DOI: 10.55415/deep-2022-0031.v1

Main Manuscript for

Structure Prediction of the whole Proteome of Monkeypox variants

Liangzhen Zheng[1,2*], Jintao Meng[2,3*], Mingzhi Lin[1], Rui Lv[4], Hongxi Cheng[1], Lixin Zou[1], Jinyuan Sun[5], Linxian Li[6,7], Ruobing Ren[4,8#], Sheng Wang[1,9#]

[1] Shanghai Zelixir Biotech Company Ltd., Shanghai 200030, China
[2] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, 518055, China
[3] National Supercomputer Center in Shenzhen, Shenzhen, 518000, China.
[4] Shanghai Key Laboratory of Metabolic Remodeling and Health, Institute of Metabolism and Integrative Biology, Fudan University, Shanghai 200438, China
[5] CAS Key Laboratory of Microbial Physiological and Metabolic Engineering, State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, PR China
[6] Ming Wai Lau Centre for Reparative Medicine, Karolinska Institutet, Hong Kong, China;
[7] Innorna (HK) Co Ltd. 12W Science and Technology West Avenue, Hong Kong Science Park, Shatin, Hong Kong, China
[8] Shanghai Qi Zhi Institute, Shanghai 200030, China
[9] CAS Key Laboratory of Quantitative Engineering Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

[*] These authors contribute equally

[#] To whom correspondence should be addressed. Email: wangsheng@zelixir.com and renruobing@fudan.edu.cn

Data Availability

The structure models, PointSite results, and the structural alignment data are open to access through the link: https://www.zelixir.com/Monkeypox/index.html.

DOI: 10.55415/deep-2022-0031.v1

Acknowledgments

We thank Yantao Liang and Hongmin Wang from Ocean University of China to provide technical support in proteome sequence analysis.

Author Contributions

L.Z, S.W. and R.R. initiated the project. S.W. collected and analyzed all sequences. L.Z., J.M. and M.L. predicted all the protein models. J.M. developed a batch approach for large-scale protein prediction. L.Z. predicted the binding sites on the protein models. L.Z. performed structure alignment for the predicted models against PDB data. M.L., H.C. and LX.Z. developed the Web Service for the predicted data. L.L. proposed the basic idea of the two case studies. J.S. and L.Z. performed docking of the p37 protein with tecovirimat. S.W., L.Z. and R.R. analyzed p37 protein with tecovirimat and the A35R protein with the antibody. S.W. and R.R. supervised the project. L.Z., S.W. and R.R. prepared the manuscript.

Competing interests

The authors declare no competing interests.

Abstract

Recently, the monkeypox virus has begun to spread in many countries worldwide[1]. The genome sequence of the monkeypox virus, associated with the current variant leading to an outbreak, was reported. It is important to better understand the new variant for accelerating the development of vaccines and drugs. Here, we reported the structure predictions of the whole proteomes of the three monkeypox variants and annotated the potential small molecule binding regions of the proteins. Meanwhile, experimentally determined structures with high similarity to monkeypox proteins are vetted through structure-alignment algorithm. Our work would help to accelerate the development of vaccines and drugs.

Keywords

DOI: 10.55415/deep-2022-0031.v1

Monkeypox, Structure prediction, Function annotation, Structure Alignment

DOI: 10.55415/deep-2022-0031.v1

Introduction

Monkeypox virus is a rare viral zoonotic orthopoxvirus that causes a disease with symptoms similar, but less severe, to smallpox. It can be transmitted through body contact, internal mucosal surfaces, or contaminated objects. With the eradication of smallpox in 1980 and the subsequent cessation of vaccination against smallpox, monkeypox became one of the most severe poxviruses. With an incubation period from 5 to 21 days, infection of monkeypox leads to fever, swollen lymph nodes, and an extensive characteristic rash. The documented mortality rate is between 0 and 11% and has been higher among young children[2]. Except for preventing monkeypox by avoiding a primary animal-to-human transmission, vaccination could be an effective means against monkeypox infection. However, populations are more susceptible to monkeypox because of the termination of routine smallpox vaccination, which offered over 80% effective cross-protection against monkeypox[3]. Various compounds against the monkeypox virus are also under development[4].

There are two distinct clades identified: the West African clade and the Congo Basin clade, also known as the central African clade. Recently, the genome sequence of the monkeypox virus, associated with the current variant leading to an outbreak affecting multiple countries, was reported. The rapid phylogenetic analysis shows that the 2022 variant belongs to the West African clade and is most closely related to the variant from Nigeria in 2018. With the increase of monkeypox cases worldwide, it is of great significance to better understand the new variant for accelerating the development of anti-monkeypox vaccines and drugs.

Results and Discussion

Firstly, we collected the genetic sequences of the well-characterized monkeypox virus variants, 1996 Congo virus strain (Zaire-96-I-16, ID: NC_003310.1), 2018 West African strain (MPXVUK_P3, ID: MT903345.1), and 2022 West African strain MPXV_USA_2022_MA001, ID: ON563414.3), from NCBI databank to generate the whole proteome datasets. We followed the NCBI databank's open reading frame (ORF) to extract 191, 190, and 190 proteins from 1996, 2018, and 2022 strains, respectively. Then we used AF2-Batch, which is the batch-mode AlphaFold2 framework, to predict the 3D structure models of all proteins we acquired (Figure 1A). In brief, we

DOI: 10.55415/deep-2022-0031.v1

reimplemented the AlphaFold2 structure prediction protocol[5] by firstly decomposing the computation workflow into multiple sequence alignment, end-to-end inference, and structure refinement and parallelizing the calculations with MPI coding on Slurm based supercomputational infrastructures. Meanwhile, we rewrote the end-to-end structural module along with the TensorFlow backend to avoid multiple compilation of the JAX library. It enables over 10000 structure predictions per day on an A100 GPU workstation with ten 50-cores CPU nodes, around ten times the speed of the original AlphaFold2 pipeline. AF2-Batch improved the capacity to quickly predict a large number of protein structures, including genome-to-proteome functional studies and systematic mutated protein structure prediction. Aimed at the recent emergence of monkeypox cases, we immediately released the structure models at the website for free use facilitate further studies (see Data availability part).

After completing the protein structure predictions, we implemented the deep PointSite model[6] to annotate the potential binding regions for small molecules on protein surfaces (Figure 1B). Based on the top-ranking structure model, the models provided the possibility that each atom of the protein composes the binding region of small molecules. The results are also released for public use on the website. Here, we chose one of the well-characterized pox proteins, P37, to present the PointSite result. P37 homolog protein, which plays a central role in forming the enveloped viral particle in the smallpox virus, is a validated target for anti-poxviral medication (Figure 1D). The closer the value to 1, the more likely the atom is included in the binding region. FDA approved Tecovirimat as the first anti-poxviral drug in 2018[7] (Figure 1E). However, the detailed recognition mechanism of Tecovirimat on P37 is unclear. We used PointSite to predict the potential binding site of monkeypox P37. Then we docked Tecovirimat in the putative pocket (Figure 1F). It is shown that Tecovirimat fits the pocket well, and the predicted binding energy by AutoDock Vina[8] is about -8.0 kcal/mol (Figure 1G). This algorithm may help us quickly select proteins with possible small molecule binding sites for further drug development targeting other poxviral proteins.

DOI: 10.55415/deep-2022-0031.v1

To better study the conservative characteristics of monkeypox proteins, we generated structural alignments, which showed protein lists with similar structures against a subset of the Protein Data Bank database (PDB70)[9] for each monkeypox protein (Figure 1C). The structure-based protein alignment algorithm tool, DeepAlign[10], was applied to rank the similarity according to the DeepScore. This function may help us annotate unknown proteins' functions by structural similarities. Here, we used the protein A35R to present the result (Figure 1H). The structure alignment list of A35R, particularly the globular domain, shares a high similarity with the A33R of the vaccinia virus (PDB ID: 4LQF) (Figure 1I)[11]. This structure was reported to show the A33R protein complex with an antibody A2C7 (Figure 1J). A33R is a well-known extracellular enveloped virus (EEV)-specific type II membrane glycoprotein. Since it plays a critical role in efficient EEV formation and helps the long-range viral spread in the host, A33 is a potential target for neutralizing antibody development targeting EEV. Similarly, A35R of the monkeypox virus is also a target for therapeutic antibody development to inhibit viral spread.

In summary, we predicted more than 600 structures and added functional annotations of proteins from monkeypox virus proteomes for public use. We provided profound annotations using the PointSite algorithm and labeled the small molecule binding regions with high confidence for all the 600+ predicted structures. Meanwhile, experimentally determined structures with high similarity to monkeypox proteins are vetted through the structure-alignment algorithm. We hope our work could accelerate the development of monkeypox vaccines, neutralizing antibodies, and therapeutic drugs.

References

[1] Kozlov M. Monkeypox goes global: why scientists are on alert. Nature. 2022 Jun;606(7912):15-16. https://doi.org/10.1038/d41586-022-01421-8.

[2] Jezek Z, Szczeniowski M, Paluku KM, Mutombo M. Human monkeypox: clinical features of 282 patients. J Infect Dis. 1987 Aug;156(2):293-8. https://doi.org/10.1093/infdis/156.2.293. [3] Fine PE, Jezek Z, Grab B, Dixon H. The transmission potential of monkeypox virus in human populations. Int J Epidemiol. 1988 Sep;17(3):643-50. https://doi.org/10.1093/ije/17.3.643.

[4] Smee DF. Progress in the discovery of compounds inhibiting orthopoxviruses in animal models.

Antivir Chem Chemother. 2008;19(3):115-24. https://doi.org/10.1177/095632020801900302.

[5] Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver,

D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. Nature 2021, 596 (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

[6] Yan X, Lu Y, Li Z, Wei Q, Gao X, Wang S, Wu S, Cui S. PointSite: A Point Cloud Segmentation Tool for Identification of Protein Ligand Binding Atoms. J Chem Inf Model. 2022 May 27. https://doi.org/10.1021/acs.jcim.1c01512.

[7] Grosenbach, D. W.; Honeychurch, K.; Rose, E. A.; Chinsangaram, J.; Frimm, A.; Maiti, B.; Lovejoy, C.; Meara, I.; Long, P.; Hruby, D. E. Oral Tecovirimat for the Treatment of Smallpox. New England Journal of Medicine **2018**, 379 (1), 44–53. https://doi.org/10.1056/NEJMoa1705688.

[8] Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. J. Comput. Chem. **2009**, NANA. https://doi.org/10.1002/jcc.21334.

[9] Steinegger, M.; Meier, M.; Mirdita, M.; Vöhringer, H.; Haunsberger, S. J.; Söding, J. HH-Suite3 for Fast Remote Homology Detection and Deep Protein Annotation. BMC Bioinformatics **2019**, 20 (1), 473. https://doi.org/10.1186/s12859-019-3019-7.

[10] Wang S, Ma J, Peng J, Xu J. Protein structure alignment beyond spatial proximity. Sci Rep. 2013;3:1448. https://doi.org/10.1038/srep01448.

[11] Matho, M. H.; Schlossman, A.; Meng, X.; Benhnia, M. R.-E.-I.; Kaever, T.; Buller, M.; Doronin, K.; Parker, S.; Peters, B.; Crotty, S.; Xiang, Y.; Zajonc, D. M. Structural and Functional Characterization of Anti-A33 Antibodies Reveal a Potent Cross-Species Orthopoxviruses Neutralizer. PLOS Pathogens **2015**, 11 (9), e1005148. https://doi.org/10.1371/journal.ppat.1005148.

DOI: 10.55415/deep-2022-0031.v1

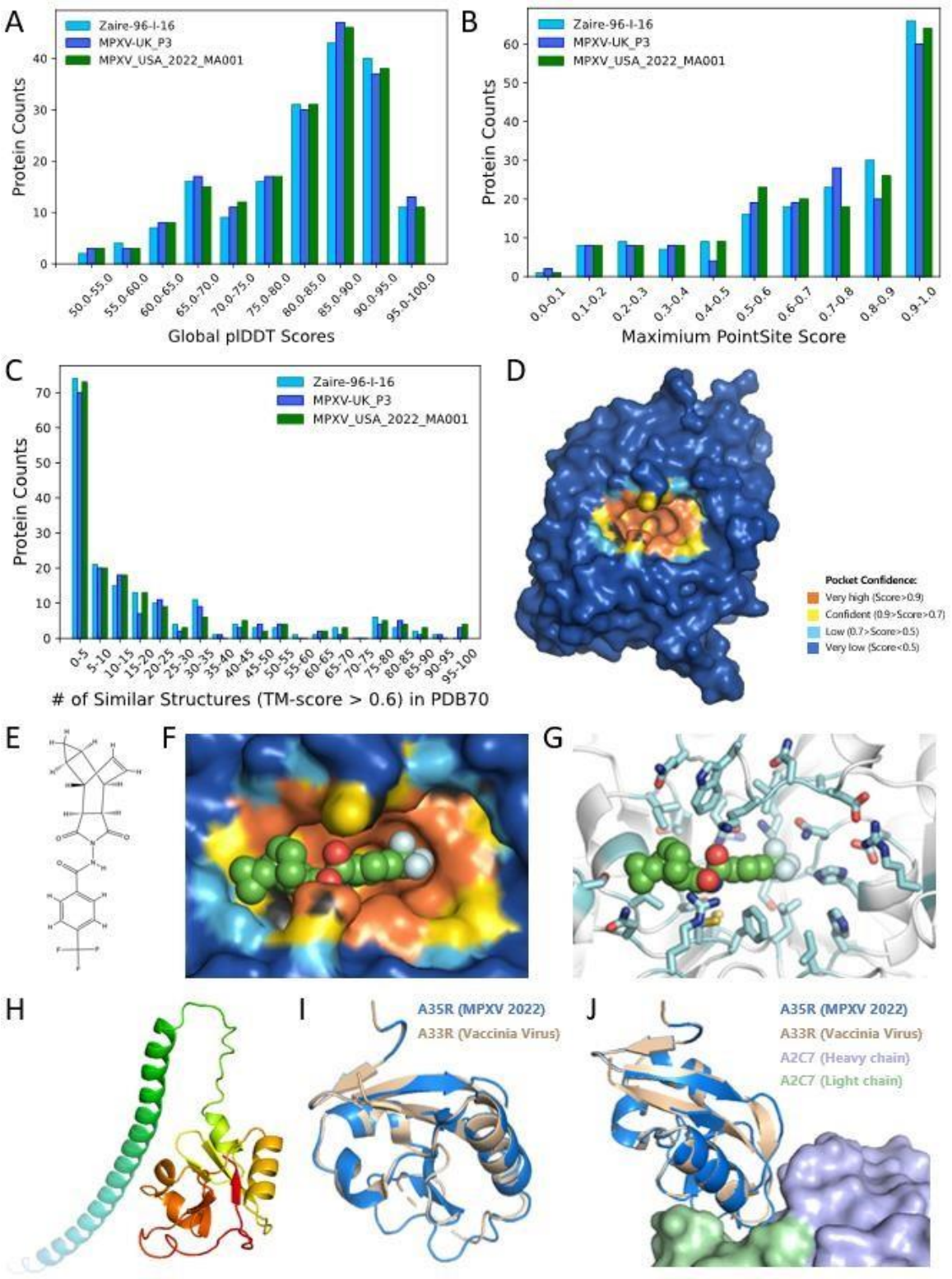

Figure 1. A) The plDDT score distribution of the top-ranked structure models of monkeypox virus 178 proteomes. The higher the score, the more confident the model is. B) The PointSite score

distribution of the top-ranked structure models of monkeypox virus proteomes. The higher the score,

DOI: 10.55415/deep-2022-0031.v1

the more possible there exists at least one small molecule binding pocket. The panel indicates that at least one-third of the proteins may have small molecule binding pockets for three strains. C) The number of similar structures (TM-score > 0.6) distribution of the top-ranked predicted models of monkeypox virus proteomes. D) The PointSite prediction of the P37 binding pocket. Different colors indicate the confidence of the prediction. The closer the value to 1, the more likely the atom is included in the binding region. E) The chemical structure of Tecovirimat. F) Tecovirimat fits the pocket of P37 well. Tecovirimat is shown in the green sphere. G) The detailed structure of the pocket of P37 with Tecovirimat. H) The predicted full-length structure of A35R by AF2-batch. The rainbow color shows the main chain from the N terminus to the C terminus. I) The structure alignment of A35R from monkeypox virus (marine color) 2022 strain and A33R from Vaccinia virus (wheat color). J) The recognition of A33R from Vaccinia virus and the antibody A2C7. The heavy chain of A2C7 is light blue, and the light chain of A2C7 is pale green.