sequence alignment and phylogenetic tree.

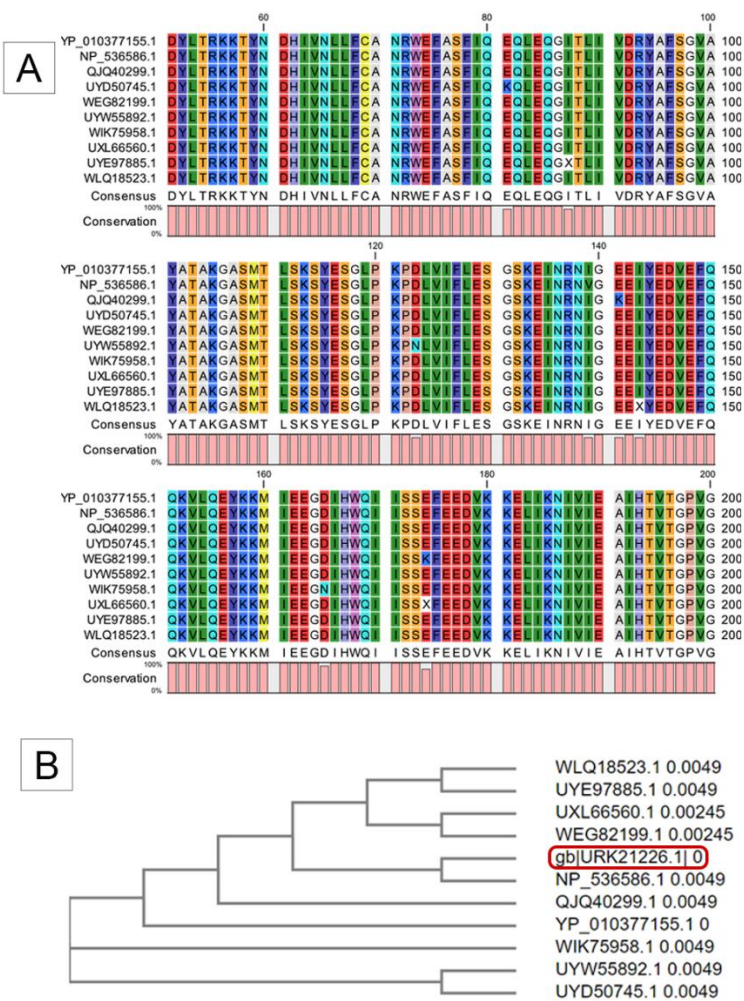
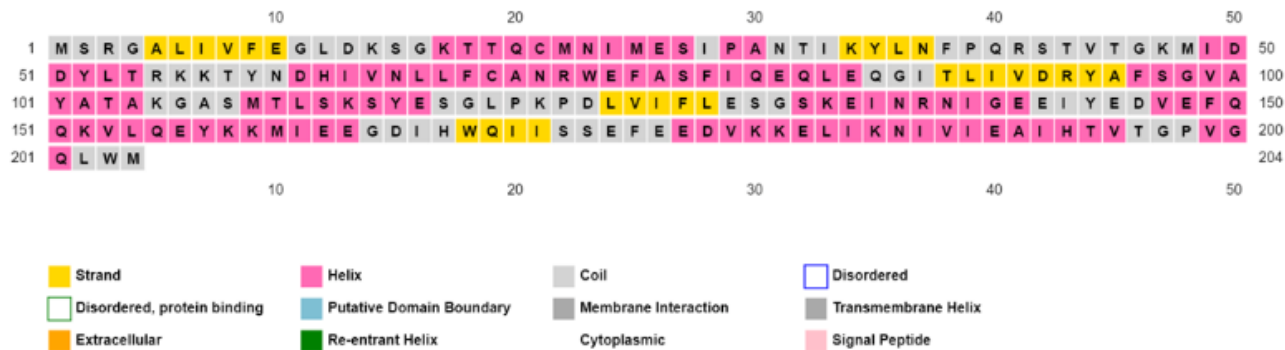


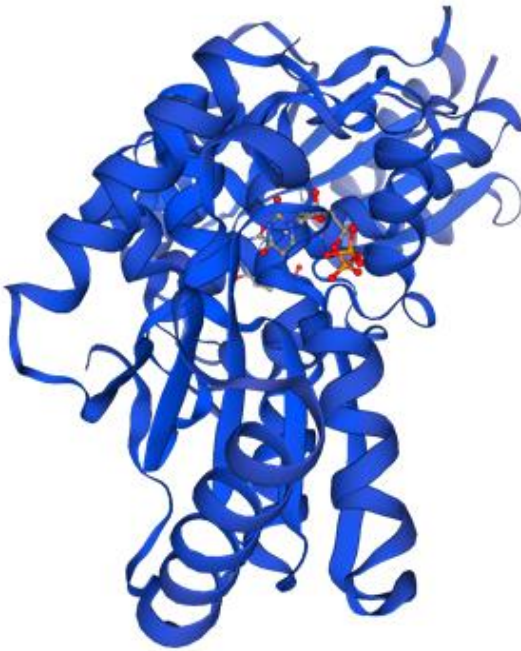
Figure 1A: Multiple Sequence Alignment using CLC Sequence Viewer.  
Figure 1B: Phylogenetic tree created by Clustal Omega.

Figure 2. Secondary structure prediction of the Hypothetical targeted protein using PSI-PRED server. Yellow and pink colors are indicating different parts of the Hypothetical protein.

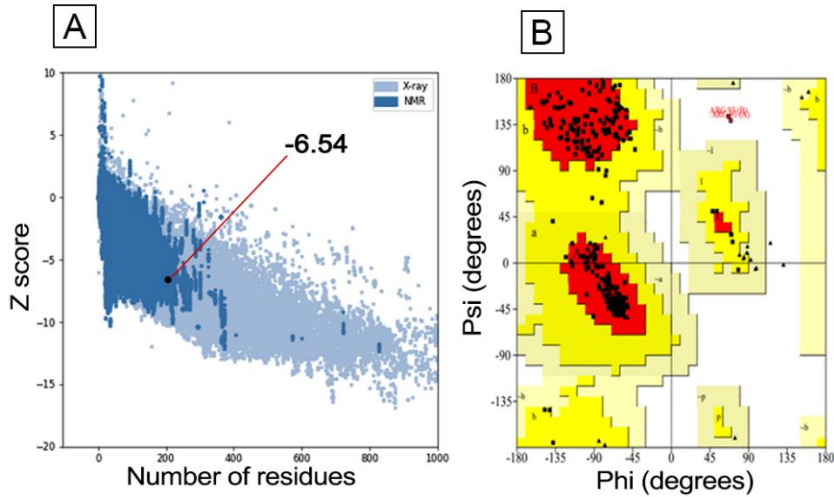




**Figure 3.** 3D Structure determination using Swiss Model.

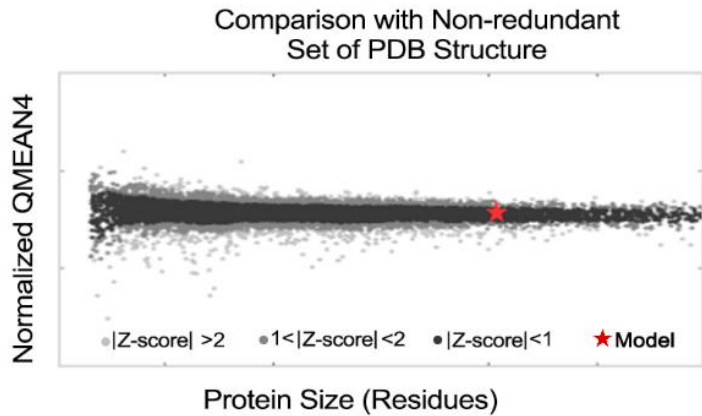


**Figure 4.** Quality Check.



**Figure 4A:** Z scores of the template protein using ProSA server.  
: **Figure 4B** Ramachandran Plot analysis.

**Figure 5.** Determination of QMEAN Z- Scores of Selected Protein.

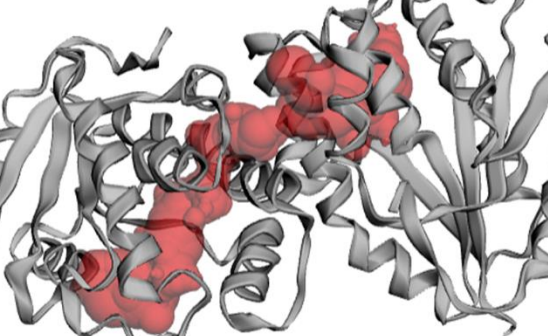




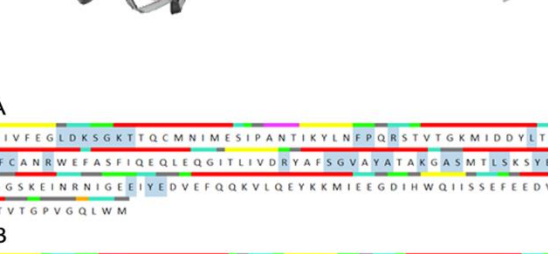
**SYSTEM PROPERTIES**

Number: 1  
 Element: 1  
 Assembly: 1  
 B-factor: 0.03  
 Position: 1  
 System: 1  
 Type: 1  
 Unit: 1  
 Value: 1  
 Weight: 1  
 X-axis: 1  
 Y-axis: 1  
 Z-axis: 1  
 X2-axis: 1  
 Y2-axis: 1  
 Z2-axis: 1  
 X3-axis: 1  
 Y3-axis: 1  
 Z3-axis: 1  
 X4-axis: 1  
 Y4-axis: 1  
 Z4-axis: 1  
 X5-axis: 1  
 Y5-axis: 1  
 Z5-axis: 1  
 X6-axis: 1  
 Y6-axis: 1  
 Z6-axis: 1  
 X7-axis: 1  
 Y7-axis: 1  
 Z7-axis: 1  
 X8-axis: 1  
 Y8-axis: 1  
 Z8-axis: 1  
 X9-axis: 1  
 Y9-axis: 1  
 Z9-axis: 1  
 X10-axis: 1  
 Y10-axis: 1  
 Z10-axis: 1  
 X11-axis: 1  
 Y11-axis: 1  
 Z11-axis: 1  
 X12-axis: 1  
 Y12-axis: 1  
 Z12-axis: 1  
 X13-axis: 1  
 Y13-axis: 1  
 Z13-axis: 1  
 X14-axis: 1  
 Y14-axis: 1  
 Z14-axis: 1  
 X15-axis: 1  
 Y15-axis: 1  
 Z15-axis: 1  
 X16-axis: 1  
 Y16-axis: 1  
 Z16-axis: 1  
 X17-axis: 1  
 Y17-axis: 1  
 Z17-axis: 1  
 X18-axis: 1  
 Y18-axis: 1  
 Z18-axis: 1  
 X19-axis: 1  
 Y19-axis: 1  
 Z19-axis: 1  
 X20-axis: 1  
 Y20-axis: 1  
 Z20-axis: 1  
 X21-axis: 1  
 Y21-axis: 1  
 Z21-axis: 1  
 X22-axis: 1  
 Y22-axis: 1  
 Z22-axis: 1  
 X23-axis: 1  
 Y23-axis: 1  
 Z23-axis: 1  
 X24-axis: 1  
 Y24-axis: 1  
 Z24-axis: 1  
 X25-axis: 1  
 Y25-axis: 1  
 Z25-axis: 1  
 X26-axis: 1  
 Y26-axis: 1  
 Z26-axis: 1  
 X27-axis: 1  
 Y27-axis: 1  
 Z27-axis: 1  
 X28-axis: 1  
 Y28-axis: 1  
 Z28-axis: 1  
 X29-axis: 1  
 Y29-axis: 1  
 Z29-axis: 1  
 X30-axis: 1  
 Y30-axis: 1  
 Z30-axis: 1  
 X31-axis: 1  
 Y31-axis: 1  
 Z31-axis: 1  
 X32-axis: 1  
 Y32-axis: 1  
 Z32-axis: 1  
 X33-axis: 1  
 Y33-axis: 1  
 Z33-axis: 1  
 X34-axis: 1  
 Y34-axis: 1  
 Z34-axis: 1  
 X35-axis: 1  
 Y35-axis: 1  
 Z35-axis: 1  
 X36-axis: 1  
 Y36-axis: 1  
 Z36-axis: 1  
 X37-axis: 1  
 Y37-axis: 1  
 Z37-axis: 1  
 X38-axis: 1  
 Y38-axis: 1  
 Z38-axis: 1  
 X39-axis: 1  
 Y39-axis: 1  
 Z39-axis: 1  
 X40-axis: 1  
 Y40-axis: 1  
 Z40-axis: 1  
 X41-axis: 1  
 Y41-axis: 1  
 Z41-axis: 1  
 X42-axis: 1  
 Y42-axis: 1  
 Z42-axis: 1  
 X43-axis: 1  
 Y43-axis: 1  
 Z43-axis: 1  
 X44-axis: 1  
 Y44-axis: 1  
 Z44-axis: 1  
 X45-axis: 1  
 Y45-axis: 1  
 Z45-axis: 1  
 X46-axis: 1  
 Y46-axis: 1  
 Z46-axis: 1  
 X47-axis: 1  
 Y47-axis: 1  
 Z47-axis: 1  
 X48-axis: 1  
 Y48-axis: 1  
 Z48-axis: 1  
 X49-axis: 1  
 Y49-axis: 1  
 Z49-axis: 1  
 X50-axis: 1  
 Y50-axis: 1  
 Z50-axis: 1  
 X51-axis: 1  
 Y51-axis: 1  
 Z51-axis: 1  
 X52-axis: 1  
 Y52-axis: 1  
 Z52-axis: 1  
 X53-axis: 1  
 Y53-axis: 1  
 Z53-axis: 1  
 X54-axis: 1  
 Y54-axis: 1  
 Z54-axis: 1  
 X55-axis: 1  
 Y55-axis: 1  
 Z55-axis: 1  
 X56-axis: 1  
 Y56-axis: 1  
 Z56-axis: 1  
 X57-axis: 1  
 Y57-axis: 1  
 Z57-axis: 1  
 X58-axis: 1  
 Y58-axis: 1  
 Z58-axis: 1  
 X59-axis: 1  
 Y59-axis: 1  
 Z59-axis: 1  
 X60-axis: 1  
 Y60-axis: 1  
 Z60-axis: 1  
 X61-axis: 1  
 Y61-axis: 1  
 Z61-axis: 1  
 X62-axis: 1  
 Y62-axis: 1  
 Z62-axis: 1  
 X63-axis: 1  
 Y63-axis: 1  
 Z63-axis: 1  
 X64-axis: 1  
 Y64-axis: 1  
 Z64-axis: 1  
 X65-axis: 1  
 Y65-axis: 1  
 Z65-axis: 1  
 X66-axis: 1  
 Y66-axis: 1  
 Z66-axis: 1  
 X67-axis: 1  
 Y67-axis: 1  
 Z67-axis: 1  
 X68-axis: 1  
 Y68-axis: 1  
 Z68-axis: 1  
 X69-axis: 1  
 Y69-axis: 1  
 Z69-axis: 1  
 X70-axis: 1  
 Y70-axis: 1  
 Z70-axis: 1  
 X71-axis: 1  
 Y71-axis: 1  
 Z71-axis: 1  
 X72-axis: 1  
 Y72-axis: 1  
 Z72-axis: 1  
 X73-axis: 1  
 Y73-axis: 1  
 Z73-axis: 1  
 X74-axis: 1  
 Y74-axis: 1  
 Z74-axis: 1  
 X75-axis: 1  
 Y75-axis: 1  
 Z75-axis: 1  
 X76-axis: 1  
 Y76-axis: 1  
 Z76-axis: 1  
 X77-axis: 1  
 Y77-axis: 1  
 Z77-axis: 1  
 X78-axis: 1  
 Y78-axis: 1  
 Z78-axis: 1  
 X79-axis: 1  
 Y79-axis: 1  
 Z79-axis: 1  
 X80-axis: 1  
 Y80-axis: 1  
 Z80-axis: 1  
 X81-axis: 1  
 Y81-axis: 1  
 Z81-axis: 1  
 X82-axis: 1  
 Y82-axis: 1  
 Z82-axis: 1  
 X83-axis: 1  
 Y83-axis: 1  
 Z83-axis: 1  
 X84-axis: 1  
 Y84-axis: 1  
 Z84-axis: 1  
 X85-axis: 1  
 Y85-axis: 1  
 Z85-axis: 1  
 X86-axis: 1  
 Y86-axis: 1  
 Z86-axis: 1  
 X87-axis: 1  
 Y87-axis: 1  
 Z87-axis: 1  
 X88-axis: 1  
 Y88-axis: 1  
 Z88-axis: 1  
 X89-axis: 1  
 Y89-axis: 1  
 Z89-axis: 1  
 X90-axis: 1  
 Y90-axis: 1  
 Z90-axis: 1  
 X91-axis: 1  
 Y91-axis: 1  
 Z91-axis: 1  
 X92-axis: 1  
 Y92-axis: 1  
 Z92-axis: 1  
 X93-axis: 1  
 Y93-axis: 1  
 Z93-axis: 1  
 X94-axis: 1  
 Y94-axis: 1  
 Z94-axis: 1  
 X95-axis: 1  
 Y95-axis: 1  
 Z95-axis: 1  
 X96-axis: 1  
 Y96-axis: 1  
 Z96-axis: 1  
 X97-axis: 1  
 Y97-axis: 1  
 Z97-axis: 1  
 X98-axis: 1  
 Y98-axis: 1  
 Z98-axis: 1  
 X99-axis: 1  
 Y99-axis: 1  
 Z99-axis: 1  
 X100-axis: 1  
 Y100-axis: 1  
 Z100-axis: 1  
 X101-axis: 1  
 Y101-axis: 1  
 Z101-axis: 1  
 X102-axis: 1  
 Y102-axis: 1  
 Z102-axis: 1  
 X103-axis: 1  
 Y103-axis: 1  
 Z103-axis: 1  
 X104-axis: 1  
 Y104-axis: 1  
 Z104-axis: 1  
 X105-axis: 1  
 Y105-axis: 1  
 Z105-axis: 1  
 X106-axis: 1  
 Y106-axis: 1  
 Z106-axis: 1  
 X107-axis: 1  
 Y107-axis: 1  
 Z107-axis: 1  
 X108-axis: 1  
 Y108-axis: 1  
 Z108-axis: 1  
 X109-axis: 1  
 Y109-axis: 1  
 Z109-axis: 1  
 X110-axis: 1  
 Y110-axis: 1  
 Z110-axis: 1  
 X111-axis: 1  
 Y111-axis: 1  
 Z111-axis: 1  
 X112-axis: 1  
 Y112-axis: 1  
 Z112-axis: 1  
 X113-axis: 1  
 Y113-axis: 1  
 Z113-axis: 1  
 X114-axis: 1  
 Y114-axis: 1  
 Z114-axis: 1  
 X115-axis:

**A**



**B**



**Chain A**

MSRGALIVFEGLDKSGKTTQCMNIMESIPANTIKYLNFPQRSTVTGKMIDDYLTRKKTYND  
 HIVNLLFCANRWEFASFQIQLEQGGITLIVDRYAFSGVAYATAKGASMTLSKSYESGLPKPD  
 LVIFLESQSKENRRNIGEEIYEDVEFQQKVLQEYKKMIEEGDIHWQIISSEFEEDVKKELIKN  
 VIEAHTVTGPGVGLWM

**Chain B**

SRGALIVFEGLDKSGKTTQCMNIMESIPANTIKYLNFPQRSTVTGKMIDDYLTRKKTYNDH  
 VNNLLFCANRWEFASFQIQLEQGGITLIVDRYAFSGVAYATAKGASMTLSKSYESGLPKPDV  
 IFLESQSKENRRNIGEEIYEDVEFQQKVLQEYKKMIEEGDIHWQIISSEFEEDVKKELIKNIV  
 FAHTVTGPGVGLWM

**Figure 7B:** Information on Amino acid for HP. Red-lined sections are the active region of chain A and Chain B.

During the July 19, 2023 update, the U.S. death count was updated from 44 to 45 to reflect a monkeypox death that occurred in March 2023 [14]. The first human case of monkeypox was recorded in 1970. Prior to the 2022 outbreak, mpox monkeypox had been reported in people in several central and western African countries [15]. Unfortunately, it is getting more severe day by day.

we investigated the hypothetical protein accession no. UWO30788.1 of monkeypox. Our aim is The main objective of this study is to get a deep knowledge of the structural and biological function of this HP (hypothetical protein) through an in-silico approach. It may help in understanding the mechanism of bacterial pathogenicity and help us with some ways of prevention. Various computational resources and strategies were applied in this paper to characterize HP URK21226.1.1. This protein is a larger molecule containing 204 amino acids, and its molecular weight is around

23290.69Da. [Table 1]. The gravity value is -0.243. This value indicates that the protein contains a non-polar amino acid. Here, the height highest amount of protein is glutamic acid (9.8%), which is a hydrophobic amino acid. Isoleucine is an isomer of leucine that contains two chiral carbon atoms. Proline is unique among the standard amino acids because it has neither free  $\alpha$ -amino nor free  $\alpha$ -carboxyl groups [16]. This feature made the whole protein hydrophobic.

Domain and motif analysis indicated that the HP belongs to the superfamily Chordopoxviridae and in the family named poxviridae based on the Pfam and InterProScan results. Viruses of the Poxviridae family are the largest and more complex viruses infecting animals. A common feature of chordopoxviruses is that, during host infection, of the host, they sometimes they colonize, replicate, and induce pathology in the skin and, to a lesser extent, in selected mucosae. Contact of infectious material with broken or lacerated skin is a common route of poxvirus transmission. The most well-known poxviruses are species of the genus Orthopoxvirus [17, 18].

Identifying mistakes in experimental and theoretical models of protein structures is a critical issue in structural biology. The ProSA program (Protein Structure Analysis) is an established tool that has with a large user base and is frequently employed in the refinement and also validation of hypothetical protein structures and in prediction and modeling. In particular, the quality scores of a protein are displayed in the context of all known protein structures, and problematic parts of a structure are shown and highlighted in a 3D molecule viewer. The structure is evaluated using a distance-based pair potential and a potential that captures the solvent exposure of protein residues. From these energies, Two characteristics of the input structure from these energies are taken and displayed in this study. The z-score and plot are its residue energies [19, 20]. The z-score determines the overall model quality (figure 5). Its value is displayed that the z-score of all experimentally determined protein chains in the current HP (URK21226.1.1). The score we got was around -6.58. [online link] ProSA-web z-scores of all protein chains in PDB determined by X-ray crystallography (light blue) or NMR spectroscopy (dark blue) with respect to connecting their length. Normally Usually found for native proteins of similar size. The other plot (figure 4A) shows local model quality by

plotting energies as a function of amino acid sequence position. In general, positive values correspond to problematic or erroneous parts of the input structure [19-21].

Computer Atlas of Surface Topography of Proteins (CASTp) aims to provide a comprehensive and detailed quantitative characterization of the topographic features of the protein. Usually, the enzyme subunit has one active site that is capable of binding site. It depends on the subsequent amino acids, which the determine protein's shape the shape of the protein (i.e., the structure of the active site) and, hence, the specificity of that protein.<sup>22</sup>

The most active site is visualized in Figure 7 (a), and the predominant amino acid is represented in Figure 7(b) represents the predominant amino acid. Meanwhile, the protein-protein interaction network explored some surrounding proteins mostly that could not be characterized. Its pocket area is a 1069.032A<sup>2</sup> Solvent Accessible (SA) surface area, and a total volume of 630.865 A<sup>3</sup> amino acids, and identity is 98.77%, which means meaning the template is near to the target template. It occupies a good area of negative volume. The negative volume is the space encompassed by the atoms that form these geometric and topological features, such as pockets, cavities, and channels, which were shown only through the representation of surface atoms during formation [12, 23]. SWISS-MODEL server successfully passed all of the model quality assessment tools like PROCHECK, QMEAN, and ERRAT. Unfortunately, the VARIFY 3D result came failed. Because 74.94% of the residues have averaged 3D-1D score  $\geq 0.1$ . The overall quality factor of ERRAT is 97.68. The Ramachandran plot also carries an impactful result also, which is as shown in the Figure 4(b). It can be mentioned as a good-quality of structure [24].

## Conclusion

Pathogens harboring the Chordopox\_A20R protein pose a significant threat to human health, making understanding its functions imperative. While considerable progress has been made in recent years towards elucidating the roles of the Chordopoxvirus\_A20R protein, numerous structural and functional aspects and its effectors remain shrouded in mystery. Our study represents a pioneering effort in characterizing the Chordopoxvirus\_A20R protein from the Monkeypox virus, delving into its structural and functional attributes. The annotation of this

hypothetical protein holds promise for informing the development of effective drugs or vaccines. By unraveling this protein's mysteries, we aim to shed light on potential targets for antiviral interventions. However, it's important to note that our findings represent just the beginning; further research and experimental validations are essential to solidify our understanding of this pivotal protein. We emphasize the critical need for continued investigation into Chordopox\_A20R and its effectors, not only within the context of monkeypox but also across other pathogenic microorganisms. This information is very essential for viral metabolic pathways, disease progression, drug development, and disease control strategies, as they provide insights into the mechanisms underlying pathogenicity and potential avenues for therapeutic intervention. Through sustained research endeavors, we aspire to pave the way toward more effective approaches for combating infectious diseases caused by pathogens housing the Chordopox\_A20R protein. Further research works are needed to find out and confirm our study about crucial protein

#### Disclosure of potential conflicts of interest

The author receives no financial support from any institute or organization.

#### Contributors

- Dr. Md. Abunasar Miah planned, designed and guided this manuscript.
- Sadia Hossain Jeba, Mohammad Arfatul Hasan, Abrar Hossain Chowdhury and
- A.S.M. Faisal have performed the entire manuscript.
- Nur Islam and Most. Ummay Salma Khatun has revised the manuscript thoroughly.

The first and the last authors mainly drafted the manuscript and all authors equally contributed to edit the language.

#### References

- 1- McCollum AM, Damon IK. Human Monkeypox. *Clin Infect Dis*. 2014;58(2):260-267. doi:10.1093/cid/cit703
- 2- Moore MJ, Rathish B, Zahra F. Mpox(Monkeypox). In: StatPearls. Treasure Island (FL): StatPearls Publishing; May 3, 2023.
- 3- Mitja O, Ogoina D, Titanji BK, Galvan C, Muyembe JJ, Marks M, et al. Monkeypox. *The Lancet*. 2022;401(10370). doi:10.1016/S0140-6736(22)02075-X
- 4- Mukherjee D, Roy S, Singh V, Gopinath S, Pokhrel NB, Jaiswal V. Monkeypox an emerging global health threat during the COVID-19 time. *Ann Med Surg (Lond)*. 2022;79: 104075. doi:10.1016/j.amsu.2022.104075
- 5- Adler H, Gould S, Hine P, Snell LB, Wong W, Houlihan CF, et al. Clinical features and management of human monkeypox: a retrospective observational study in the UK. *Lancet Infect Dis*. 2022;22. doi:10.1016/S1473-3099(22)00228-6
- 6- The Lancet Infectious Diseases null. Monkeypox: a neglected old foe. *Lancet Infect Dis*. 2022; 22(7):913. doi:10.1016/S1473-3099(22)00377-2
- 7- Binz PA, Shofstahl J, Vizcaino JA, Barsnes H, Chalkley RJ, Menschaert G, et al. Proteomics standards initiative extended FASTA format. *J proteome Res*. 2019;18(6):2686-2692. doi:10.1021/acs.jproteome.9b00064
- 8- Bateman A. The Pfam protein families database. *Nucleic Acid Res*. 2004;32(90001):138D-141. doi:10.1093/nar/gkh121
- 9- Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterproScan 5: genome-scale protein function classification. *Bioinforma Oxf Engl*. 2014; 30(9):1236-1240. doi:10.1093/bioinformatics/btu031
- 10- Marchler-Bauer A, Bryant SH, CD-Search: protein domain annotations on the fly. *Nucleic Acids Res*. 2004; 32: W327-W331. doi: 10.1093/nar/gkh454
- 11- Walker SD, McEldowney S, Molecular docking: a potential tool to aid ecotoxicity

- testing in environment risk assessment of pharmaceuticals. *Chemosphere*. 2013;93(10):2568-2577. doi:10.1016/j.chemosphere.2013.09.074
- 12- Tian W, Chen C, Lei X, Zhao J, Liang J. CASTp 3.0: computer atlas of surface topology of proteins. *Nucleic Acids Res*. 2018;46(W1):W363-W367. doi : 10.1093/nar/gky473
  - 13- Rizvi SMD, Hussain T, Moin A, Dixit SR, Mandal SP, Adnan M, et al. Identifying the Most Potent Dual-Targeting Compound(s) against 3CLprotease and NSP15exonuclease of SARS-CoV-2 from *Nigella sativa*: Virtual Screening via Physicochemical Properties, Docking and Dynamic Simulation Analysis. *Processes*. 2021;9(10):1814. doi:10.3390/pr9101814
  - 14- Outbreak cases and data |Mpox | Poxvirus | CDC. <https://www.cdc.gov/poxvirus/mpox/response/2022/index.html>. Published January 11, 2024. Accessed February 17, 2024.
  - 15- About Mpox | Mpox | Poxvirus | CDC. <https://www.cdc.gov/poxvirus/mpox/about/index.html>. Published September 26, 2023. Accessed February 17, 2024.
  - 16- Zhu C, Gao Y, Li H, Meng S, Li L, Francisco JS, et al. Characterizing hydrophobicity of amino acid side chains in a protein environment via measuring contact angle of a water nanodroplet on planar peptide network. *Proc Natl Acad Sci U S A*. 2016;113(46):12946-12951. doi:10.1073/pnas.1616138113
  - 17- McVey DS, Kennedy M, Chengappa MM, Wilkes R, eds. *Veterinary Microbiology*. 1st ed. Wiley; 2022. doi:10.1002/9781119650836
  - 18- Afonso PP, Silva PM, Schnellrath LC, Jesus, D. M., Hu J., Yang Y, et al. Biological characterization and next-generation genome sequencing of the unclassified Cotia virus SPAn232 (Poxviridae). *J Virol*. 2012;86(9):5039-5054. doi:10.1128/JVI.07162-11
  - 19- Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*. 2007;35(2):W407-410. doi:10.1093/nar/gkm290
  - 20- Sippl MJ. Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J Mol Biol*. 1990;213(4):859-883. doi:10.1016/s0022-2836(05)80269-4
  - 21- Sippl MJ. Knowledge-based potentials for proteins. *Curr Opin Struct Biol*. 1995;5(2):229-235. doi:10.1016/0959-440x(95)80081-6
  - 22- Binkowski TA, Naghibzadeh S, Liang J. CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic Acids Res*. 2003;31(13):3352-3355. doi:10.1093/nar/gkg512
  - 23- Edelsbrunner H, Mücke EP. Three-dimensional alpha shapes. *ACM Trans Graph*. 1994;13(1):43-72. doi:10.1145/174462.156635
  - 24- Rabbi MF, Akter SA, Hasan MJ, Amin A. In Silico Characterization of a Hypothetical Protein from *Shigella dysenteriae* ATCC 12039 Reveals a Pathogenesis-Related Protein of the Type-VI Secretion System. *Bioinform Biol Insights*. 2021;15:11779322211011140. Published 2021 Apr 22. doi:10.1177/11779322211011140