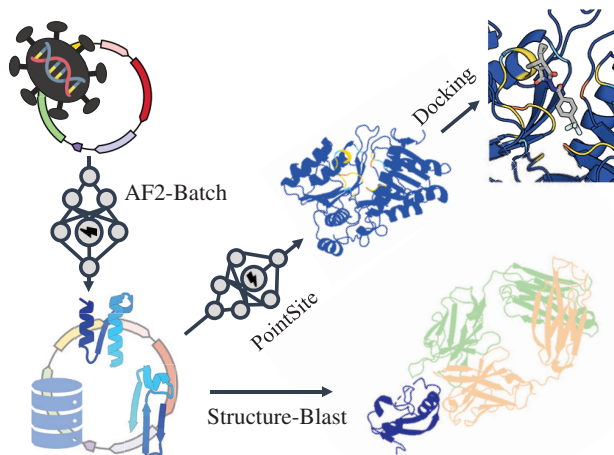


Structure prediction of the entire proteome of monkeypox variants

Graphical abstract



Highlights

- AF2-Batch predicted over 600 protein structures from three monkeypox virus strains.
- PointSite predicted all proteins' small-molecule-binding pockets to facilitate virtual compound screening.
- Similar structures in the Protein Data Bank were selected and aligned for conserved antigenic epitope analysis.

Authors

Liangzhen Zheng, Jintao Meng, Mingzhi Lin, Rui Lv, Hongxi Cheng, Lixin Zou, Jinyuan Sun, Linxian Li, Ruobing Ren, Sheng Wang

Correspondence

wangsheng@zelixir.com
(S. Wang);
renruobing@fudan.edu.cn
(R. Ren)

In brief

This work predicted and analyzed the protein structures of three monkeypox virus proteomes. Among them, some proteins contain the potential small-molecule-binding sites or conserved antigenic epitope for developing anti-monkeypox agents.

Short Communication

Structure prediction of the entire proteome of monkeypox variants

Liangzhen Zheng^{a,b,1}, Jintao Meng^{b,c,1}, Mingzhi Lin^a, Rui Lv^d, Hongxi Cheng^a, Lixin Zou^a, Jinyuan Sun^e, Linxian Li^{f,g}, Ruobing Ren^{d,h,*}, Sheng Wang^{a,i,*}

^aShanghai Zelixir Biotech Company Ltd., Shanghai 200030, China

^bShenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, 518055, China

^cNational Supercomputer Center in Shenzhen, Shenzhen, 518000, China

^dShanghai Key Laboratory of Metabolic Remodeling and Health, Institute of Metabolism and Integrative Biology, Fudan University, Shanghai 200438, China

^eCAS Key Laboratory of Microbial Physiological and Metabolic Engineering, State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

^fMing Wai Lau Centre for Reparative Medicine, Karolinska Institutet, Hong Kong, China

^gInnorna (HK) Co Ltd. 12W Science and Technology West Avenue, Hong Kong Science Park, Shatin, Hong Kong, China

^hShanghai Qi Zhi Institute, Shanghai 200030, China

ⁱCAS Key Laboratory of Quantitative Engineering Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

¹These authors contribute equally.

*Correspondence: wangsheng@zelixir.com (S. Wang); renruobing@fudan.edu.cn (R. Ren)

Received: 8 June 2022; Revised: 16 June 2022; Accepted: 17 June 2022

Published online: 28 June 2022

DOI: 10.15212/AMM-2022-0017

Recently, the monkeypox virus has begun to spread in many countries worldwide [1]. The genome sequence of the monkeypox virus variant responsible for the current outbreak has been reported, thus providing an important resource for better understanding the new variant and accelerating vaccine and drug development. Here, we report structure predictions of the whole proteomes of three monkeypox variants, with annotation of potential small-molecule-binding regions of the proteins. Experimentally determined structures with high similarity to monkeypox proteins were vetted through a structure-alignment algorithm. Our work should help accelerate the development of vaccines and drugs.

monkeypox, structure prediction, function annotation, structure alignment

Monkeypox virus is a rare viral zoonotic orthopoxvirus that causes a disease with symptoms similar to, but less severe than, smallpox. It can be transmitted through body contact, internal mucosal surfaces, or contaminated objects [2, 3]. With the eradication of smallpox in 1980 and the subsequent cessation of vaccination against smallpox, monkeypox became one of the most severe poxviruses. After an incubation period of 5–21 days, monkeypox infection leads to fever, swollen lymph

nodes, and an extensive characteristic rash [2, 4]. The documented mortality rate is between 0% and 11%, and has been reported to be higher among young children [5]. Beyond preventing monkeypox through avoiding primary animal-to-human transmission, vaccination may be effective against monkeypox infection [6]. However, some populations in which routine smallpox vaccination has been terminated are more susceptible to monkeypox, because such vaccination provides >80% effective cross-protection against monkeypox [7]. Various compounds against the monkeypox virus are also under development [8].

Two distinct clades have been identified: the West African clade and the Congo Basin clade, also known as the Central African clade [3]. Recently, the genome sequence of the monkeypox virus variant associated with the current outbreak affecting multiple countries has been reported. Rapid phylogenetic analysis has indicated that the 2022 variant belongs to the West African clade and is most closely associated with the variant from Nigeria in 2018. With the increase in monkeypox cases worldwide, better understanding of the new variant is important for accelerating the development of anti-monkeypox vaccines and drugs.

First, we collected the genetic sequences of the well-characterized monkeypox virus variants—the 1996 Congo virus strain (Zaire-96-I-16, ID: NC_003310.1), 2018 West African strain (MPXV-UK_P3, ID: MT903345.1),

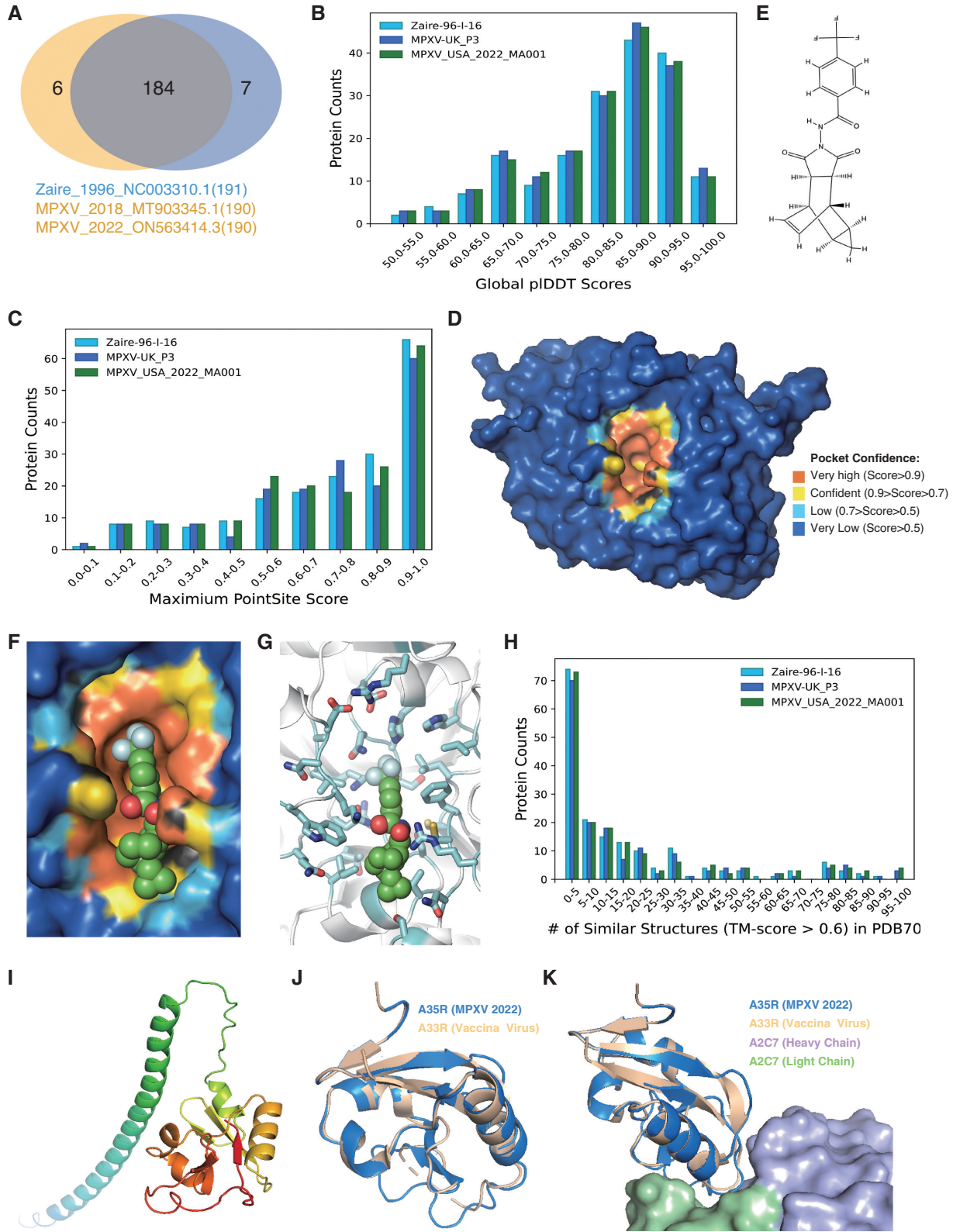


Figure 1 | A) Major differences among 1996 Congo virus strain (Zaire-96-I-16, ID: NC_003310.1), 2018 West African strain (MPXV-UK_P3, ID: MT903345.1), and 2022 West African strain MPXV_USA_2022_MA001, ID: ON563414.3) proteomes. B) pIDDT score distribution of the top-ranked structure models of monkeypox virus proteomes. Higher scores indicate higher confidence

Short Communication

of the model. C) The PointSite score distribution of the top-ranked structure models of monkeypox virus proteomes. Higher scores indicate greater likelihood of the existence of at least one small-molecule-binding pocket. The panel indicates that at least one-third of the proteins may have small-molecule-binding pockets in the three strains. D) 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 to be included in the binding region. E) Chemical structure of tecovirimat. F) Tecovirimat fits the pocket of P37 well. Tecovirimat is shown as a green sphere. G) Detailed structure of the P37 pocket with bound tecovirimat. H) Number of similar structures (TM-score > 0.6) and distribution of the top-ranked predicted models of monkeypox virus proteomes. I) Predicted full-length structure of A35R by AF2-batch. Rainbow coloration indicates the main chain from the N terminus to the C terminus. J) Structure alignment of A35R from the monkeypox virus (marine color) 2022 strain and A33R from vaccinia virus (wheat color). K) 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.

and 2022 West African strain MPXV_USA_2022_MA001, ID: ON563414.3)—from the NCBI databank to generate whole proteome datasets. We followed the NCBI databank's open reading frame to extract 191, 190, and 190 proteins from the 1996, 2018, and 2022 strains, respectively. The 2018 West African strain showed a very similar proteome to that of the 2022 strain, with correspondence among all 190 proteins. However, seven proteins in the 1996 strain do not exist in the 2018 and 2022 strains (Figure 1A). For example, a gene named BR-209 in the 1996 Congo virus strain encodes a full-length 326 amino acid (A.A.) protein, which is composed of an N-terminal fragment of 210 A.A. and a C-terminal fragment of 126 A.A. However, the West African strains contain a one-base insertion near the N terminus and a four-base deletion, thus causing two frameshifts yielding a new protein composed of an N-terminal 163 A.A. fragment and a C-terminal 132 A.A. fragment. Because BR-209 may function as an interleukin-1 β (IL-1 β) binding protein that prevents IL-1 β from interacting with the IL-1 receptor, the differences between BR-209 of the Congo versus West African strains of monkeypox may affect virulence [9]. We then used AF2-Batch, the batch-mode AlphaFold2 framework, to predict 3D structural models of all analyzed proteins (Figure 1B). In brief, we reimplemented the AlphaFold2 structure-prediction protocol [10] by first decomposing the computation workflow into multiple sequence alignment, end-to-end inference, and structure refinement, then 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. This pipeline enables more than 10000 structure predictions to be made per day on an A100 GPU workstation with ten 50-core CPU nodes, at a speed approximately ten times that of the original AlphaFold2 pipeline. AF2-Batch improved the ability to quickly predict many protein structures, including genome-to-proteome functional studies and structure prediction of systematically mutated proteins. Because of the recent emergence of monkeypox cases, we immediately released the structure models on the website <https://www.zelixer.com/Monkeypox/index.html>, allowing free use to facilitate further studies (see Data availability).

After completing the protein structure predictions, we implemented the deep PointSite model [11] to annotate the potential binding regions for small molecules on protein surfaces (Figure 1C). On the basis of the top-ranking structure model, the models indicated the likelihood of each atom of the protein to compose small-molecule-binding regions. The results have also been released for public use on the website. Here, we chose one of the well-characterized pox proteins, P37, to present the PointSite results. 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 to be part of the binding region. Tecovirimat, the first FDA approved anti-poxviral drug, was approved in 2018 [12] (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). Tecovirimat fits the pocket well, and the predicted binding energy by AutoDock Vina [13] is approximately -8.0 kcal/mol (Figure 1G). This algorithm may aid in rapid selection of proteins with possible small-molecule-binding sites for further drug development targeting other poxviral proteins.

To better study the conserved characteristics of monkeypox proteins, we generated structural alignments against a subset of the Protein Data Bank database (PDB70) [14] for each monkeypox protein, thus obtaining lists of proteins with similar structures (Figure 1H). The structure-based protein-alignment-algorithm tool DeepAlign [15] was applied to rank the similarity according to the DeepScore. This function may aid in annotation of unknown proteins' functions on the basis of structural similarities. Here, we used the protein A35R to present the results (Figure 1I). The structural alignment list of A35R, particularly the globular domain, shares high similarity with the A33R of vaccinia virus (PDB ID: 4LQF) (Figure 1J) [16]. This structure shows the A33R protein complex with an antibody A2C7 (Figure 1K). A33R is a well-known extracellular-enveloped virus (EEV)-specific type II membrane glycoprotein. Because it plays a critical role in efficient EEV formation and facilitates long-range viral spread in hosts, A33 is a potential target for development of neutralizing antibodies

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 extensive annotations by using the PointSite algorithm, and labeled the small-molecule-binding regions with high confidence for all 600+ predicted structures. Meanwhile, experimentally determined structures with high similarity to monkeypox proteins were vetted through the structure-alignment algorithm. We hope that our work will accelerate the development of monkeypox vaccines, neutralizing antibodies, and therapeutic drugs.

DATA AVAILABILITY

The structure models, PointSite results, and structural alignment data can be openly accessed through the following link: <https://www.zelixir.com/Monkeypox/index.html>.

ACKNOWLEDGEMENTS

We thank Yantao Liang and Hongmin Wang from Ocean University of China for providing 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 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 bound to tecovirimat, and A35R protein bound to 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.

REFERENCES

- [1] Kozlov M: Monkeypox Goes Global: Why Scientists Are on Alert. *Nature* 2022, 606:15–16. <https://doi.org/10.1038/d41586-022-01421-8>.
- [2] *Multi-country monkeypox outbreak: situation update*. <https://www.who.int/emergencies/disease-outbreak-news/item/2022-DON390>.
- [3] Petersen BW, Damon IK: Smallpox, Monkeypox, and Other Poxvirus Infections. In *Goldman-Cecil Medicine*. 26th edition. Philadelphia, PA: Elsevier; 2020.
- [4] CDC. Signs and Symptoms Monkeypox. May 11, 2015.
- [5] Ježek Z, Szczeniowski M, Paluku KM, Mutombo M: Human Monkeypox: Clinical Features of 282 Patients. *Journal of Infectious Diseases* 1987, 156:293–298.
- [6] McCollum AM, Damon IK: Human Monkeypox. *Clinical Infectious Diseases* 2014, 58:260–267.
- [7] Fine PEM, Jezek Z, Grab B, Dixon H: The Transmission Potential of Monkeypox Virus in Human Populations. *International Journal of Epidemiology* 1988, 17:643–650.
- [8] Smee DF: Progress in the Discovery of Compounds Inhibiting Orthopoxviruses in Animal Models. *Antiviral Chemistry and Chemotherapy* 2008, 19:115–124.
- [9] Chen N, Li G, Liszewski MK, Atkinson JP, Jahrling PB, Feng Z, et al.: Virulence Differences between Monkeypox Virus Isolates from West Africa and the Congo Basin. *Virology* 2005, 340:46–63.
- [10] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al.: Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* 2021, 596:583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
- [11] Yan X, Lu Y, Li Z, Wei Q, Gao X, Wang S, et al.: PointSite: A Point Cloud Segmentation Tool for Identification of Protein Ligand Binding Atoms. *J. Chem. Inf. Model.* 2022, 62:2835–2845. <https://doi.org/10.1021/acs.jcim.1c01512>.
- [12] Grosenbach DW, Honeychurch K, Rose EA, Chinsangaram J, Frimm A, Maiti B, et al.: Oral Tecovirimat for the Treatment of Smallpox. *New England Journal of Medicine* 2018, 379:44–53. <https://doi.org/10.1056/NEJMoa1705688>.
- [13] El-Hachem N, Haibe-Kains B, Khalil A, Kobeissy FH, Nemer G: AutoDock and AutoDockTools for Protein-Ligand Docking: Beta-Site Amyloid Precursor Protein Cleaving Enzyme 1(BACE1) as a Case Study. In *Neuroproteomics*. Volume 1598. Edited by Kobeissy FH, Stevens SM. Springer New York, NY: Methods in Molecular Biology; 2017:391–403. https://doi.org/10.1007/978-1-4939-6952-4_20.
- [14] Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J: HH-Suite3 for Fast Remote Homology Detection and Deep Protein Annotation. *BMC Bioinformatics* 2019, 20:473. <https://doi.org/10.1186/s12859-019-3019-7>.
- [15] 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>.
- [16] Matho MH, Schlossman A, Meng X, Benhnia MREI, Kaefer T, Buller M, et al.: Structural and Functional Characterization of Anti-A33 Antibodies Reveal a Potent Cross-Species Orthopoxviruses Neutralizer. *PLOS Pathogens* 2015, 11:e1005148. <https://doi.org/10.1371/journal.ppat.1005148>.