Biological and Biomedical Journal

Journal homepage: https://bbj.journals.ekb.eg

MOPO-Database of Predicted 3D Structure of Monkeypox Virus

Ahmed G. Soliman[1*], Mohamed M. Waly[2], Hafsa A. Gwidah[3], Yaser M. Hassan[4*]

[1]Biotechnology program, Faculty of Agriculture, Ain Shams ،Cairo, Egypt.
[2]Biotechnology program, Faculty of Science, Helwan University, Cairo, Egypt.
[3]Biotechnology program, Faculty of Science, Tanta University, Tanta, Egypt.
[4]Biotechnology program, Faculty of Science, Ain Shams University, Cairo, Egypt.

ARTICLE INFO

Received: 16/12/2024
Accepted: 27/1/2025

Corresponding author:
Ahmed G. Soliman, Ms.c
E-mail: ahmedgamal_soliman@agr.asu.edu.eg
Mobile: (+2) 106 557 5397

P-ISSN: 2974-4334
E-ISSN: 2974-4324
DOI: 10.21608/bbj.2025.338874.1055

ABSTRACT

Monkeypox (Mpox) virus has been a neglected tropical zoonotic disease for more than 50 years. Due to its rapidly spreading outbreak in more than 100 countries and territories, recently, becomes a focus of global attention. It is a contagious zoonotic disease caused by the *Orthopoxvirus*, endemic in Central and West Africa, and has symptoms similar to that caused by smallpox viruses. Symptoms of Mpox typically include fever, chills, rash, and lesions on the genitals or face. Hundreds of human clinical samples have undergone next-generation sequencing (NGS), with raw data deposited in public repositories, since the global outbreak in 2022. However, sequence analysis for in-depth investigation of virus evolution is still hampered by the lack of a complete protein database for the Mpox virus. Viral protein sequence data are essential for scientific prevention and epidemic management. Although many databases include protein structures, there is a significant gap in viral protein structures. To address this critical gap, the Mpox protein (MOPO) database was developed, which integrates 180 predicted 3D protein structures of the Mpox virus. Using the AlphaFold server, we predicted and validated protein structures with high template modeling scores (pTM). This database serves as a repository for all predicted protein isoforms of the Mpox virus. All predicted protein structures were uploaded to the database, and it became possible for any researcher or person interested in the field of viruses to obtain the structure of any protein, which is a quick and useful way for researchers to enhance clinical research for infected patients.

Keywords: AlphaFold, Mpox, Protein Structure Prediction, Viral Protein 3D Modeling, Virus database**.**

1. Introduction

Viruses are a broad group of biological entities that are mandatory cellular parasites. Most viruses do not share any common genes or gene families, unlike cellular organisms (Villarreal, 2008). This calls into question virus ancestry and evolution. Furthermore, the number of viruses on Earth is massive (more than 10^{31} particles) (Hendrix et al., 1999; Mushegian, 2020). It's predicted that they carry between 10^{8} and 10^{10} unique genes (Koonin et al., 2023). Most viral variety is currently unexplored, and nothing is known about the function of most viral genes that have been sequenced (Krishnamurthy and Wang, 2017). Viral genes not only encode a huge number of various functions, which results in a great diversity of viral genomes, but viral genes also constitute diverse clusters of genes with related functions (Kuchibhatla et al., 2014). For a long time, viruses have been a primary cause of many infectious diseases. Thus, understanding

the molecular knowledge of viruses is crucial to developing improved vaccines, designing novel anti-viral agents, and comprehending how viruses enter the body (Sharma et al., 2011). Viral sequence data can help in the traceability and evolution of viruses (Lu et al., 2020), the creation of viral nucleic acid detection reagents, vaccines, and medications (Lundstrom, 2015, 2018; Zhang and Holmes, 2020), and play an important role in clinical assistant diagnosis and scientific prevention, as well as control of epidemics by providing an important scientific basis.

In 1958, at the Statens Serum Institute (SSI) in Copenhagen, Denmark, two outbreaks of a pox-like disease in cynomolgus monkeys occurred. The novel virus discovered by the SSI was dubbed "monkeypox (Mpox) virus", since it infected the captive monkeys imported from Singapore. The first identification of the Mpox, which is nearly related to the Variola Virus, was a member of the *Orthopoxvirus* genus in 1959 (Arita and Henderson, 1968). In the following ten years, eight further outbreaks of Mpox struck captive monkeys imported to the Netherlands and the USA from Southeast Asia (Arita and Henderson, 1968). Close interaction between humans and the large monkey community interaction in the area through activities including hunting, cooking, and playing with live or dead animals (Ladnyj et al., 1972). For the first time in human history in 1970, a 9-month-old male newborn in the Democratic Republic of the Congo (DRC) region contracted Mpox (Ladnyj et al., 1972). Four human Mpox outbreaks were documented in Africa between 1970 and 2017 (Antunes et al., 2022). Since then, sporadic outbreaks have been reported in some countries in western and central Africa, mainly among children in rural rainforest areas (Sun et al., 2024). In 2003, the Mpox outbreak obtained worldwide attention when it extended outside the African continent, especially to the United States (Reed et al., 2004). In the 20 years that followed, a few cases were documented (Alakunle et al., 2020). As the 2017 Nigerian epidemic persisted in 2018 and 2019, many cases connected to travel from Nigeria to Singapore and the United Kingdom emerged (Antunes et al., 2022). Because of this, the World Health Organization (WHO) declared the dengue outbreak a “public health emergency of international concern” on July 23, 2022, a total of 91,123 confirmed dengue cases and 663 probable cases, including 157 deaths, have been reported to WHO from 115 countries and regions as of September 30, 2023. Concerns have been heightened because the number of patients infected with dengue has increased dramatically recently. By November 10, 2022, the unprecedented pandemic had spread to 110 nations and reached 79,231 cases.

Some rodents and monkeys in central Africa are the Mpox virus natural stewards. Early human infections are usually caused by contact with infected animals, which includes coming into contact with mucous membranes, bodily fluids, tissues, or undercooked meat. Additionally, scratches or bites from infected animals can spread the infection (Harris, 2022). It is thought that direct contact with respiratory droplets from infected people is the route of human-to-human transmission (Walter and Malani, 2022; Upadhayay et al., 2022; Beeson et al., 2023), besides vertical transmission (Adler and Taggart, 2022; Billioux et al., 2022). Unlike other outbreaks, this latest Mpox infection was the greatest known pandemic to occur outside of Africa. Previously, a Mpox infection could only be identified after contact with infected animals or travel to a Mpox-affected area (Durski et al., 2018; Lu et al., 2023; Wang et al., 2023). Nonetheless, in the present outbreak, human-to-human sexual contact rather than contact with infected animals or travel is linked to the majority of Mpox infections. Male homosexuals or bisexuals have been the victims of Mpox epidemics in the previous two years, according to reported cases. According to a research study, 41% of gay or bisexual men infected with HIV constitute 98% of these infected cases. Furthermore, the anal and vaginal areas accounted for 73% of the lesions that were seen (Del Rio and Malani, 2022). The symptoms of Mpox (Villarreal, 2008) extend from 14 to 21 days, while the incubation period is from 7 to 14 days (Nolen et al., 2016; Guarner et al., 2022). Due to the extended incubation time, it might be difficult to get an accurate diagnosis, which could delay receiving medical assistance, exacerbate the illness, and spread the virus

further (Accordini et al., 2023; Reda et al., 2023). Numerous institutes worldwide have sequenced hundreds of clinical samples since the 2022 epidemic and deposited them in the National Center for Biotechnology Information (NCBI) Virus repository.

Mpox virus is a large virus, the 197-kb DNA genome of which consists of two linear double-stranded strands within a dumbbell-shaped inner core and an outer envelope around it (Shchelkunov et al., 2001, 2002; Chen et al., 2005; Moss, 2013). Mpox encodes 209 predicted open reading frames (Choi et al., 2022). A highly conserved centric region in the Mpox genome has encoding genes for structural proteins, necessary enzymes, and genes involved in viral replication. Variable terminal ends of the genome encode genes related to host range, viral defense, and viral/host interactions (Shchelkunov et al., 2001, 2002; Chen et al., 2005; Moss, 2013). Hairpin loops covalently connect inverted terminal repeats (ITRs) at the distal extremities of the genome (Shchelkunov et al., 2001, 2002; Chen et al., 2005; Moss, 2013). Virion entrance into the host cell by fusion with the plasma or endocytic membrane marks the start of the viral life cycle. Replication occurs in the cytoplasm within membrane-bound foci known as viral "factories" upon uncoating. Each cell may produce about 10,000 genomes, but half end up enclosed in virions (Moss, 2013). Based on their geographic origins, two primary Mpox clades have been identified phylogenetically: the Congo Basin and West Africa, these clades are distinguished by their considerable disparities in case fatality rates (>10% and <1%, respectively), which can be attributed to a few different virulence genes (Shchelkunov et al., 2001; Antunes et al., 2022; Chen et al., 2005). These lineages have further sublineages, such as A.1, A.2, B.1, B.2, and more (Sklenovská and Van Ranst, 2018).

The immense genetic variety of viruses, which is reflected in a less than 30% amino acid sequence identity between previously identified and newly found viruses, presents difficulties for sequence-based annotation and classification (Terzian et al., 2021; Kuchibhatla et al., 2014). On the other hand, protein structures have a direct impact on function and are thus more conserved, which makes them useful for studying viral processes (Bamford et al., 2005; Illergård et al., 2009). Therefore, the availability of viral protein structures is important for viral annotation through the detection of structural similarities. Thanks to recent developments in computational protein structure prediction (Baek et al., 2021; Jumper et al., 2021), hundreds of millions of protein structures are now accessible via resources like the AlphaFold Server (Varadi et al., 2022, 2024). Samples in the NCBI Protein repository are a collection of sequences deposited from several sources, including translations from annotated coding regions in GenBank and RefSeq. To promote bioinformatics and assist experimental researchers working on the Mpox Virus, a complete data repository of Mpox's proteins was established.

In this study, the Mpox protein database (MOPO) was developed as a comprehensive database derived from a large set of viral proteins obtained from the NCBI Protein and predicted by the AlphaFold Server. The database contains information on the Mpox virus and the history of it. Furthermore, it consists of viral protein-derived data designed to serve as a comprehensive solution for the retrieval and analysis of viral proteins from all viral groups and species whose 3D structures haven't been solved. MOPO is freely available online at WWW.mopo.pro

2. Materials and Methods

2.1. Data collection and database construction

The Mpox virus's genome assembly (ViralProj15142) RefSeq (NC_003310.1) was selected from the NCBI genome repository under Taxid: 10244 (https://www.ncbi.nlm.nih.gov/datasets/genome/) and was chosen to download his proteins (n = 180) as FASTA files. This specific genome was selected for several reasons: first, it is one of the first genomes known to humanity during a Mpox outbreak; second, it is one of the most virulent genomes (Adler and Taggart, 2022; Antunes et al., 2022; Upadhayay et al., 2022; Beeson et al., 2023), and one of the most important reasons is that this sequence has been annotation submitted by NCBI RefSeq. The current taxonomic

nomenclature for MPVX with 9 basic ranks as of the ICTV 2021 release is as follows: Domain: Varidnaviria; Kingdom: Bamfordvirae; Phylum: Nucleocytoviricota; Class: Pokkesviricetes; Order: Chitovirales; Family: Poxviridae; Subfamily: Chordopoxvirinae; Genus: *Orthopoxvirus*; Species: *Monkeypoxvirus*. An electron micrograph of two Mpox virions revealed a dumbbell-shaped inner core (Fowotade et al., 2018). We have a representative genome of the Mpox virus, a reference genome (NC_003310), and ZAI-96-I-16 (MPV-ZAI) isolated from a patient during the 1996 outbreak in Zaire. The linear, double-stranded, 196,858-base-pair sequence encodes 190 open reading frames (ORFs) with a highly conserved central region and terminal variable regions enclosed in hairpin loops at the ends, as indicated by voltage gating chloride channels (CLC) (Shen-Gunther et al., 2022). We collected protein names, their accession numbers, their function, and their FASTA sequence files in order to use FASTA sequences in predicting proteins and using the name, accession number, and function to help researchers when searching or when choosing a specific protein and to help us collect all the data in an organized manner in the database to display it on the site in a simplified manner.

2.2. Predicting the 3D structure of proteins

Protein prediction techniques and tools have been developed, and there are many tools in this area, most notably the AlphaFold server. AlphaFold's online server (https: //alphafoldserver.com) has been used to predict the protein structures. The prediction protein sequence was obtained from the FASTA sequence file uploaded and compiled from each protein page on the NCBI and UniProt websites and used as input for each modeling task. Following the published AlphaFold pipeline, AlphaFold predicted the 3D structure of protein on the result page, and the result was available to download as a zip file via the download button. After the predicted 3D protein structure was created, it was downloaded, and each protein was assembled with its own FASTA file.

2.3 Protein file preparation and visualization

After the model prediction for each protein was completed, it was downloaded and decompressed to the crystallographic information file (.cis) format. This format is not commonly used or widely used, so it was necessary to convert the format from this format to the Protein Data Bank (.pdb) format to facilitate its use in different applications and save time for researchers if they downloaded multiple files simultaneously. We selected the protein model file with the highest predicted template modeling (pTM) score and converted it from the (.cis) format to the (.pdb) format. The protein files were prepared and visualized using BIOVIA Discovery Studio Visualizer.

3. Results

3.1 Selection of Mpox virus genome

A comprehensive database was established from the genome assembly of the Mpox virus (ViralProj15142; RefSeq: NC_003310.1), obtained from the NCBI genome repository (Taxid: 10244). This genome was selected due to its historical significance as one of the earliest and most virulent strains linked to Mpox outbreaks, particularly the ZAI-96-I-16 strain isolated during the 1996 outbreak in Zaire, and this one of the most studied and well-characterized genomes of the *Orthopoxvirus* genus, making it an ideal candidate for comprehensive protein analysis. A total of 180 proteins encoded by this genome were extracted from the NCBI genome repository in FASTA format. Each protein was meticulously annotated with its corresponding accession number, name, and functional description to create a structured dataset that researchers can easily access and reference. The database was constructed with the following goals in mind:

Comprehensive Functional Annotation: Each protein entry was linked to its biological function.

Accession Number Integration: Including accession numbers facilitates cross-referencing with external databases such as UniProt, PDB, and Pfam, providing a multi-layered dataset where users can access evolutionary and functional information beyond the current study.

Organized FASTA Sequence Files: The FASTA format for each protein sequence was retained to allow easy access for downstream computational analysis, including sequence alignment,

evolutionary studies, and structural prediction. This simplifies the workflow for researchers aiming to conduct homology modeling, domain analysis, or epitope mapping for vaccine development.

This well-organized database serves as a foundational resource for further computational and experimental studies on Mpox proteins. By combining structural and functional data, we provide a platform that allows researchers to efficiently select specific proteins for deeper investigation, whether for structural biology, drug discovery, or vaccine development purposes. The database is designed to be an open resource for the scientific community, with its potential applications extending into comparative genomics, viral evolution studies, and therapeutic target identification.

3.2 3D Structure prediction

To gain insights into the structural characteristics of Mpox virus proteins, the AlphaFold server was employed to predict the 3D structures of the 180 proteins extracted from the Mpox virus genome (NC_003310.1). AlphaFold, recognized for its cutting-edge accuracy in protein structure prediction, was chosen for this study due to its ability to predict high-resolution structures with limited experimental data. The protein sequences, obtained in FASTA format from NCBI, served as input for AlphaFold's deep learning algorithm. The AlphaFold pipeline first performed multiple sequence alignments (MSAs) using homologous sequences from multiple sequence databases to construct evolutionary profiles for each Mpox protein. Combined with structural templates, these MSAs were then utilized to predict accurate 3D models for each protein. AlphaFold uses a deep learning model to predict the distances between pairs of amino acids and the angles between chemical bonds in the protein. The aligned sequences help the neural network model in AlphaFold to make more accurate predictions about the distances and angles between amino acids. Including homologous sequences in MSA helps reduce noise and improve the signal for the deep learning model. The predicted distances and angles are used to assemble a 3D model of the protein. AlphaFold generated five models for each protein, ranked according to the pTM score (a measure of predicted structure accuracy generated by AlphaFold), an essential metric for assessing model confidence. The top-ranked prediction is displayed on the result page, and all samples, along with their associated confidences, are available to download as a zip file via the Download button, as shown in Figure 1. After the predicted 3D protein structure was created, it was downloaded, and selected Model 0 (The top-ranked model) (Abramson et al., 2024).

Figure 1. AlphaFold results for Rifampicin resistance protein

A comprehensive analysis of the predicted structures revealed that the vast majority of proteins exhibited high pTM scores, as shown in Figure 2, indicative of the reliability and accuracy of the models, which is considered a threshold for high-confidence predictions in structural biology.

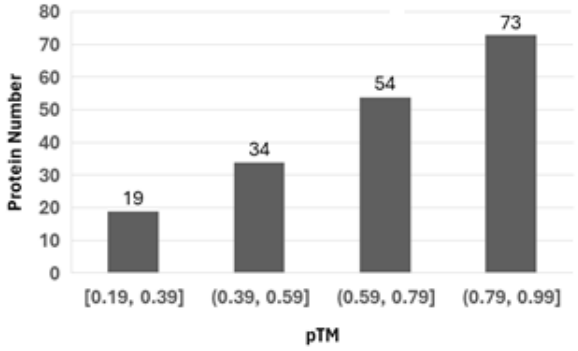

Fig. 2. Average protein number ranked according to the pTM score

For instance, the following examples highlight the structural accuracy of proteins:

Ribonucleoside-diphosphate reductase (2): Ribonucleoside-diphosphate reductase holoenzyme provides the precursors necessary for viral DNA synthesis. Allows virus growth in

non-dividing cells. Catalyzed the biosynthesis of deoxyribonucleotides from the corresponding ribonucleotides and exhibited a pTM score of 0.95. Hydroxysteroid dehydrogenase: This enzyme catalyzes the oxidative conversion of Delta5-ene-3-beta-hydroxy steroid and the oxidative conversion of ketosteroids. The 3-beta-HSD enzymatic system plays a crucial role in the biosynthesis of all classes of hormonal steroids. During viral infection, steroid production contributes to virulence by inhibiting the host inflammatory response, exhibiting a pTM score of 0.95 as shown in Figure 3.

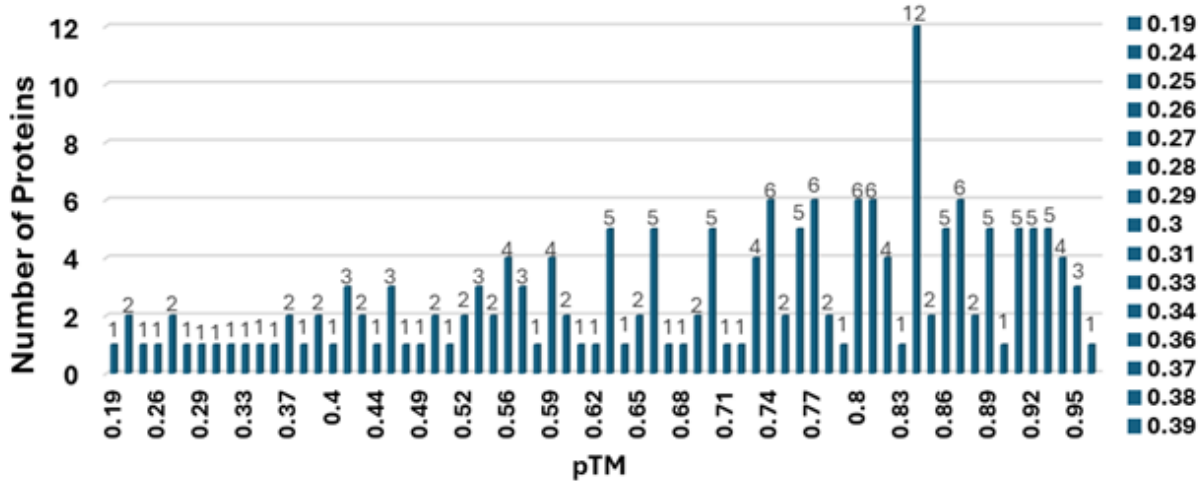

Fig. 3. Total pTM score for each AlphaFold generated protein

Each 3D model generated by AlphaFold was subsequently downloaded in a compressed (.cis) format, ensuring the integrity of the data while minimizing storage demands. The predicted structures, coupled with their respective FASTA sequences, offer a powerful resource for understanding the molecular mechanisms of Mpox virus proteins. The AlphaFold-generated models represent a significant advancement in the structural biology of Mpox proteins, providing unprecedented insights into their conformations, active sites, and potential functional domains. These structures lay the groundwork for future computational and experimental studies, including molecular docking, drug discovery, and vaccine design. The high-confidence predictions also offer a valuable resource for comparative structural analysis within the *Orthopoxvirus* genus, facilitating the identification of conserved regions that could serve as universal drug or vaccine targets.

3.3 Design of database:

A key outcome of this study was the development of a robust and user-friendly database that consolidates both structural and functional information for 180 proteins encoded by the Mpox virus genome. The database design was driven by the need for a comprehensive resource that facilitates easy access to Mpox viral protein sequences, structural models, and functional annotations, enabling researchers to perform targeted queries and analyses efficiently. Once the protein structures were predicted by AlphaFold, the resulting files were downloaded in the crystallographic information (.cis) format. Given that the. cis format is not widely supported by most molecular modeling and visualization software; it was necessary to convert these files into the more universally accepted Protein Data Bank (pdb) format. This conversion was critical to ensure compatibility with downstream applications, such as molecular docking, molecular dynamics simulations, and structural bioinformatics studies. The conversion process was with each protein model being transformed from. cis to .pdb format without compromising data integrity or structural accuracy. The database was designed with an intuitive and user-friendly interface to facilitate ease of access and navigation for researchers with varying levels of computational expertise; the following key elements were incorporated into the database design:

FASTA Sequence files: Each Mpox protein was represented by its corresponding FASTA sequence, allowing users to retrieve raw sequence data for alignment, phylogenetic analysis, or further modeling. These sequences were cataloged along with their associated accession numbers and protein names, ensuring clear traceability to the NCBI genome repository and other external databases.

Structural models in.pdb Format: After predicting the 3D structures of the Mpox proteins using AlphaFold, the models were converted from the. cis format to the more widely accepted.pdb format to enhance compatibility with common molecular visualization and analysis tools. These models were made available in the database, allowing researchers to download and analyze protein structures directly.

Functional annotations: Each protein entry in the database is linked to its predicted or experimentally validated function based on published literature and sequence homology comparisons.

Downloadable content: The database provides downloadable access to both FASTA sequences and 3D protein models in .pdb format. This feature enables researchers to quickly obtain data for integration into their computational pipelines for molecular docking, dynamics simulations, or evolutionary studies.

The design of the database allows for future scalability, enabling the inclusion of additional viral proteins from related poxviruses or newly sequenced Mpox strains.

The structural models and functional data available in the database provide a valuable resource for the scientific community, particularly for researchers focused on studying Mpox virus pathogenesis, viral protein function, and antiviral development. By centralizing this information, the database significantly reduces the time and effort required for protein characterization, offering a critical tool for accelerating research into the prevention and treatment of Mpox. MOPO, is a database for the Mpox that is designed to facilitate access and analysis of the proteins of the Mpox. The database is a helpful tool for researchers investigating Mpox, since it is designed to function as a centralized repository for 3D protein structures, functional annotations, and protein interaction data. By centralizing this information, the database significantly reduces the time and effort required for protein characterization, offering a critical tool for accelerating research into the prevention and treatment of Mpox.

4. Discussion

Mpox is a severe viral disease caused by a virus from the Poxviridae family, which contains double-stranded DNA. The first case of Mpox was discovered in 1970 when a 9-month-old newborn male in the Democratic Republic of the Congo was infected with the Mpox virus. The virus has since spread outside Africa to several countries to date. Furthermore, it is essential to understand the fast-progressing transmission of the virus infection and its genetic features of the evolving virus by studying the mutation rate. In this study, we collected proteins from the Mpox virus genome sequence dataset contained in NCBI Genome Assembly ViralProj15142, which was submitted and annotated with high quality by the underlying NCBI RefSeq database, so we have comprehensive knowledge of each gene. Complete genome annotations are particularly important for understanding the biology of the virus and developing any future treatments or vaccines we may seek. We found 180 coding sequences (CDS) proteins in the virus genome and downloaded their FASTA sequences, which play critical roles in viral replication, immune evasion, and pathogenesis. They have been presented and explained in high quality by the NCBI Database. Each protein identifier has been linked to the genome through the NCBI Protein Database with its accession number. We used the AlphaFold server to predict the 3D structures of the 180 encoded Mpox proteins, as shown in Figure 4. Based on amino acid sequences, the AlphaFold deep learning-based algorithm is extremely accurate in predicting protein structures. We developed three-dimensional (3D) models of Mpox proteins to facilitate thorough structural study. We have introduced MOPO, a database of 180 Mpox genome sequence proteins, which is designed to facilitate access to and analysis of Mpox proteins. This database was created due to the increasing

number of infected cases and the WHO interest in this deadly disease and issuing more than one resolution regarding it. MOPO is the ideal solution for researchers and those looking for a solution or vaccine to solve this deadly disease. This database is the solution to combat this disease. The prediction values for each protein are reviewed to ensure that the published protein is valid before it is submitted for use. The database is a useful tool for researchers investigating Mpox because it is designed to serve as a central repository for 3D protein structures and functions, as shown in Figure 5. It is available through the following link: www.mopo.pro

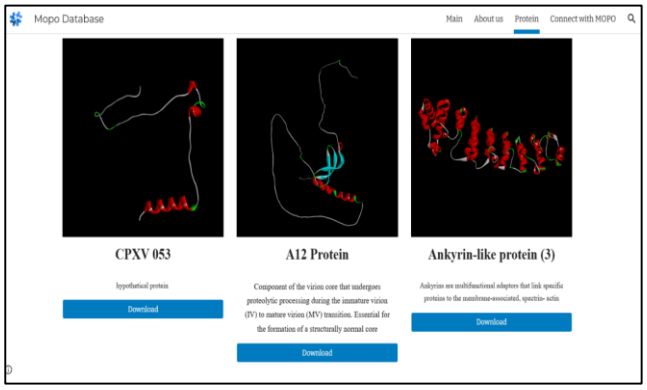

Fig. 4. Important elements into the database design

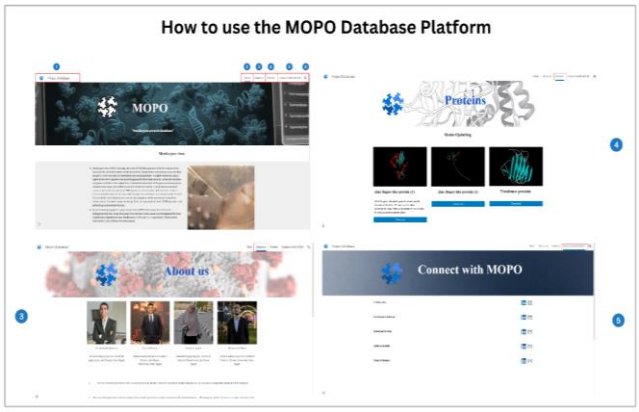

Fig. 5. A snapshot a MOPO web interface, showing detailed information with tabs and options, (1) Main tab: the name of the database, (2) the main of the database web interface platform, (3) About us tab: illustration of the information about the authors, (4) Protein tab: illustration of the in information about the proteins and the PDB files, (5) Connect With MOPO tab: illustration the contacts of the authors, (6) Search tab: providing the search in the web interface platform.

Conclusion

The Mpox virus is a large virus with a 197-kb DNA genome containing two linear double-stranded strands within a dumbbell-shaped inner core and an outer envelope around it. Data showed that the Mpox is still evolving and spreading from person to person. Over time, the Mpox virus has undergone mutations that lead to genetic variances in the circulating strains, often known as lineages. The properties of the virus, including its transmission and the intensity of symptoms in infected individuals, can be influenced by genetic variations, such as mutations in the Mpox protein structures. The problem is that the Mpox virus's protein structures are currently unavailable. In order to solve this problem, we are building a database because the protein structure is essential for comprehending the function of the virus. The goal of this project is to offer crucial insights that will facilitate further study and the creation of new treatments. In this study, a rapid and simplified means of profiling Mpox proteins will advance Mpox research to serve a wide range of research and clinical applications. Genome selection for the Mpox virus The NCBI Genome database contains the sequence genome database of the Mpox virus (Genome Assembly ViralProj15142), which was submitted and well annotated by the NCBI RefSeq reviewers. Protein IDs and FASTA sequences were collected. Then, the AlphaFold server was used to get predicted protein structures. Future plans include the addition of new strains from Mpox, and that's a wide variety of protein docking procedures that can greatly advance our knowledge of viral mechanics. This method not only broadens our understanding of the intricate ways in which viruses behave but also creates new opportunities for potential therapeutic approaches.

Author contributions: Ahmed G. Soliman provided constructive suggestions for this work and developed the database. Mohamed M. Waly extracted the data, developed the database, and wrote the manuscript. Hafsa A. Gwidah and Yaser M. Hassan contributed to data visualization and manuscript writing. All authors have read and agreed to the published version of the manuscript.

Funding: Not applicable.
Institutional review board statement: Not applicable.
Informed consent statement: Not applicable.
Data availability statement: The source code for the MOPO database website has been uploaded to www.mopo.pro
Conflicts of interest: The authors declare no conflict of interest.

5. Reference

Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung C-C O',Neil M, Reiman D, Tunyasuvunakool K, Wu Z, Žemgulytė A, Arvaniti E, … Jumper JM. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, *630*(8016), 493–500.

Accordini S, Cordioli M, Pomari E, Tacconelli E, Castilletti C. 2023. People with asymptomatic or unrecognized infection potentially contribute to monkeypox virus transmission. *The Lancet Microbe*, *4*(4), e209.

Adler H, Taggart R. 2022. Monkeypox exposure during pregnancy: What does UK public health guidance advise? *The Lancet*, 400(10362), 1509.

Alakunle E, Moens U, Nchinda G, Okeke MI. 2020. Monkeypox Virus in Nigeria: Infection Biology, Epidemiology, and Evolution. Viruses, 12(11), 1257.

Antunes F, Cordeiro R, Virgolino A. 2022. Monkeypox: From A Neglected Tropical Disease to a Public Health Threat. Infectious Disease Reports, 14(5), 772–783.

Arita I, Henderson DA. Smallpox and monkeypox in non-human primates. Bull World Health Organ. 1968;39(2):277-283.

Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, Millán C, Park H, Adams C, Glassman CR, DeGiovanni A, Pereira JH, Rodrigues AV, Van Dijk AA, Ebrecht AC, … Baker D. 2021. Accurate prediction of protein structures and interactions using a three-track neural network. Science, 373(6557), 871–876.

Bamford DH, Grimes JM, Stuart DI. 2005. What does structure tell us about virus evolution? Current Opinion in Structural Biology, 15(6), 655–663.

Beeson A, Styczynski A, Hutson CL, Whitehill, F, Angelo KM, Minhaj FS, Morgan C, Ciampaglio K, Reynolds MG, McCollum A M, Guagliardo SAJ. 2023. Mpox respiratory transmission: The state of the evidence. *The Lancet Microbe*, *4*(4), e277–e283.

Billioux BJ, Mbaya OT, Sejvar J, Nath A. 2022. Potential complications of monkeypox. The Lancet Neurology, 21(10), 872.

Chen N, Li G, Liszewski MK, Atkinson JP., Jahrling PB, Feng Z, Schriewer J, Buck C, Wang C, Lefkowitz EJ, Esposito JJ, Harms, T, Damon IK, Roper RL, Upton C, Buller RML 2005. Virulence differences between monkeypox virus isolates from West Africa and the Congo basin. Virology, 340(1), 46–63.

Choi CH, No JS, Kim JW, Lee M, Shin H, Choi MM, Yi H, Rhie, G. 2022. Complete Genome Sequence of Monkeypox Virus Strain MPXV-ROK-P1-2022 Isolated from the First Monkeypox Patient in the Republic of Korea. Microbiology Resource Announcements, 11(11), e00853-22.

Del Rio C, Malani PN. 2022. Update on the Monkeypox Outbreak. JAMA, 328(10), 921.

Durski, KN, McCollum, AM, Nakazawa Y, Petersen BW, Reynolds MG, Briand S, Djingarey MH, Olson V, Damon IK, Khalakdina A, 2018. Emergence of Monkeypox—West and Central Africa, 1970–2017.

Fowotade A, Fasuyi TO, Bakare RA. 2018. Re-emergence of monkeypox in Nigeria: A cause for concern and public enlightenment. *African* Journal of Clinical and Experimental Microbiology, 19(4), 307.

Guarner J, Del Rio C, Malani, PN. 2022. Monkeypox in 2022-What Clinicians Need to Know. JAMA, 328(2), 139.

Harris E. 2022. What to Know About Monkeypox. JAMA, 327(23), 2278.

Hendrix RW, Smith MCM, Burns RN, Ford ME, Hatfull, GF. 1999. Evolutionary relationships among diverse bacteriophages and prophages:

All the world's a phage. Proceedings of the National Academy of Sciences, 96(5), 2192–2197.

Illergård K, Ardell DH, Elofsson A. 2009. Structure is three to ten times more conserved than sequence-A study of structural response in protein cores. *Proteins:* Structure, Function, and Bioinformatics, 77(3), 499–508.

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 SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, … Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589.

Koonin EV, Krupovic, M, Dolja VV. 2023. The global virome: How much diversity and how many independent origins? *Environmental Microbiology*, *25*(1), 40-44.

Krishnamurthy SR, Wang D. 2017. Origins and challenges of viral dark matter. Virus Research, 239, 136–142.

Kuchibhatla, DB, Sherman WA, Chung BYW, Cook S, Schneider G, Eisenhaber B, Karlin D G. 2014. Powerful Sequence Similarity Search Methods and In-Depth Manual Analyses Can Identify Remote Homologs in Many Apparently "Orphan" Viral Proteins. Journal of Virology, 88(1), 10–20.

Lu J, Xing H, Wang C, Tang M, Wu C, Ye F, Yin L, Yang Y, Tan W, Shen L. 2023. Mpox (formerly monkeypox): Pathogenesis, prevention and treatment. Signal Transduction and Targeted Therapy, 8(1), 458.

Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, … Tan W. 2020. Genomic characterization and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. The Lancet, 395(10224), 565–574.

Lundstrom K. 2015. RNA-based drugs and vaccines. Expert Review of Vaccines, 14(2), 253–263.

Lundstrom K. 2018. Latest Development on RNA-Based Drugs and Vaccines. *Future Science OA*, *4*(5), FSO300.

Ladnyj ID, Ziegler P, Kima E. 1972. A human infection caused by monkeypox virus in Basankusu Territory, Democratic Republic of the Congo. Bull World Health Organ. 46(5):593-7

Moss B. 2013. Poxvirus DNA Replication. Cold Spring Harbor Perspectives in Biology, 5(9), a010199–a010199.

Mushegian AR. 2020. Are There 10^{31} Virus Particles on Earth, or More, or Fewer? Journal of Bacteriology, 202(9).

Nolen LD, Osadebe L, Katomba J, Likofata J, Mukadi D, Monroe B, Doty J, Hughes CM, Kabamba J, Malekani J, Bomponda PL, Lokota JI, Balilo MP, Likafi T, Lushima RS, Ilunga BK, Nkawa F, Pukuta, E, Karhemere, S, … Reynolds, MG. 2016. Extended Human-to-Human Transmission during a Monkeypox Outbreak in the Democratic Republic of the Congo. Emerging Infectious Diseases, 22(6), 1014–1021.

Reda, A, El-Qushayri, AE, Shah J. 2023. Asymptomatic monkeypox infection: A call for greater control of infection and transmission. The Lancet Microbe, 4(1), e15–e16.

Reed, KD, Melski JW, Graham MB, Regnery RL, Sotir, MJ, Wegner MV, Kazmierczak JJ, Stratman, EJ, Li Y, Fairley JA, Swain GR, Olson VA, Sargent EK, Kehl SC, Frace MA, Kline R, Foldy SL, Davis JP, Damon IK. 2004. The Detection of Monkeypox in Humans in the Western Hemisphere. New England Journal of Medicine, 350(4), 342–350.

Sharma OP, Jadhav A, Hussain A, Kumar MS. 2011. VPDB: Viral Protein Structural Database. Bioinformation, 6(8), 324-326.

Shchelkunov SN, Totmenin AV, Babkin IV, Safronov PF, Ryazankina, OI, Petrov NA, Gutorov VV, Uvarova, EA, Mikheev MV, Sisler JR, Esposito JJ, Jahrling PB, Moss B, Sandakhchiev, LS. 2001. Human monkeypox and smallpox viruses: Genomic comparison. FEBS Letters, 509(1), 66–70.

Shchelkunov SN, Totmenin AV, Safronov PF., Mikheev MV, Gutorov VV, Ryazankina OI,

Petrov NA, Babkin, IV, Uvarova EA, Sandakhchiev LS, Sisler JR., Esposito JJ, Damon IK, Jahrling PB, Moss B. 2002. Analysis of the Monkeypox Virus Genome. *Virology, 297*(2), 172–194.

Shen-Gunther J, Cai H, Wang Y. 2022. A Customized Monkeypox Virus Genomic Database (MPXV DB v1.0) for Rapid Sequence Analysis and Phylogenomic Discoveries in CLC Microbial Genomics. Viruses, 15(1), 40.

Sklenovská N, Van Ranst M. 2018. Emergence of Monkeypox as the Most Important Orthopoxvirus Infection in Humans. Frontiers in Public Health, 6, 241.

Sun Y, Nie W, Tian D, Ye Q. 2024. Human monkeypox virus: Epidemiologic review and research progress in diagnosis and treatment. Journal of Clinical Virology, 171, 105662.

Terzian P, Olo Ndela E, Galiez C, Lossouarn J, Pérez Bucio RE, Mom R, Toussaint A, Petit MA, Enault F. 2021. PHROG: Families of prokaryotic virus proteins clustered using remote homology. NAR Genomics and Bioinformatics, 3(3), lqab067.

Upadhayay S, Arthur R, Soni D, Yadav P, Navik, U, Singh R, Gurjeet Singh T, Kumar P. 2022. Monkeypox infection: The past, present, and future. International Immunopharmacology, 113, 109382.

Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, Žídek A, Green T, Tunyasuvunakool K, Petersen S, Jumper J, Clancy E, Green R, Vora A, Lutfi M, … Velankar S. 2022. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Research, 50(D1), D439–D444.

Varadi M, Bertoni D, Magana P, Paramval U, Pidruchna I, Radhakrishnan M, Tsenkov M, Nair S, Mirdita M, Yeo J, Kovalevskiy O, Tunyasuvunakool K, Laydon A, Žídek A, Tomlinson H, Hariharan D, Abrahamson J, Green T, Jumper J, … Velankar S. 2024. AlphaFold Protein Structure Database in 2024: Providing structure coverage for over 214 million protein sequences. Nucleic Acids Research, 52(D1), D368–D375.

Villarreal LP. 2008. Evolution of Viruses. In *Encyclopedia of Virology* (pp. 174-184). Elsevier.

Walter K, Malani PN. 2022. What Is Monkeypox? JAMA, 328(2), 222.

Wang Y, Leng P, Zhou H. 2023. Global transmission of monkeypox virus—A potential threat under the COVID-19 pandemic. Frontiers in Immunology, 14, 1174223.

Zhang YZ, Holmes EC. 2020. A Genomic Perspective on the Origin and Emergence of SARS-CoV-2. Cell, 181(2), 223–227.