Comprehensive Mutational Landscape Analysis of Monkeypox Virus Proteome

Tugba Ozaktas[1], Ayten Dizkirici[2], Alessandra Carbone[3,4], Mustafa Tekpinar[5*]

[1]Department of Molecular Biology, Faculty of Science, Van Yüzüncü Yıl University, 65080, Van, Türkiye.

[2]Department of Molecular Biology, Faculty of Science, Van Yüzüncü Yıl University, 65080, Van, Türkiye.

[3]Sorbonne Université, CNRS, IBPS, Laboratoire de Biologie Computationnelle et Quantitative - UMR 7238, 4 place Jussieu, 75005 Paris, France.

[4]Institut Universitaire de France, Paris 75005, France.

[5]Department of Physics, Faculty of Science, Van Yüzüncü Yıl University, 65080, Van, Türkiye.

Abstract

We present a comprehensive computational analysis of the single point mutational landscapes of the Monkeypox virus (MPXV) proteome. We reconstructed full single-point mutational landscapes of 171 MPXV proteins using two advanced mutational effect predictors, ESCOTT and iGEMME, selected for their superior performance on viral proteins, assessed by benchmarking against the experimental data in the ProteinGym (v1.0.0) dataset. A recent MPXV strain

[*] Corresponding author.

E-mail: tekpinar@buffalo.edu

Phone: +90 538 986 39 10

sequenced in July 2024 was used as the reference genome. Multiple sequence alignments and protein structures were generated using Colabfold v1.5.5, and the predicted structures were evaluated with pLDDT metric, secondary structure predictions, and comparisons with available experimental data, ensuring high confidence in the structural models. To prioritize the most mutation-sensitive proteins within the large MPXV proteome as prime candidates for drug or vaccine development, we introduced a novel, interpretable metric: Average Gene Mutation Sensitivity (AGMS). Among the top 20 identified proteins, several were membrane-associated proteins, expected to be important for viral interactions with the hosts. This analysis provides a valuable resource for assessing the impact of new MPXV variants and guiding therapeutic strategies. This pioneering study underscores the significance of understanding MPXV evolution in the context of the ongoing global health crisis and offers a robust computational framework to support this effort.

Keywords

Mpox, monkeypox, virus, genome, proteome, protein, mutation, variant.

Introduction

World Health Organization declared Mpox outbreak a public health emergency of international concern on August 14, 2024. 99518 people were infected with causative agent of Mpox, monkeypox virus (MPXV), and 207 of the infected patients died of the disease as of August 6, 2024 according the CDC (https://www.cdc.gov/poxvirus/mpox/response/2022/world-map.html). Understanding impact of mutations in MPXV is of utmost importance both for assessing evolution of the virus and development of novel vaccines or therapies against this emerging health threat. Unfortunately, there are no systematic investigations of variant effects with experimental methods like deep mutational scanning. Therefore, the computational prediction of the single point mutational landscape of the MPXV genome represents a critical frontier in understanding the virus's evolution, epidemiology, and potential responses to therapeutic interventions.

The MPXV has a large double-stranded DNA genome of approximately 197 kb in size, encoding around 180 proteins (1, 2). Compared to other orthopoxviruses, the Mpox genome is larger, with a higher number of open reading frames (3). Despite the fact that the mutational landscape of MPXV is characterized by a relatively slow mutation rate compared to RNA viruses, significant variations have been observed in recent strains, particularly those emerging post-2022. This discrepancy underscores the importance of ongoing surveillance and genomic analysis to monitor the virus's evolution and potential adaptive changes that could affect public health responses (4, 5). By leveraging computational tools to predict mutational impacts, researchers

can better understand the dynamics of MPXV and inform strategies for vaccine development and therapeutic interventions. The computational prediction of MPXV's mutational landscape not only enhances our understanding of the virus's evolutionary patterns but also serves as a vital resource for public health initiatives aimed at controlling future outbreaks.

Due to these reasons, we first determined the best computational variant prediction algorithm for viruses utilizing the viral proteins in a publicly available experimental dataset called ProteinGym (v1.0.0). Our study revealed that ESCOTT (6) and iGEMME (6, 7) methods are the best mutational effect prediction algorithms for viral proteins. We generated input multiple sequence alignments (MSA) and structure files with Colabfold v1.5.5. We assessed structure qualities with respect to a few available experimental structures and evaluated structure qualities over the entire genome with the pLDDT (predicted Local Distance Difference Test) (8, 9). Then, we calculated complete single point mutational landscapes of 171 MPXV proteins with these methods using Colabfold generated MSAs and structure files. Finally, we analyzed sensitivity of each gene to mutations with a metric called 'Average Gene Mutation Sensitivity' and identified the top 20 MPXV genes as potential targets for therapeutic development.

Methods

Reference MPXV Genome

Since we wanted to use a recent strain of the MPXV, we selected https://www.ncbi.nlm.nih.gov/nuccore/PQ008840.1, which has been sequenced in July, 2024. We generated fasta files for 175 proteins. OPG003, OPG025, OPG108, OPG147 genes were eliminated from the sequence set due to long unknown segments. The list of the 171 remaining proteins is given in Table S1 of Supplementary Information.

MSA and Structure Files

We obtained multiple sequence alignment files and protein structures using Colabfold v1.5.5 web server (10, 11). We removed the gaps in query sequences of the MSA files and converted them from a3m format to fasta format with reformat utility tool of HHSuite (12). Number of sequences in each MSA file is given in Table S1 of the Supplementary Information. We visualized 3D structures of the proteins with Pymol (13). Secondary structure calculations were performed with biotite Python library and dssp (14, 15). Colorblind friendly viridis colormap for Pymol (16) was obtained from https://github.com/smsaladi/pymol_viridis and AlphaFold (11) pLDDT colormap for Pymol was obtained from https://github.com/BobSchiffrin/pymol_scripts/blob/main/colour_af_plddt.py.

Calculation of Full Mutational Landscapes with ESCOTT and iGEMME

We used PRESCOTT program to calculate ESCOTT and iGEMME scores of all single point amino acid mutations of 171 genes (6, 7). iGEMME is an improved version of GEMME utilizing only evolutionary information to determine impact of mutations while ESCOTT uses both evolutionary and structural information for the same purpose (6, 7). The docker image containing version 1.6.0 of PRESCOTT at https://hub.docker.com/r/tekpinar/prescott-docker was used both for ESCOTT and iGEMME computations.

ESCOTT raw scores can be between [-12, 2], where low scores indicate high mutational impact, and high scores no effect. The raw scores are not uniform across different proteins. Due to this reason, we sort all scores in the ESCOTT/iGEMME matrix by their rank and divide the rank position by the total number of mutations in the matrix to obtain a score between zero and one. Since we want low scores (close to zero) to indicate no mutational impact and high scores (close to one) suggest high mutational impact, we subtract the final scores from 1. Ranksorted ESCOTT/iGEMME scores are obtained using this method.

Average Gene Mutation Sensitivity (AGMS)

Raw ESCOTT or iGEMME scores are minmax normalized to obtain values between zero and one. All values are subtracted from one in order to make zero values to indicate no effect and values close to one to indicate high impact upon the mutations. Then, Average Minmaxed Mutation Score (AMMS) is calculated by summing up all mutation scores divided by 20 for each amino acid position in a protein. To be more precise, AMMS is a per residue score associated to each position in the protein sequence. The AMMS is used to calculate AGMS for a protein with N amino acids as follow:

$$AGMS = \frac{\sum_{i=1}^{N} AMMS}{N}$$

AGMS is a score between zero and one that can help prioritize drug or vaccine targets. Low values suggest low sensitivity to mutations for the gene, while scores close to one indicate high sensitivity to mutations. It should be noted that if 'ranksorted' values had been used, AGMS would always be 0.5 and that's why we used minmax normalized AMMS scores to calculate AGMS.

Results and Discussion

Selecting the Best Mutational Effect Prediction Algorithm for Viral Proteins

Extended ProteinGym dataset (v1.0.0) contains 217 deep mutational scanning substitution experiments from various organisms (17). Moreover, the dataset contains theoretical predictions from 62 methods for all of the experimental measurements. ProteinGym v1.0.0 includes 28 virus experiments containing numerous multiple and single point mutations (Table S2). Of these 28 experiments, only three of them (highlighted with red text color in Table S2) contain multiple mutations. When Spearman correlation performance is ordered by taxon, GEMME (7) and TranseptEVE family of models (L, M and S models) (18) are the best performers for mutational effect prediction of viral proteins (see the benchmarks at https://proteingym.org/benchmarks for the performance of the other 58 methods). In addition to the forementioned methods, we investigated performances of two recently developed methods called ESCOTT and iGEMME. We calculated all mutations for 28 viral proteins with these methods using the MSA and structure files provided by ProteinGym to make the comparisons sufficiently fair. Our results show that ESCOTT is the best method both for predicting effect of combined mutations (single+multiple mutations) (Figure 1 A) and single point mutations only (Figure 1 B) while iGEMME becomes the second best method. ESCOTT and iGEMME are very fast and they can be run on CPUs compared to TranceptEVE requiring a recent generation of GPUs.

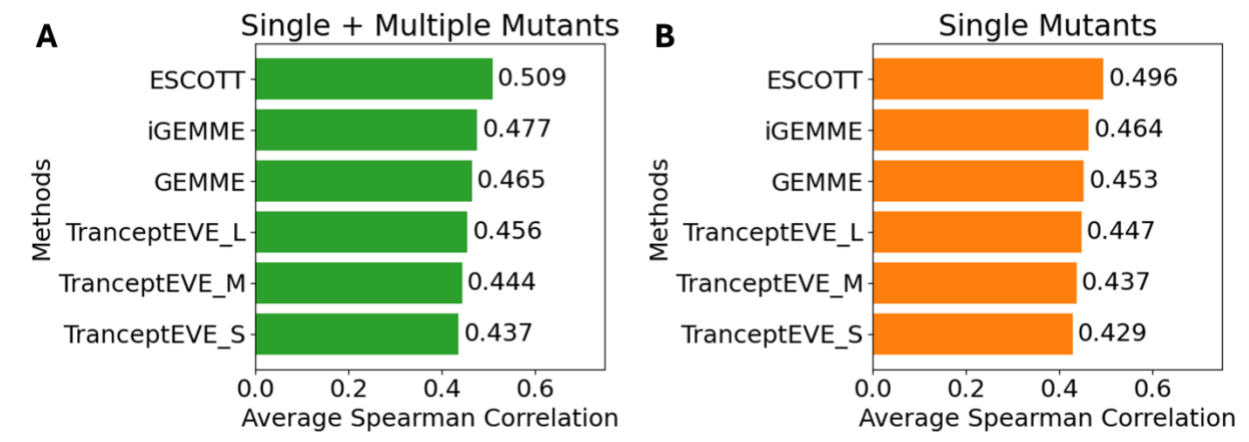


Figure 1. Performance comparison of ESCOTT and iGEMME with the other four methods. A) Average Spearman correlation performance of the best six methods over 28 deep mutational scanning experiments of viral proteins from ProteinGym v1.0.0 dataset containing both single and multiple point mutations. B) Average Spearman correlation performance of the best six methods over 25 deep mutational scanning experiments of viral proteins from the same dataset containing only single point mutations.

Quality of MPXV Protein Structures

Both ESCOTT and iGEMME require an MSA file as input. ESCOTT utilizes structural information coming from PDB format files in addition to evolutionary information coming from an MSA file. Since quality of structure prediction may affect the mutational landscape calculations, it is important to check it. Due to this reason, we investigated if there are experimental structures of MPXV proteins. We obtained experimental structures of A42R profilin-like protein (PDB ID: 4qwo) (19) and DNA polymerase (PDB ID: 8wpe) (20) of MPXV from Protein Databank (21). We superimposed them with the Colabfold predicted structures of OPG171 and OPG071 using Pymol, respectively (Figure 2 A and B). RMSD was 0.36 Å for A42R profilin-like protein and it was 1.89 Å for DNA polymerase, which suggests a good prediction. pLDDT is another metric used to check the quality of predictions (8, 9). When we inspect pLDDT values projected onto these two structures, we can see that majority of pLDDT values for amino acids are greater than 90 (shown as dark blue), indicating a good prediction. Since we wanted to find the quality of structure prediction for 171 proteins, we collected pLDDT for all amino acids in all of them and plotted them as a pie chart (Figure 2 C). We observed that 75.5% of pLDDT values are very high or high (dark blue or light blue). Moreover, we wanted to see if the predicted protein structures contain coils, which is an indicator of disorder in a protein. To do this, we determined secondary structure of all amino acids and plotted their distributions as a pie chart (Figure 2 D). We observed that coils constitute only 24.9% of the secondary structure types. In conclusion, available experimental structures, pLDDT values and secondary structure analysis suggest that the quality of the predicted structures are good for utilization in mutational effect predictions of all proteins encoded in the MPXV genome.

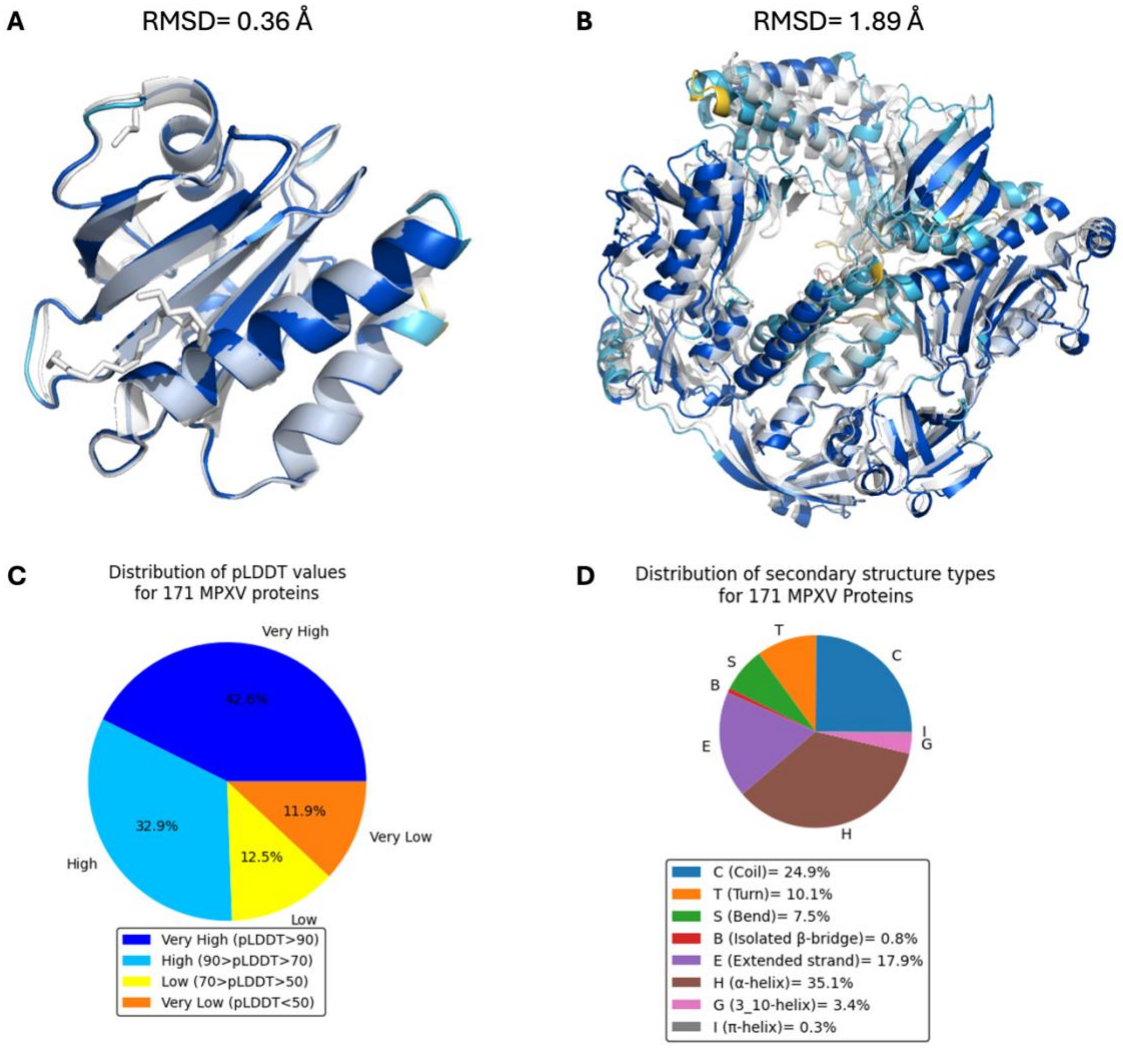


Figure 2. Protein structure quality investigation with pLDDT and secondary structure analysis. A) Experimental structure (PDB ID: 4qwo, transparent white cartoon representation) vs Colabfold generated OPG171 structure for profilin domain protein B) Experimental structure (PDB ID: 8wpe, transparent white cartoon representation) vs Colabfold generated OPG071 structure for MPXV DNA polymerase C) Pie chart for distribution of pLDDT values of 171 MPXV proteins. D) Pie chart for distribution of secondary structures for 171 MPXV proteins.

Single Point Mutational Landscape of MPXV Proteins

We calculated single point mutational landscapes of 171 MPXV proteins both with ESCOTT and iGEMME methods. We were able to obtain full mutational landscape of 171 proteins with ESCOTT and 169 proteins with iGEMME (two proteins, namely OPG076 and OPG158, failed to run with iGEMME due to insufficient sequence diversity in the MSA files). Raw and ranksorted score files as well as mutational landscapes in PNG image format for all proteins can be downloaded using publicly available zenodo repository from https://doi.org/10.5281/zenodo.13365666. Moreover,

the raw and ranksorted ESCOTT/iGEMME output files can be uploaded to PRESCOTT web server (http://prescott.lcqb.upmc.fr/visualisation.php) for an interactive visualization of the results.

As an example of the output, we present here the ranksorted impact of mutations to all possible amino acids for virus core cysteine proteinase, whose main role is to degrade peptides or proteins (Figure 3 A). Also, we provide the raw scores obtained with ESCOTT in Figure S1. iGEMME ranksorted scores (Figure S2 A) and raw scores (Figure S3) are also quite similar to ESCOTT scores for this protein. A visual inspection of the ESCOTT and iGEMME figures shows that the results are quite similar for the proteinase. Since we wanted to quantify the similarity for all proteins, we calculated Spearman correlation between ESCOTT scores and iGEMME scores for 169 proteins (Figure S4). The lowest Spearman value is greater than 0.75, which suggests high similarity and agreement of the two methods.

In addition to full mutational landscapes, we calculated AMMS for each position and added them to occupancy column of a PDB file for all proteins as explained in Average Gene Mutation Sensitivity (AGMS) subsection of the Methods section. We projected these per residue sensitivities onto the proteinase structure (Figure 3 B for ESCOTT and Figure S2 B for iGEMME). In many viral proteinases, there is a catalytic dyad (or triad) that function like a chemical scissor (22). Catalytic dyad of this proteinase is H241 and C328. It is clear both from the dark colors in Figure 3 B and the 2D plot given in Figure 3 C that the catalytic dyad amino acids are some of most sensitive positions to mutations.

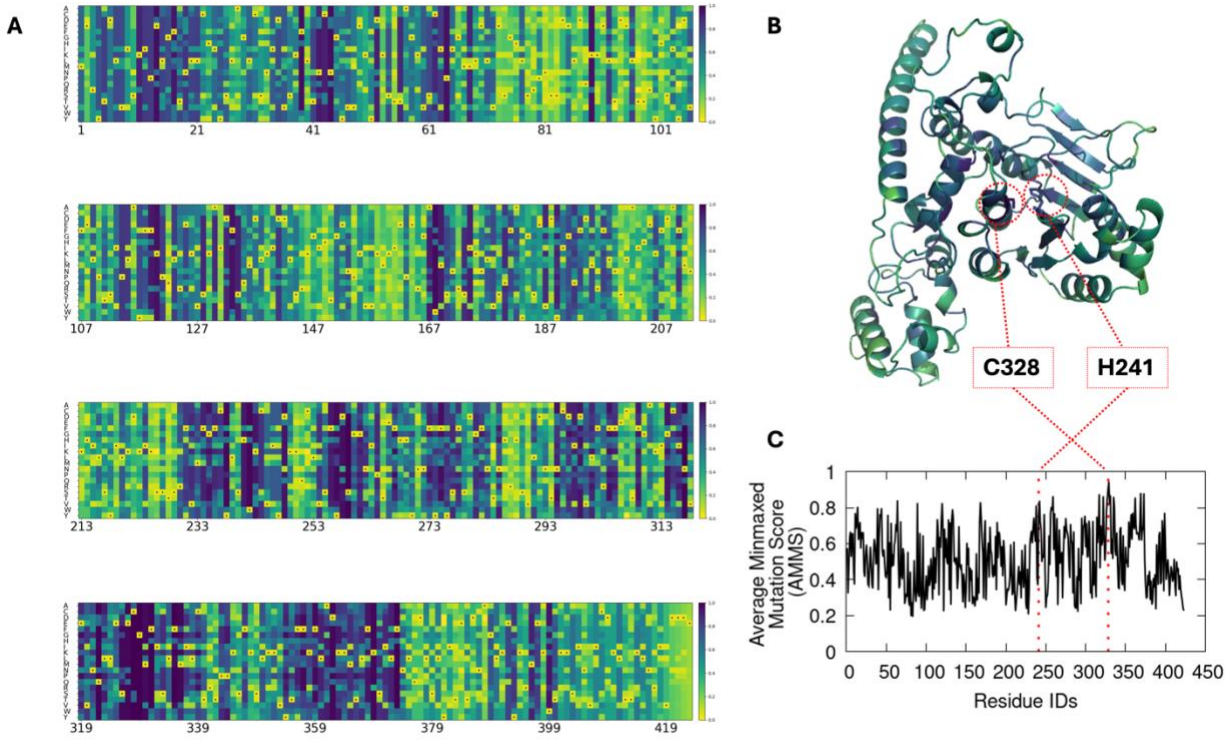


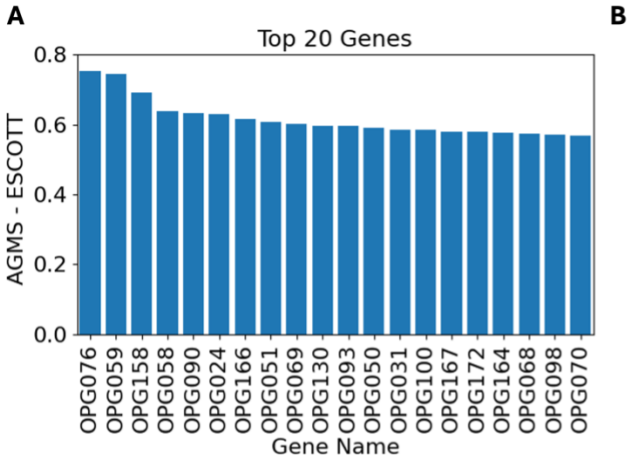
Figure 3. Mutational landscape of virus core cycteine proteinase (OPG083) and its projections obtained with ESCOTT. A) Single point mutational landscape of virus core cycteine proteinase (OPG083) calculated with ESCOTT is given in ranksorted scale (0, 1) and viridis colormap. Yellow color (values close to zero) indicates no effect while dark blue (values close to one) indicates a high effect for the mutations. Dotted squares denote the original amino acid in the protein sequence. B) Average Minmaxed Mutation Score (AMMS) for each position projected onto Colabfold generated protein structure (OPG083). Catalytic dyad (H241 and C328) of the proteinase is shown in licorice representation in red dotted circles and it has a dark color in viridis colormap, suggesting high importance of the positions. C) 2D plot of the Average Minmaxed Mutation Score (AMMS) vs residue IDs. Locations of H214 and C328 are highlighted with red vertical dashed lines.

Average Gene Mutation Sensitivity (AGMS)

We calculated AGMS using ESCOTT results for 171 proteins and we selected 20 genes with the highest scores (Figure 4 A). We found that Spearman correlation of protein length with AGMS is -0.134, which suggests that there is (almost) no correlation between these two parameters (Figure S5). Six of the top 20 proteins -including the first OPG076 (Mature Virion (MV) membrane entry fusion complex (EFC) component)- are on the membrane of the virus, which constitutes the frontier of interactions with the host proteins. OPG076 is an important component because studies utilizing variants of EFC proteins have shown that viruses without these proteins exhibit significantly reduced infectivity and impaired membrane fusion. For instance, recombinant viruses deficient in specific EFC proteins demonstrated very low levels of membrane fusion when tested in vitro (23, 24). The second gene in the top 20 is OPG059, Cytochrome C oxidase. A recent

study on influenza virus highlights its role in viral replication (25). OPG158 (A32.5L) gene encodes a protein similar to the vaccinia virus A32 protein, which is involved in virion membrane assembly (26, 27). One study showed that repression of A32 gene results in noninfectious, DNA-deficient, spherical, enveloped particles in the vaccinia virus (28). Another study found that a vaccinia virus strain lacking the B14R ortholog (similar to A32.5L) had reduced virulence, suggesting that A32.5L may also impact monkeypox virus pathogenesis, depending on the infection route (26). It is interesting to note that a hypothetical protein (OPG166) protein appears in Figure 4 and this protein might be considered as a new drug target. To summarize, we can see that many proteins in the top 20 conduct important functions in viruses according to the literature and our analysis may reveal some unconsidered new target proteins.

Furthermore, we calculated AGMS using iGEMME scores for 169 proteins (Figure S6) and its correlation with protein length (Figure S7). Common genes determined both by ESCOTT and iGEMME are OPG059, OPG058, OPG090, OPG024, OPG166, OPG051, OPG069, OPG130, OPG093, OPG050, OPG167, OPG172 and OPG164. Finally, we plotted the bottom 20 genes both with ESCOTT and iGEMME containing the genes the least sensitive to mutations according to their AGMS scores (Figure S8 and Figure S9).



B

Gene	Product
OPG076	MV membrane EFC component
OPG059	Cytochrome C oxidase
OPG158	A32.5L
OPG058	Protein F14 (1)
OPG090	DNA-directed RNA polymerase 7 kDa subunit
OPG024	Retroviral pseudoprotease-like protein
OPG166	Hypothetical protein
OPG051	CPXV054 protein
OPG069	Myristoylated protein E7
OPG130	A5L protein-like
OPG093	Late transcription factor VLTF-1
OPG050	CPXV053 protein
OPG031	C4L/C10L-like family protein
OPG100	IMV membrane protein J1
OPG167	CD47-like protein
OPG172	Type-I membrane glycoprotein
OPG164	IEV transmembrane phosphoprotein
OPG068	IMV membrane protein E6
OPG098	Nucleic acid binding protein VP8/L4R
OPG070	Membrane protein E8

Figure 4. A) Top 20 genes with the highest average gene mutation sensitivity (AGMS) obtained from ESCOTT calculations. B) Gene names and their products for the top 20 genes.

Conclusions

We calculated entire single point mutational landscape of 171 MPXV genes using the state-of-the art ESCOTT and iGEMME methods. ESCOTT/iGEMME mutational landscapes can be used to asses potential impact of new MPXV protein variants. Furthermore, we calculated AGMS for all genes in the dataset. AGMS values can also be quite useful for drug and vaccine design studies since the proteins with high AGMS values can be used as potential drug/vaccine targets.

We should note that we limited our comparison with ProteinGym v.1.0.0 and the computational methods presented therein. Some recent methods like AlphaMissense were excluded due to three reasons (29): 1) Lack of the publicly available code 2) The aim of AlphaMissense was to calculate pathogenicity of human mutations rather than other organisms. 3) Requirement of specialized hardware such as high-end graphical processing units. Moreover, we are not aware of any deep mutational scanning dataset of MPXV proteins as of writing if this paper and therefore, we do not have a complete mutational dataset for MPXV proteins for a direct comparison with our data. Despite these limitations, we present these mutational landscapes calculated with two state-of-the-art methods due to urgency of the situation with the hope that it will be useful to scientific and clinical community.

Acknowledgement

The numerical calculations reported in this paper were partially performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources). We would like to thank them for providing us these excellent computational resources and service.

Data and Code Availability

All data generated for this study is available at https://doi.org/10.5281/zenodo.13365666.

For each one of the proteins, we have the following files:

1. Raw ESCOTT/iGEMME scores data file
2. Raw ESCOTT/iGEMME scores image in png format
3. Ranksorted ESCOTT/iGEMME scores data file
4. Ranksorted ESCOTT/iGEMME scores image in png format
5. PDB file containing pLDDT values in Bfactor column and AMMS values in occupancy column.

References

1. Karagoz A, Tombuloglu H, Alsaeed M, Tombuloglu G, AlRubaish AA, Mahmoud A, et al. Monkeypox (mpox) virus: Classification, origin, transmission, genome organization, antiviral drugs, and molecular diagnosis. J Infect Public Health. 2023;16(4):531-41.
2. Giorgi FM, Pozzobon D, Di Meglio A, Mercatelli D. Genomic and transcriptomic analysis of the recent Mpox outbreak. Vaccine. 2024;42(7):1841-9.
3. Yu Z, Zou X, Deng Z, Zhao M, Gu C, Fu L, et al. Genome analysis of the mpox (formerly monkeypox) virus and characterization of core/variable regions. Genomics. 2024;116(1):110763.
4. Zhang S, Li YD, Cai YR, Kang XP, Feng Y, Li YC, et al. Compositional features analysis by machine learning in genome represents linear adaptation of monkeypox virus. Frontiers in genetics. 2024;15:1361952.
5. Chakraborty C, Bhattacharya M, Sharma AR, Dhama K. Evolution, epidemiology, geographical distribution, and mutational landscape of newly emerging monkeypox virus. Geroscience. 2022;44(6):2895-911.
6. Tekpinar M, David L, Henry T, Carbone A. PRESCOTT: a population aware, epistatic and structural model accurately predicts missense effect. medRxiv. 2024:2024.02.03.24302219.
7. Laine E, Karami Y, Carbone A. GEMME: a simple and fast global epistatic model predicting mutational effects. Mol Biol Evol. 2019.
8. Mariani V, Biasini M, Barbato A, Schwede T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics. 2013;29(21):2722-8.
9. Guo HB, Perminov A, Bekele S, Kedziora G, Farajollahi S, Varaljay V, et al. AlphaFold2 models indicate that protein sequence determines both structure and dynamics. Sci Rep. 2022;12(1):10696.
10. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nat Meth. 2022;19(6):679-82.
11. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021.
12. Steinegger M, Meier M, Mirdita M, Vohringer H, Haunsberger SJ, Soding J. HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinformatics. 2019;20(1):473.
13. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.3r1. 2010.
14. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983;22(12):2577-637.
15. Kunzmann P, Hamacher K. Biotite: a unifying open source computational biology framework in Python. BMC Bioinformatics. 2018;19(1):346.
16. Saladi SM, Maggiolo AO, Radford K, Clemons WM. Structural biologists, let's mind our colors. bioRxiv. 2020:2020.09.22.308593.
17. Parigger L, Krassnigg A, Grabuschnig S, Gruber K, Steinkellner G, Gruber CC. AI-assisted structural consensus-proteome prediction of human monkeypox viruses isolated within a year after the 2022 multi-country outbreak. Microbiol Spectr. 2023;11(6):e0231523.

18. Notin P, Van Niekerk L, Kollasch AW, Ritter D, Gal Y, Marks DS. TranceptEVE: Combining Family-specific and Family-agnostic Models of Protein Sequences for Improved Fitness Prediction. bioRxiv. 2022:2022.12.07.519495.
19. Minasov G, Inniss NL, Shuvalova L, Anderson WF, Satchell KJF. Structure of the Monkeypox virus profilin-like protein A42R reveals potential functional differences from cellular profilins. Acta Crystallogr F Struct Biol Commun. 2022;78(Pt 10):371-7.
20. Wang X, Ma L, Li N, Gao N. Structural insights into the assembly and mechanism of mpox virus DNA polymerase complex F8-A22-E4-H5. Mol Cell. 2023;83(23):4398-412.e4.
21. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000;28(1):235-42.
22. Dougherty WG, Semler BL. Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. Microbiol Rev. 1993;57(4):781-822.
23. Kao CF, Liu CY, Hsieh CL, Carillo KJD, Tzou DL, Wang HC, et al. Structural and functional analyses of viral H2 protein of the vaccinia virus entry fusion complex. J Virol. 2023;97(12):e0134323.
24. Laliberte JP, Weisberg AS, Moss B. The membrane fusion step of vaccinia virus entry is cooperatively mediated by multiple viral proteins and host cell components. PLoS Pathog. 2011;7(12):e1002446.
25. He J, Huang H, Li B, Li H, Zhao Y, Li Y, et al. Identification of cytochrome c oxidase subunit 4 isoform 1 as a positive regulator of influenza virus replication. Front Microbiol. 2022;13:862205.
26. Chen NH, Li GY, Liszewski MK, Atkinson JP, Jahrling PB, Feng ZH, et al. Virulence differences between monkeypox virus isolates from West Africa and the Congo basin. Virology. 2005;340(1):46-63.
27. Ferrareze PAG, Pereira ECRA, Thompson CE. Genomic characterization and molecular evolution of human monkeypox viruses. Arch Virol. 2023;168(11):278.
28. Cassetti MC, Merchlinsky M, Wolffe EJ, Weisberg AS, Moss B. DNA packaging mutant: repression of the vaccinia virus A32 gene results in noninfectious, DNA-deficient, spherical, enveloped particles. J Virol. 1998;72(7):5769-80.
29. Cheng J, Novati G, Pan J, Bycroft C, Zemgulyte A, Applebaum T, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science. 2023;381(6664):eadg7492.