RESEARCH

In silico structural and functional characterization of hypothetical proteins from Monkeypox virus

Kajal Gupta[1*]

Abstract

Background Monkeypox virus is a small, double-stranded DNA virus that causes a zoonotic disease called Monkeypox. The disease has spread from Central and West Africa to Europe and North America and created havoc in some countries all around the world. The complete genome of the Monkeypox virus Zaire-96-I-16 has been sequenced. The viral strain contains 191 protein-coding genes with 30 hypothetical proteins whose structure and function are still unknown. Hence, it is imperative to functionally and structurally annotate the hypothetical proteins to get a clear understanding of novel drug and vaccine targets. The purpose of the study was to characterize the 30 hypothetical proteins through the determination of physicochemical properties, subcellular characterization, function prediction, functional domain prediction, structure prediction, structure validation, structural analysis, and ligand binding sites using Bioinformatics tools.

Results The structural and functional analysis of 30 hypothetical proteins was carried out in this research. Out of these, 3 hypothetical functions (Q8V547, Q8V4S4, Q8V4Q4) could be assigned a structure and function confidently. Q8V547 protein in Monkeypox virus Zaire-96-I-16 is predicted as an apoptosis regulator which promotes viral replication in the infected host cell. Q8V4S4 is predicted as a nuclease responsible for viral evasion in the host. The function of Q8V4Q4 is to prevent host NF-kappa-B activation in response to pro-inflammatory cytokines like TNF alpha or interleukin 1 beta.

Conclusions Out of the 30 hypothetical proteins of Monkeypox virus Zaire-96-I-16, 3 were annotated using various bioinformatics tools. These proteins function as apoptosis regulators, nuclease, and inhibitors of NF-Kappa-B activator. The functional and structural annotation of the proteins can be used to perform a docking with potential leads to discover novel drugs and vaccines against the Monkeypox. In vivo research can be carried out to identify the complete potential of the annotated proteins.

Keywords Monkeypox, Monkeypox virus, Hypothetical proteins, Bioinformatics analysis, Drug target identification

Background

Monkeypox (MPX) is a viral disease caused by the Monkeypox virus (MPXV), which can spread from vertebrates to humans and vice versa. It was earlier prevalent in Central and West Africa. Due to its similarity with symptoms in smallpox, MPX was also known as "Monkey smallpox." The major symptoms of MPX are high fever, headache, lymphadenopathy, and systemic blisters and pustules [1]. The mortality rate of Monkeypox is 1–10% [2, 3]. Since May 2022, Monkeypox cases have been found not only in the countries where the disease was endemic but also in certain countries of Europe and North America. From 1 January 2022 to 22 August 2022, 41,664 laboratory-confirmed cases and 12 deaths were reported in the world, according to the Multi-country outbreak of Monkeypox report published by the World Health Organization (WHO). Ten countries

*Correspondence:
Kajal Gupta
gupta.kajal49@gmail.com
[1]Department of Biochemistry, Daulat Ram College, University of Delhi, Delhi, India

including the USA, Spain, Brazil, and others account for 88.9% of the cases reported globally (Fig. 1) [4].

The increased emergence of the disease has created havoc among a wider audience leading to higher demand for potential drugs and vaccines. There are currently no FDA-approved drugs for Monkeypox. The FDA-approved drugs for smallpox tecovirimat and brincidofovir are administered to patients with Monkeypox. However, low viral resistance barrier can render the drug useless when used indiscriminately. An FDA-approved vaccine JYNNEOS vaccine is being marketed for prevention of Monkeypox in individuals older than 18 years and susceptible to the infection [5]. Thus, it has become essential to study the genome of the Monkeypox virus in-depth for understanding new potential drug and vaccine targets in MPXV which can be administered safely to individuals of all age groups. In this research paper, we studied the hypothetical proteins in MPXV to discover new drug and vaccine targets in MPXV.

Monkeypox virus (MPXV) is a poxvirus belonging to the *Orthopoxvirus* genus. The 200×250 nm sized virus particle has an oval or brick shape [6] and produces two infectious particles during replication: intracellular mature viral particle and extracellular enveloped viral particle. The structure of an intracellular mature viral particle comprises a lipoprotein envelope around the viral core and a lateral body rich in proteins. It is quite stable in the external environment and is released in the environment by cell lysis. It usually aids the virus in disease transmission between different animals. On the other hand, an extracellular enveloped viral particle comprises a lipid membrane wrapped around the intracellular mature viral particle which is formed from the transport Golgi apparatus or endosomes [7].

The replication cycle of poxvirus is completed in the cytoplasm. The host cell invasion takes place in 3 steps: adsorption, membrane fusion, and core invasion. Although specific cell receptors are not known for MPXV invasion, it has been identified that in the case of vaccinia virus (VACV), which is similar to MPXV, the adsorption on the cell surface occurs by four viral proteins namely D8, A27, A26, and H3. D8 binds to chondroitin [8], A26 binds to laminin [9], and A26 and H3 bind to heparan [10] mediating the adsorption of the virus on the host cell surface, which is followed by membrane fusion and core invasion.

The viral proteins A16, A21, A28, F9, G3, G9, H2, J5, L1, L5, and O3 form an entry fusion complex in VACV to allow the invasion of intracellular mature viral particle and Extracellular enveloped viral particle in the host cell. Except for O3, all the other viral proteins also play an important role in poxvirus replication [11, 12]. Following the membrane invasion, the viral core enters the cytoplasm and is de-hulled by the action of certain viral proteins like A16L, A21L, A28L, F9L, G3L, G9R, H2R, J5L, and L5R, to initiate viral biosynthesis [12, 13].

Monkeypox virus Zaire-96-I-16(MPXV-ZAI) [NCBI:txid619591] has a non-segmented, linear, double-stranded DNA genome consisting of 196.858 kilobase pairs (Kbp) [NCBI Reference Sequence: NC_003310.1]. The genome comprises a total of 191 genes and 175 protein clusters [14]. There are 190 non-overlapping ORFs of ≥ 60 amino acids, and the GC content is approximately 31.1%. The coding region of MPXV-ZAI is around 195,118 bp long. *Homo sapiens* are the known hosts for MPXV [15]. One hundred ninety-one viral segments of MPXV-ZAI can be classified into various categories according to their function. Thirty proteins out of these unknown viral segments are hypothetical proteins, which
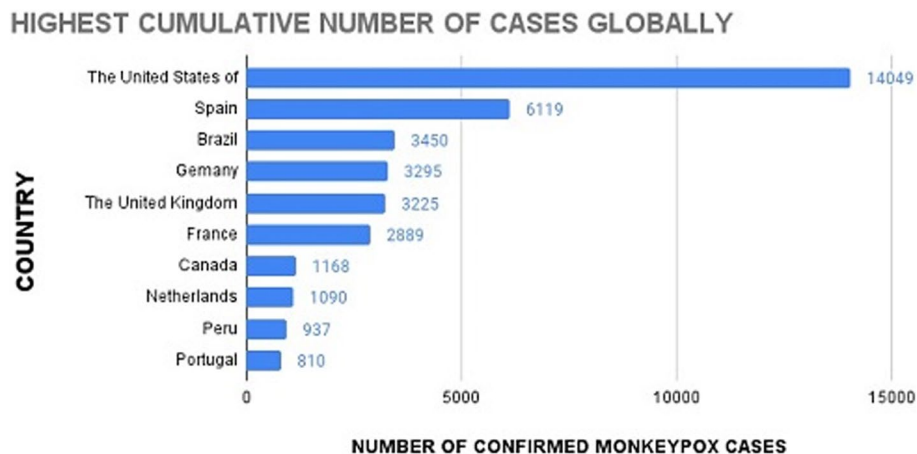

Fig. 1 Ten countries with highest cumulative number of Monkeypox cases globally (World Health Organization)

have been studied in this research. The protein length, Uniprot Id, and locus tag are listed in Table S1 [16].

Methods
Sequence retrieval
We analyzed the genome of the Monkeypox virus Zaire-96-I-16(MPXV-ZAI)[NCBI:txid619591] and found 161 protein-coding genes (http://www.ncbi.nlm.nih.gov/genome) and a total of 30 proteins as hypothetical proteins (HPs). The FASTA sequences of all 30 HPs proteins were retrieved from Uniprot (http://www.uniprot.org) [17].

Physicochemical properties of HPs
The physicochemical parameters of all 30 HPs were studied using Expasy's ProtParam server [18] (https://web.expasy.org/protparam/), which were then used for theoretical measurements of various parameters such as molecular weight, theoretical isoelectric point, extinction coefficient [19], instability index [20], aliphatic index, and grand average of hydropathicity (GRAVY) [21]. The extinction coefficient measures the amount of light that proteins absorb at a certain wavelength. Theoretical estimation of the stability of a protein in a test tube can be known by the instability index. The relative volume occupied by aliphatic side chain amino acids in the protein can be known by the aliphatic index of a protein. The GRAVY score for a peptide or protein is calculated as the sum of the hydropathy values of all of the amino acids, divided by the number of residues in the protein sequence. The predicted properties of HPs are listed in Table S2.

Subcellular localization and transmembrane helices prediction
The subcellular location of a protein usually determines the function of proteins. Thus, the prediction of subcellular localization from protein sequence can be useful to determine the protein function which can further aid in developing antiviral drugs and vaccines. The protein can be present in the outer membrane, inner membrane, capsid, periplasm, cytoplasm, or extracellular space [22]. The subcellular localization of HPs were predicted using Virus-PLoc [23–26], TMHMM [27–29], and HMMTOP [30, 31].

Virus-PLoc (http://www.csbio.sjtu.edu.cn/bioinf/virus/) predicts the subcellular location of viral proteins within a host and virus-infected cells. The predictor classifies the viral protein locations in the following categories: (1) cytoplasm, (2) endoplasmic reticulum, (3) extracellular, (4) inner capsid, (5) nucleus, (6) outer capsid, and (7) plasma membrane. The prediction approach is by fusing PseAA (pseudo amino acid) composition.

TMHMM (https://services.healthtech.dtu.dk/service.php?TMHMM-2.0) predicts the transmembrane helices in proteins. The TMHMM prediction tool gives information about the most probable location and orientation of transmembrane helices in the protein sequence. It works on the algorithm called N-best that sums over all paths through the model with the same location and direction of the helices.

HMMTOP (http://www.enzim.hu/hmmtop/index.php) predicts the localization of helical transmembrane segments and the topology of helical transmembrane proteins. The predicted subcellular localization of HPs is listed in Table S3.

Function prediction using sequence analogy
The most basic step to understanding the function of an unknown protein is by looking for its structural homologs in different genomics and proteomics-based databases. This approach works on the hypothesis that an unknown protein with sequence analogy to a known protein may have a similar function as the known protein. The sequence similarity search was performed via protein BLAST (pBLAST) against the non-redundant database. Generally, HPs contain low identity as compared to other known or annotated proteins [32, 33]. The predicted functions according to sequence analogy through pBLAST are listed in Table S4.

Function and functional domain prediction through sequence analysis
Precise function of a protein can be determined using the information about functional domain in the HPs, various bioinformatics tools like SMART [34, 35], MOTIF-SCAN [36], INTERPROSCAN [37–39], pfp-fundseqe [25, 40, 41], PFAM [42–46], CATH [47, 48], and SUPERFAMILY [49, 50] were used to detect functional domains in HPs and classify the 30 HPs into family and superfamily.

SMART (a Simple Modular Architecture Research Tool) identifies and annotates the genetically mobile domains and analyses domain architectures.

Motif-Scan (including HAMAP profiles, Prosite patterns, Pfam HMMs (local models), and Pfam HMMs (global models)) were used to find the motifs in a protein sequence.

InterProScan uses protein sequence in FASTA format to predict the family to which the unknown protein belongs and the domain present in it. The predicted functional domains in HPs are listed in Table S5.

Structure prediction and validation
The structure of 30 HPs were predicted using Phyre2 Batch processing [51]. Phyre2 utilizes an extensive remote homology detection method to build 3D models

of unknown proteins. Seven protein models were confidently predicted by Phyre2. The models were validated using Procheck [52–55] and SWISS-Model QMEAN-DisCo scoring function [56].The QMEANDisCo scoring function of 7 predicted models is listed in Table S6.

Function annotation using structural analysis
Protein function predicted from structure gives more valuable information than protein function predicted through sequence analysis [57]. The structural analysis of 7 predicted protein models was done using DeepFRI [58] and COACH [59, 60].

The structure-based molecular function and biological process predicted using DeepFRI are listed in Table S7.

The ligand and consensus binding residues of predicted models are listed in Table S8.

Results
Monkeypox virus Zaire-96-I-16 (MPXV-ZAI) comprises 191 protein-coding genes. Out of these, 30 proteins are regarded as hypothetical proteins. In this study, these proteins were studied using various bioinformatics tools for annotating their function in viral infection. The amino acid length of 30 hypothetical proteins ranges from 64 for shortest protein to 737 for longest protein. Expasy's ProtParam tool was used to estimate the physicochemical properties of the HPs. The subcellular localization and

presence of transmembrane helices revealed the presence of most of the HPs in the inner layer of the host cell and a few in the Endoplasmic reticulum and host cytoplasm. Approximately half of the HPs were predicted as transmembrane proteins with 1–2 transmembrane helices. The function of 5 hypothetical proteins could be predicted with confidence (≥ 4 software predictions). The function of these proteins and their subcellular localization are listed in Table 1.

The function of additional 8 proteins could be predicted with uncertainty ($= 3$ software predictions). The function of these proteins and their subcellular localization are listed in Table 2.

The structure prediction and structure analysis of 30 hypothetical proteins were done. Out of 30 proteins, the structure of 7 proteins were confidently predicted ($> 65\%$ of residues modeled with confidence $> 90\%$). The structure of these 7 proteins were validated and analyzed through various bioinformatics software.

Discussion
In this study, we carried out the structural and functional annotation of 30 hypothetical proteins from the Monkeypox virus Zaire-96-I-16(MPXV-ZAI). Physicochemical properties predicted through Expasy's ProtParam server revealed that the isoelectric point of the HPs ranges from 3.48 to 9.49. The isoelectric point is the pH at which the

Table 1 Hypothetical proteins with confidently predicted function (≥ 4 software prediction) and subcellular localization

S.NO	Protein name	UniprotKB identifier	Function	Virus-PLoc prediction
1	Hypothetical protein MPXVgp033	Q8V547	Apoptosis regulators	Endoplasmic Reticulum
2	Hypothetical protein MPXVgp037	Q8V543	Telomere binding protein	Endoplasmic Reticulum
3	Hypothetical protein MPXVgp043	Q8V537	Cell-to-cell spread of the virus	Inner layer
4	Hypothetical protein MPXVgp081	Q8V501	Early virion morphogenesis and formation or elongation of crescent membranes	Plasma membrane
5	Hypothetical protein MPXVgp165	Q8V4S4	Nuclease	Plasma membrane

Table 2 Hypothetical proteins with confidently predicted function ($= 3$ software prediction) and subcellular localization

S.NO	Protein name	UniprotKB identifier	Function	Virus-PLoc prediction
1	Hypothetical protein MPXVgp016	Q8V563	Inhibitor of host NF-kappa-B activation	Endoplasmic reticulum
2	Hypothetical protein MPXVgp026	Q8V554	Immunosuppressive activities	Endoplasmic reticulum
3	Hypothetical protein MPXVgp056	Q8V524	Virion assembly	Cytoplasm
4	Hypothetical protein MPXVgp066	Q8V514	Membrane protein	Endoplasmic reticulum
5	Hypothetical protein MPXVgp067	Q8V513	Binds to the hairpin form of the viral telomeric sequence	Plasma membrane
6	Hypothetical protein MPXVgp074	Q8V506	Nuclease	Inner layer
7	Hypothetical protein MPXVgp147	Q8V4U2	Virulence	Outer capsid
8	Hypothetical protein MPXVgp185	Q8V4Q4	Prevent host NF-kappa-B activation	Cytoplasm

net charge on a protein molecule is zero. Protein at its isoelectric point is mildly insoluble, compact, and stable, which leads to protein getting crystallized [61].

The extinction coefficient of protein ranges from 1490 M-1 Cm-1 to 80,540 M-1 Cm-1 at 280 nm. Analysis of the extinction coefficient of a protein helps to determine the protein–protein interaction and protein–ligand interaction in solution [19]. The instability index provides information about the stability of a protein in the test tube. A value of greater than 40 indicates that a protein is unstable and a value of less than 40 indicates a stable protein. Out of 40 hypothetical proteins, 13 proteins were regarded as unstable and 17 proteins are stable in the test tube. The relative volume occupied by the aliphatic side chains in a protein is indicated by the Aliphatic index. It can be used to determine the thermostability of globular proteins [62].The aliphatic index of HPs ranges from 36.56 to 177.83. GRAVY (Grand Average of Hydropathy) of the HPs ranges from -1.602 to 1.223.

The detailed analysis of 3 of the HPs from the Monkeypox virus whose function and structure (through Phyre2) could be annotated with confidence is mentioned as follows.

Q8V547

Q8V547 is predicted as an apoptosis regulator that promotes viral replication in the host cell and has an immunoglobulin-like domain. Also, it may act as an inhibitor of NLR-mediated interleukin-1 beta/IL1B production in infected cells. The function is confidently predicted by 5 software. The Virus Ploc software detected its subcellular location in the endoplasmic reticulum of the virus-infected cell with no transmembrane helices. The predicted 3-dimensional structure showed that 144 residues (66% of the sequence) have been modeled with 100.0% confidence by the single highest scoring template (Figure S1), and validation through the Ramachandran plot revealed 96.50% residues in the most favored regions and 2.8% residues in additional allowed regions (Figure S2). There are no residues in the outlier region. The QMEANDisCo Global Score of the predicted model was 0.68 ± 0.07(above the cutoff range of 0.50). The structure analysis through COACH predicted the ligand for this protein to be a peptide. The Gene Ontology structure-based analysis predicted the biological process of Q8V547 in cellular metabolism and primary metabolism.

Q8V4S4

Q8V4S4 is predicted as a nuclease responsible for viral evasion. A DNA-binding domain in the protein has also been detected in the functional domain annotation. The function has been annotated confidently by 4 function prediction software. A confidently predicted 3-dimensional structure of the protein with 97% of residues modeled at > 90% confidence is selected (Figure S3). Validation of the structure through the Ramachandran plot revealed 73% in the most favored region, 18.6% residues in the additional allowed region, and 4% residues in the outlier region (Figure S4). The Swiss model QMEANDisCo Global Score of the predicted structure was 0.59 ± 0.05 (above the cutoff range of 0.50). The analysis of the predicted structure through COACH revealed GERAN-8-YL GERAN as the ligand for the protein. The Gene Ontology structure-based analysis of Q8V4S4 predicted the molecular function as organic cyclic and heterocyclic compound binding and a biological role in organic substance metabolism, primary metabolism, nitrogen compound metabolism, macromolecule metabolism, cellular metabolism, and cellular macromolecule metabolism.

Q8V4Q4

The function of Q8V4Q4 as predicted confidently by 3 software is to prevent host NF-kappa-B activation in response to pro-inflammatory cytokines like TNF alpha or interleukin 1 beta. A 4 helical cytokine binding motif is predicted by pfp-fundseqe. A 3-dimensional model with 133 residues (87% of the sequence) has been modeled with 100.0% confidence by the single highest scoring template (Figure S5). The model as validated by Ramachandran plot revealed 84.1% residues in most favored regions, 12.1% in additional allowed regions, and 1.5% in outlier region (Figure S6). QMEANDisCo Global Score of the predicted model is 0.63 ± 0.07(above the cutoff range of 0.50). The structural analysis of the model on COACH revealed 3-[3-(4-chloro-3,5-dimethylphenoxy) PROPYL]-1-benzothiophene-2-carboxylic acid as the ligand of the protein. The structure analysis of the 3D model revealed protein binding as the molecular function of the protein and regulation of cellular process as the biological process.

Conclusion

This research worked on predicting the structure and function of hypothetical proteins of Monkeypox virus Zaire-96-I-16 using bioinformatics tools and software. Three of the 30 hypothetical proteins (Q8V547, Q8V4S4, Q8V4Q4) were confidently annotated. Q8V547 protein was predicted as an apoptosis regulator which promotes viral replication in the infected host cell.

Q8V4S4 was predicted as a nuclease responsible for viral evasion in the host. The function of Q8V4Q4 was to prevent host NF-kappa-B activation in response to pro-inflammatory cytokines like TNF alpha or interleukin 1 beta. Further studies are required to unravel the complete potential of these proteins as potential drug

targets and vaccines. Since these proteins were predicted to be involved in protection and proliferation of the virus inside a host, they can prove to be good targets for drugs and vaccines for protection against the disease.

Abbreviations

MPX	Monkeypox
MPXV	Monkeypox virus
VACV	Vaccinia virus
MPXV-ZAI	Monkeypox virus Zaire-96-I-16
HPs	Hypothetical proteins
GRAVY	Grand average of hydropathicity

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s43141-023-00505-w.

> **Additional file 1: Supplementary Figure 1 (Figure S1).** 3D structure of hypothetical protein Q8V547 predicted from Phyre2. **Supplementary Figure 2 (Figure S2):** Evaluation of 3D structure of Hypothetical Protein Q8V547 through Ramachandran plot. **Supplementary Figure 3 (Figure S3):** 3D structure of hypothetical protein Q8V4S4 predicted from Phyre2. **Supplementary Figure 4 (Figure S4):** Evaluation of 3D structure of Hypothetical Protein Q8V4S4 through Ramachandran plot. **Supplementary Figure 5 (Figure S5):** 3D structure of hypothetical protein Q8V4Q4 predicted from Phyre2. **Supplementary Figure 6 (Figure S6):** Evaluation of 3D structure of Hypothetical Protein Q8V4Q4 through Ramachandran plot.
>
> **Additional file 2: Supplementary Table 1 (Table S1).** The table presents the 30 Hypothetical proteins of Monkeypox virus Zaire-96-I-16 along with the Uniprot Id, Locus tag and protein length. **Supplementary Table 2 (Table S2):** The table reports the physiochemical properties of 30 hypothetical proteins as predicted by Expasy's Protparam server. **Supplementary Table 3 (Table S3):** The table details the list of predicted subcellular location and presence of transmembrane helices in 30 Hypothetical proteins. **Supplementary Table 4 (Table S4):** The table reports the function of 30 hypothetical proteins as predicted by Protein BLAST. **Supplementary Table 5(Table S5):** The table presents the function and functional domain of 30 Hypothetical proteins as annotated by SMART,Motif Scan, pfp- fundseqe, InterProscan, PFAM, CATH and Superfamily. **Supplementary Table 6 (Table S6):** The table entails the QMEANDisCo Global Score of the predicted 3 dimensional structure of 7 Hypothetical proteins. **Supplementary Table 7 (Table S7):** The table reports the Structure-Based Molecular Function and Structure-Based Biological Process of the 7 Hypothetical proteins from there predicted 3 dimensional structure. **Supplementary Table 8 (Table S8):** The table presents the ligand binding site prediction of 7 Hypothetical proteins of Monkeypox virus.

Declarations

References

1. Gong Q, Wang C, Chuai X, Chiu S (2022) Monkeypox virus: a re-emergent threat to humans. Virologica Sinica 37(4):477–482. https://doi.org/10.1016/j.virs.2022.07.006
2. Doshi RH, Guagliardo SA, Doty JB, Babeaux AD, Matheny A, Burgado J, Townsend MB, Morgan CN, Satheshkumar PS, Ndakala N, Kanjingankolo T (2019) Epidemiologic and ecologic investigations of monkeypox, Likouala Department, Republic of the Congo, 2017. Emerg Infect Dis 25(2):281–289. https://doi.org/10.3201/eid2502.181222
3. Ogoina D, Izibewule JH, Ogunleye A, Ederiane E, Anebonam U, Neni A, Oyeyemi A, Etebu EN, Ihekweazu C (2019) The 2017 human monkeypox outbreak in Nigeria—report of outbreak experience and response in the Niger Delta University Teaching Hospital, Bayelsa State, Nigeria. PLoS One 14(4):e0214229. https://doi.org/10.1371/journal.pone.0214229
4. World Health Organization.(2022, August 24) "Multi-country outbreak of monkeypox". Retrieved from https://www.who.int/publications/m/item/multi-country-outbreak-of-monkeypox--external-situation-report--4---24-august-2022.
5. Food and Drug administration.(2023, January 2) "FDA Mpox Response". Retrieved from https://www.fda.gov/emergency-preparedness-and-response/mcm-issues/fda-mpox-response
6. Cho CT, Wenner HA (1973) Monkeypox virus. Bacteriological reviews 37(1):1–8. https://doi.org/10.1128/br.37.1.1-18.1973
7. Pickup DJ (2015) Extracellular virions: the advance guard of poxvirus infections. PLoS Pathogens 11(7):e1004904. https://doi.org/10.1371/journal.ppat.1004904
8. Matho MH, Schlossman A, Gilchuk IM, Miller G, Mikulski Z, Hupfer M, Wang J, Bitra A, Meng X, Xiang Y, Kaever T (2018) Structure–function characterization of three human antibodies targeting the vaccinia virus adhesion molecule D8. J Biol Chem. 293(1):390–401. https://doi.org/10.1074/jbc.M117.814541
9. Chiu WL, Lin CL, Yang MH, Tzou DLM, Chang W (2007) Vaccinia virus 4c (A26L) protein on intracellular mature virus binds to the extracellular cellular matrix laminin. J virol 81(5):2149–2157. https://doi.org/10.1128/JVI.02302-06
10. Singh K, Gittis AG, Gitti RK, Ostazeski SA, Su HP, Garboczi DN (2016) The vaccinia virus H3 envelope protein, a major target of neutralizing antibodies, exhibits a glycosyltransferase fold and binds UDP-glucose. J Virol 90(10):5020–5030. https://doi.org/10.1128/JVI.02933-15
11. Schin AM, Diesterbeck US, Moss B (2021) Insights into the organization of the poxvirus multicomponent entry-fusion complex from proximity analyses in living infected cells. J Virol 95(16):e00852-e921. https://doi.org/10.1128/JVI.00852-21
12. Senkevich TG, Ojeda S, Townsley A, Nelson GE, Moss B (2005) Poxvirus multiprotein entry–fusion complex. Proc Nat Acad Sci 102(51):18572–18577. https://doi.org/10.1073/pnas.0509239102
13. Brown E, Senkevich TG, Moss B (2006) Vaccinia virus F9 virion membrane protein is required for entry but not virus assembly, in contrast to the related L1 protein. J virol 80(19):9455–9464. https://doi.org/10.1128/JVI.01149-06
14. Schoch CL, Ciufo S, Domrachev M, Hotton CL, Kannan S, Khovanskaya R, Leipe D, Mcveigh R, O'Neill K, Robbertse B, Sharma S(2020). NCBI

Taxonomy: a comprehensive update on curation, resources and tools. Database(Oxford).https://doi.org/10.1093/database/baaa062

15. Shchelkunov SN, Totmenin AV, Babkin IV, Safronov PF, Ryazankina OI, Petrov NA, Gutorov VV, Uvarova EA, Mikheev MV, Sisler JR, Esposito JJ (2001) Human monkeypox and smallpox viruses: genomic comparison. FEBS letters 509(1):66–70. https://doi.org/10.1016/S0014-5793(01)03144-1

16. Genome. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004 – [cited 2022 August 27]. Available from: https://www.ncbi.nlm.nih.gov/genome/

17. The UniProt Consortium(2022), UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Research, gkac1052.https://doi.org/10.1093/nar/gkac1052

18. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic acids res 31(13):3784–8. https://doi.org/10.1093/nar/gkg563

19. Gill SC, Von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal biochem 182(2):319–26. https://doi.org/10.1016/0003-2697(89)90602-7

20. Guruprasad K, Reddy BB, 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):155–161. https://doi.org/10.1093/protein/4.2.155

21. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J mol biol 157(1):105–32. https://doi.org/10.1016/0022-2836(82)90515-0

22. Naveed M, Tehreem S, Usman M, Chaudhry Z, Abbas G (2017) Structural and functional annotation of hypothetical proteins of human adenovirus: prioritizing the novel drug targets. BMC res notes 10(1):1–6. https://doi.org/10.1186/s13104-017-2992-z

23. Chou KC, Shen HB (2008) Cell-PLoc: a package of Web servers for predicting subcellular localization of proteins in various organisms. Nat protoc 3(2):153–62. https://doi.org/10.1038/nprot.2007.494

24. Shen HB, Chou KC (2007) Virus-PLoc: a fusion classifier for predicting the subcellular localization of viral proteins within host and virus-infected cells. Biopolymers 85(3):233–240. https://doi.org/10.1002/bip.20640

25. Chou KC (2005) Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes. Bioinformatics 21(1):10–9. https://doi.org/10.1093/bioinformatics/bth466

26. Shen HB, Chou KC (2006) Ensemble classifier for protein fold pattern recognition. Bioinformatics 22(14):1717–1722. https://doi.org/10.1093/bioinformatics/btl170

27. Möller S, Croning MD, Apweiler R (2001) Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics 17(7):646–53. https://doi.org/10.1093/bioinformatics/17.7.646

28. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J mol biol 305(3):567–80. https://doi.org/10.1006/jmbi.2000.4315

29. Sonnhammer EL, Von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol 6:175–182. https://doi.org/10.1006/jmbi.2000.4315

30. Tusnády GE, Simon I (1998) Principles governing amino acid composition of integral membrane proteins: applications to topology prediction. J Mol Biol 283:489–506. https://doi.org/10.1006/jmbi.1998.2107

31. Tusnády GE, Simon I (2001) The HMMTOP transmembrane topology prediction server. Bioinformatics 17:849–850. https://doi.org/10.1093/bioinformatics/17.9.849

32. Mahram A, Herbordt MC (2010) Fast and accurate NCBI BLASTP: acceleration with multiphase FPGA-based prefiltering. InProceedings of the 24th ACM International Conference on Supercomputing, pp 73–82. https://doi.org/10.1145/1810085.1810099

33. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–10. https://doi.org/10.1016/S0022-2836(05)80360-2

34. Letunic I, Doerks T, Bork P (2012) SMART 7: recent updates to the protein domain annotation resource. Nucleic acids res 40(D1):D302-5. https://doi.org/10.1093/nar/gkr931

35. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P (2000) SMART: a web-based tool for the study of genetically mobile domains. Nucleic acids res 28(1):231–4. https://doi.org/10.1093/nar/28.1.231

36. Pagni M, Ioannidis V, Cerutti L, Zahn-Zabal M, Jongeneel CV, Hau J, Martin O, Kuznetsov D, Falquet L (2007) MyHits: improvements to an interactive resource for analyzing protein sequences. Nucleic Acids Res 35:W433-7. https://doi.org/10.1093/nar/gkm352

37. Venkataraman A, Chew TH, Hussein ZA, Shamsir MS (2011) A protein short motif search tool using amino acid sequence and their secondary structure assignment. Bioinformation 7(6):304. https://doi.org/10.6026/007/97320630007304

38. Zdobnov EM, Apweiler R (2001) InterProScan–an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17(9):847–848. https://doi.org/10.1093/bioinformatics/17.9.847

39. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–40. https://doi.org/10.1093/bioinformatics/btu031

40. Shen HB, Chou KC (2009) Predicting protein fold pattern with functional domain and sequential evolution information. J Theor Biol 256(3):441–6. https://doi.org/10.1016/j.jtbi.2008.10.007

41. Shen HB, Chou KC (2006) Ensemble classifier for protein fold pattern recognition. Bioinformatics 22(14):1717–22. https://doi.org/10.1093/bioinformatics/btl170

42. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer EL, Tosatto SC, Paladin L, Raj S, Richardson LJ, Finn RD (2021) Pfam: the protein families database in 2021. Nucleic acids res 49(D1):D412-9. https://doi.org/10.1093/nar/gkaa913

43. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA (2016) The Pfam protein families database: towards a more sustainable future. Nucleic acids res 44(D1):D279-85. https://doi.org/10.1093/nar/gkv1344

44. Bateman A, Birney E, Durbin R, Eddy SR, Finn RD, Sonnhammer EL (1999) Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. Nucleic acids res 27(1):260–2. https://doi.org/10.1093/nar/27.1.260

45. Sonnhammer EL, Eddy SR, Birney E, Bateman A, Durbin R (1998) Pfam: multiple sequence alignments and HMM-profiles of protein domains. Nucleic acids res 26(1):320–2. https://doi.org/10.1093/nar/26.1.320

46. Sonnhammer EL, Eddy SR, Durbin R (1997) Pfam: a comprehensive database of protein domain families based on seed alignments. Proteins 28(3):405–20. https://doi.org/10.1002/(SICI)1097-0134(199707)28:3%3c405::AID-PROT10%3e3.0.CO;2-L

47. Kundsen M, Wiuf C (2010) The CATH database. Hum genomics 4(3):207–212. https://doi.org/10.1186/1479-7364-4-3-207

48. Pearl FM, Lee D, Bray JE, Buchan DW, Shepherd AJ, Orengo CA (2002) The CATH extended protein-family database: providing structural annotations for genome sequences. Protein Sci 11(2):233–244. https://doi.org/10.1110/ps.16802

49. Wilson D, Pethica R, Zhou Y, Talbot C, Vogel C, Madera M, Chothia C, Gough J (2009) SUPERFAMILY—sophisticated comparative genomics, data mining, visualization and phylogeny. Nucleic acids res 37(suppl_1):D380-6. https://doi.org/10.1093/nar/gkn762

50. Wilson D, Madera M, Vogel C, Chothia C, Gough J (2007) The SUPERFAMILY database in 2007: families and functions. Nucleic acids res 35(suppl_1):D308-13. https://doi.org/10.1093/nar/gkl910

51. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat protoc 10(6):845–58. https://doi.org/10.1038/nprot.2015.053

52. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK - a program to check the stereochemical quality of protein structures. J Applied Crystallogr 26:283–291. https://doi.org/10.1107/S0021889809009944

53. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8:477–486. https://doi.org/10.1007/BF00228148. ([PubMed id: 9008363])

54. Laskowski R A, MacArthur M W, Thornton J M (2001). PROCHECK: validation of protein structure coordinates, in International Tables of Crystallography, Volume F. Crystallography of Biological Macromolecules, eds. Rossmann M G & Arnold E, Dordrecht, Kluwer Academic Publishers, The Netherlands, pp. 722–725.

55. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992) Stereo-chemical quality of protein structure coordinates. Proteins 12:345–364. https://doi.org/10.1002/prot.340120407. ([PubMed id: 1579569])

56. Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T (2020) QMEANDisCo—distance constraints applied on model quality estimation. Bioinformatics 36(6):1765–71. https://doi.org/10.1093/bioinformatics/btz828

57. Kumar K, Prakash A, Anjum F, Islam A, Ahmad F, Hassan M (2015) Structure-based functional annotation of hypothetical proteins from Candida dubliniensis: a quest for potential drug targets. 3 Biotech 5(4):561–76. https://doi.org/10.1007/s13205-014-0256-3

58. Gligorijević V, Renfrew PD, Kosciolek T, Leman JK, Berenberg D, Vatanen T, Chandler C, Taylor BC, Fisk IM, Vlamakis H, Xavier RJ (2021) Structure-based protein function prediction using graph convolutional networks. Nat commun 12(1):1–4. https://doi.org/10.1038/s41467-021-23303-9

59. Yang J, Roy A, Zhang Y (2013) Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics 29(20):2588–95. https://doi.org/10.1093/bioinformatics/btt447

60. Yang J, Roy A, Zhang Y (2012) BioLiP: a semi-manually curated database for biologically relevant ligand–protein interactions. Nucleic acids res 41(D1):D1096-103. https://doi.org/10.1093/nar/gks966

61. Kantardjieff KA, Rupp B (2004) Protein isoelectric point as a predictor for increased crystallization screening efficiency. Bioinformatics 20(14):2162–8. https://doi.org/10.1093/bioinformatics/bth066

62. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A(2005). Protein identification and analysis tools on the ExPASy server. The proteomics protocols handbook.571–607.https://doi.org/10.1385/1-59259-584-7:531

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.