



Elsevier has created a [Monkeypox Information Center](#) in response to the declared public health emergency of international concern, with free information in English on the monkeypox virus. The Monkeypox Information Center is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its monkeypox related research that is available on the Monkeypox Information Center - including this research content - immediately available in publicly funded repositories, with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the Monkeypox Information Center remains active.



Identification of core therapeutic targets for Monkeypox virus and repurposing potential of drugs against them: An *in silico* approach

Anshuman Sahu^{a,1}, Mahendra Gaur^{b,a,1}, Nimai Charan Mahanandia^c, Enketeswara Subudhi^d, Ranjit Prasad Swain^e, Bharat Bhusan Subudhi^{a,*}

^a Drug Development and Analysis Laboratory, School of Pharmaceutical Sciences, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, 751030, India

^b Department of Biotechnology, Punjabi University, Patiala, 147002, India

^c Division of Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, Pusa, New Delhi, 110012, India

^d Centre for Biotechnology, School of Pharmaceutical Sciences, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, 751030, India

^e Department of Pharmaceutics, School of Pharmaceutical Sciences, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, 751030, India

ARTICLE INFO

Keywords:

Monkeypox
mpox virus
Genome mining
Core proteome
Subtractive proteomics
HTVS
Repurposing
Molecular dynamics simulation
MM/PBSA

ABSTRACT

Monkeypox virus (mpox virus) outbreak has rapidly spread to 82 non-endemic countries. Although it primarily causes skin lesions, secondary complications and high mortality (1–10%) in vulnerable populations have made it an emerging threat. Since there is no specific vaccine/antiviral, it is desirable to repurpose existing drugs against mpox virus. With little knowledge about the lifecycle of mpox virus, identifying potential inhibitors is a challenge. Nevertheless, the available genomes of mpox virus in public databases represent a goldmine of untapped possibilities to identify druggable targets for the structure-based identification of inhibitors. Leveraging this resource, we combined genomics and subtractive proteomics to identify highly druggable core proteins of mpox virus. This was followed by virtual screening to identify inhibitors with affinities for multiple targets. 125 publicly available genomes of mpox virus were mined to identify 69 highly conserved proteins. These proteins were then curated manually. These curated proteins were funnelled through a subtractive proteomics pipeline to identify 4 highly druggable, non-host homologous targets namely; A20R, I7L, Top1B and VETFS. High-throughput virtual screening of 5893 highly curated approved/investigational drugs led to the identification of common as well as unique potential inhibitors with high binding affinities. The common inhibitors, i.e., bafefenterol, burixafor and eluxadolone were further validated by molecular dynamics simulation to identify their best potential binding modes. The affinity of these inhibitors suggests their repurposing potential. This work can encourage further experimental validation for possible therapeutic management of mpox.

1. Introduction

In the backdrop of ongoing COVID-19 pandemic, the detection of the first case of Monkeypox (mpox) in Europe on 7th May 2022 set the alarm for potentially another pandemic [1]. Although this has not affected society with the same magnitude, the rapid spread to more than 80 non-endemic countries has prompted the World Health Organization (WHO) to declare this outbreak a “Public Health Emergency of International Concern” on 23rd July 2022 [2]. Mpox is a reemerging zoonotic disease common to the Central and Western parts of the African continent [3]. However, the current outbreak which is the most dispersed and largest known to date, has fuelled the fear of the next COVID-19 like a

global pandemic in some circles of the international community.

Mpox virus, a double-stranded DNA virus belongs to the family *Poxviridae*, sub-family *Chordopoxviridae*, and genus *orthopox virus* that also includes the Variola virus which is responsible for smallpox [4]. Although its carriers are not limited to monkeys, its naming can be traced back to its first isolation from captive cynomolgus monkeys by a Danish lab in 1958 [5]. The first case of human mpox virus infection was documented in 1970 in the Democratic Republic of Congo, DRC (then Zaire) in an infant [6]. Mpox has since become endemic in DRC and has expanded to other Central and Western African countries [7]. The first human case and the subsequent outbreak of mpox outside Africa was documented in the USA in the year 2003 [8–11]. The human-to-human

* Corresponding author.

E-mail address: bbsubudhi@soa.ac.in (B.B. Subudhi).

¹ Equal Contribution.

transmission of mpox occurs primarily through skin lesions of infected humans or animals, respiratory droplets, body fluids and contaminated materials [7,12]. However, the transmission dynamics of the current outbreak in non-endemic countries remain cryptic as in some cases the detected cases do not have any associations with endemic regions [13]. Moreover, animal reservoirs are yet to be identified in these non-endemic regions [13]. Therefore, it has been suggested that the current epidemic in non-endemic regions might be a culmination of multiple mpox virus imports from endemic regions accompanied by silent, undocumented, cryptic inter-human transmissions [13].

Mpox virus has a double-layered cell membrane and a brick-shaped appearance. It exploits its double-stranded DNA to replicate inside the host's cytoplasm [14,15]. The genome of mpox virus is linear, approximately 197,000 bp long and includes hairpin termini. It consists of ≥ 198 non-overlapping open reading frames (ORFs). The central coding region is highly conserved and is flanked by variable regions with inverted terminal repeats on both sides [16–18]. Half of these ORFs are integral to viral replication and morphogenesis and are well conserved amongst other pox viruses [19]. The remaining half consists of “accessory” genes involved in immunomodulation, pathogenesis and host tropism. Many of them are yet to be functionally characterised [19]. As such these genes may be dispensable to the virus during replication [20]. As of now, pathogenesis and life cycle of mpox virus is very poorly understood. Nonetheless, limited *in vitro* studies suggest mpox virus can infect most mammalian cells [21,22]. The mature virion most likely attaches itself to host cell via glycosaminoglycans lining the host cell surface or modules of the extracellular matrix and external virion proteins [23,24]. Following attachment, the mature virion enters the host cell either by a low endosomal pathway or by merging with the plasma membrane in a pH-dependent manner to release the viral core into the cytoplasm [18,24]. Inside, the intracellular immature virion synthesises viral proteins to form an intracellular enveloped virion to eventually exit the host cell as an extracellular mature virion [15]. However, our knowledge of viral receptors that enables a virion to attach itself to a host cell remains elusive to this day [15].

The clinical presentation of mpox is somewhat similar to smallpox, albeit milder with three distinct phases [25]. The first phase (incubation phase) ranges from 7 to 14 days [25,26]. The second (prodrome) phase includes fever fluctuating between 38.5 °C and 40.5 °C, accompanied by muscle aches, headache, backache, chills and lymphadenopathy. This phase distinguishes it from smallpox and chickenpox. The third (rash) stage is characterised by macular rash that advances into papular, vesicular, and pustular stages resulting in crusts which eventually fall off [25]. The facial rashes can gradually progress across the body into the genitalia [27]. Complications include nutritional deterioration in patients with rashes in the oral cavity [15], permanent facial distortions upon the healing of facial lesions [28], loss of vision due to corneal infection [29], bacterial superinfections, and bronchopneumonia due to coinfection with influenza [30]. Rare serious complications include myocarditis, epiglottitis and septic shock resulting in mortality on account of the exaggerated immune response [31]. Although the infection is generally mild, immunocompromised individuals, children, pregnant women, elderly population and persons with comorbidity like HIV/AIDS may be prone to severe outcomes leading to death [13].

Two FDA approved vaccines initially developed as anecdotes against smallpox are currently being contemplated for effective immunisation against mpox. The first ACAM2000, is a next-generation vaccine similar to the discontinued Dryvax, with an efficacy of 85% [32]. MVA-BN (JYNNEOS in the U.S.) is the second next-generation vaccine developed with the modified Ankara strain (MVA) of the Vaccinia virus [33]. ACAM2000 is contraindicated in cases with atopic dermatitis, immunocompromised individuals, and pregnant women whereas the potency of MVA-BN remains to be validated in clinical trials [33,34]. Vaccinia immunoglobulin (VIG) has also been suggested from successful past experiences with smallpox [35]. Currently, tecovirimat (NCT00728689), cidofovir and brincidofovir (NCT01143181) are the

only available antivirals against mpox. Their potency against orthopox virus induced diseases in animal studies has been well documented. However, scientific data of their effectiveness in individuals infected with the currently circulating strains of mpox virus WA clade-II remains speculative [15,36]. Moreover, the development of resistance against tecovirimat and cidofovir has also been documented in orthopox viruses by virtue of acquisition of drug-resistant mutations in F13L or E9L [37]. Thus, taken together, the current therapeutic arsenal against mpox is almost deserted.

Drug repurposing offers a tangible solution, considering the immediate necessity and prohibitive time/cost involved in new drug development. Structure-based virtual screening of drug libraries is viable strategy to find the repurposing potential of drugs. However, the selection of a target and the availability of quality structures are keys to the success of this approach. With little information about the targets of mpox virus and their structure, virtual screening against mpox virus is a challenge. Nonetheless, the available genomes in public repositories are a goldmine of untapped possibilities. However, judicious selection of target is necessary for successful repurposing. Subtractive proteomics approach has generally been used to prioritise bacterial targets [38–42]. This approach has rarely been used to select viral targets as the number protein targets are usually very limited. Since, no data is available on the protein targets of mpox virus and relatively much higher number proteins targets are involved, it is worthwhile to use subtractive proteomics approach to prioritise the drug target and find potential inhibitors for repurposing against mpox disease (Fig. 1).

2. Materials and methods

2.1. Global dataset of publicly available mpox virus genomes and quality assessment

All publicly available mpox virus genome assemblies were downloaded from the National Centre for Biotechnology Information (NCBI) on 3rd June 2022.

2.2. Annotation of mpox virus genomes

To avoid artificial differences resulting from different genome annotation pipelines, the genome assemblies downloaded from GenBank were reannotated with the Prokka v 1.14.6 [43] with kingdom = Virus, genus = orthopox virus and species = ‘Monkeypox virus’.

2.3. Identification and re-annotation of mpox virus core genome

Clustering orthologs in viruses can ease the screening for drug candidates [44,45]. Therefore, to obtain the core proteome of mpox virus, the proteomes from all the genomes were clustered into groups of orthologs using the OrthoFinder v2.5.4 [46,47] with default parameters. The OrthoFinder surmises homologous regions and determines the orthogroups (OG's) by using a combination of the BLAST search and the Markov Cluster Algorithm (MCA) [47]. For this study, we defined core orthologs as ORFs present in all the strains of mpox virus. Therefore, only the ORFs shared by all (100%) the strains of mpox virus were considered as core orthologs. Further, to improve the functional annotations of all core proteins, we searched the consensus sequence of these proteins against the NCBI, UniProt and KEGG databases.

2.4. Subtractive proteomics and identification of druggable core proteome of mpox virus

We devised a suitable subtractive proteomics strategy [48] to pinpoint the highly druggable core proteins of mpox virus. We did so by screening out undesirable proteins from the core proteomic dataset of mpox virus by applying a set of rationally chosen parameters in the order listed below.

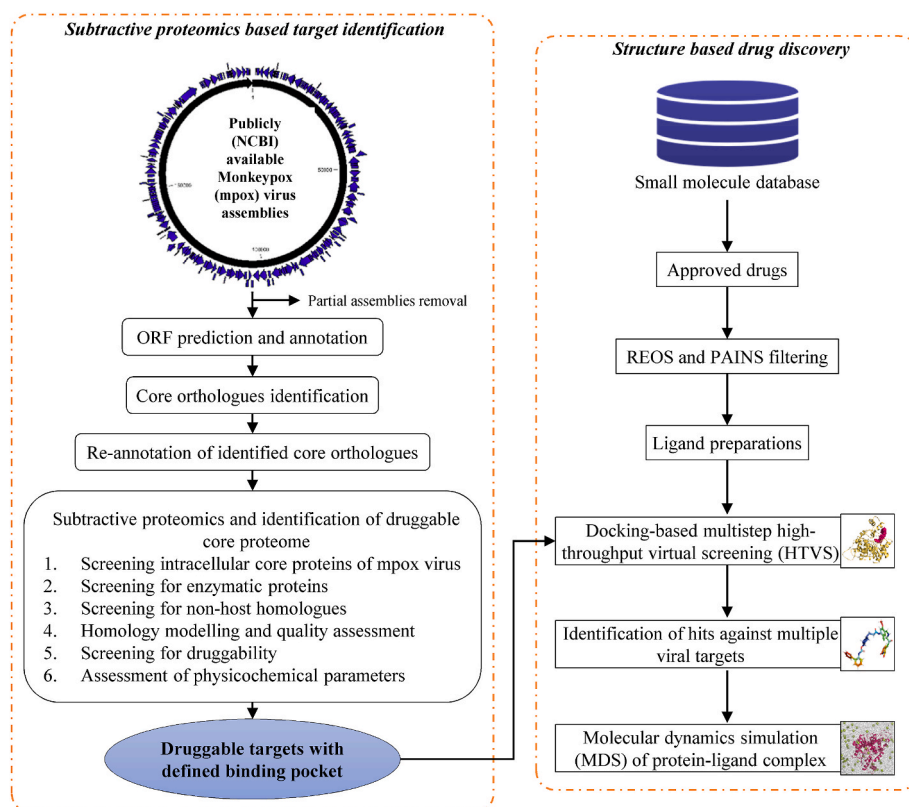


Fig. 1. Computational framework of the genome-to-drug approach used in this study. The first phase involves the identification of core orthologs/proteins from the multiple genomes of mpox virus by adopting subtractive proteomics. The second phase involves docking-based high-throughput virtual screening of a non-redundant library against the identified core orthologs for the identification of potential drug candidates followed by molecular dynamics simulation of drug-target complexes.

2.4.1. Screening intracellular core proteins of mpox

Intracellular proteins can be easily isolated and overexpressed for experimental studies as compared to membranous proteins. Taking this into view, the core proteins of mpox virus were first sorted as globular proteins and membrane-associated proteins by assessing their sub-cellular localization using the DeepTMHMM server v2.0 [49], CCTOP [50], MEMSAT-SVM [51] and TOPCONS [52]. A consensus of all the predictors was considered to interpret their sub-cellular localization. The predicted transmembrane proteins were filtered out and the globular proteins were selected for our next downstream analysis.

2.4.2. Screening for enzymatic proteins

These globular core proteins were then sorted as either enzymes or non-enzymes by using the DEEPre [53], an enzyme function predictor that employs deep learning to predict the Enzyme Commission Number (EC number) of the input protein sequence, EMBL EBI'S Enzyme portal (<https://www.ebi.ac.uk/enzymeportal/>) and scientific literature.

2.4.3. Screening of non-host homologues

The absence of shared homology with human proteins is a key feature of a viable target structure to achieve selectivity and avoid toxicity. To achieve this, the pool of mpox virus enzymes was piped into NCBI-BLASTp with an e-value of 0.0001 against the entire set of human proteomes (taxid: 9606). The list of final enzymes based on this criterion was selected and screened further to assess their druggability.

2.4.4. Homology modeling, model refinement and model quality assessment

The evaluation of a modeled 3D protein structure with high precision is a core element of computational structure prediction [54]. The rapid emergence of revolutionary protein structure prediction methods accompanied by highly efficient structure evaluation tools has paved new avenues to construct high quality protein models for computational

studies. In this study, the screened non-host homologous enzymes were modeled on the RoseTTAFold [55] server hosted by RosettaCommons (<https://rosetta.bakerlab.org>) before the assessment of their druggability. The RoseTTAFold uses a three-track neural network to fold the amino-acid sequences of a protein into a 3D model. The Molprobit [56] was then used for the optimization of stereochemistry and quality evaluation of the modeled proteins. The optimized proteins were further qualitatively evaluated using 3 independent programs including the PROCHECK [57], VERIFY3D [58] and ERRAT [59].

2.4.5. Assessment of druggability

Druggability is defined as the ability of a protein to undergo high-affinity tethering with drug-like molecules [60]. This is a key feature for prioritizing and validating putative targets in pathogens [61]. Thus, the druggability of the final list of 3D-modeled non-host homologous enzymes was assessed using the CavityPlus [62], DoGSiteScorer [63] and DeepSite [64]. We set an average pKd ≥ 6.5 , an average binding affinity of the associated pocket, for predicting druggable pockets without ligand using the CavityPlus [65]. The DoGSiteScorer identifies potential binding sites (called pockets). Proteins with pockets having druggability scores ≥ 0.8 were retained and the pocket with the highest druggability score was selected. The DeepSite [64] employs a machine learning algorithm based on deep convolutional neural networks (DCNNs) for predicting ligand-binding sites in proteins and provides cartesian coordinates of the centre of the identified binding pockets. A cut-off confidence score of 0.9 (90%) was used to predict the druggable pockets using the DeepSite. The consensus results of these three predictors were taken for predicting druggable enzymes.

2.5. Physicochemical characterization of druggable core proteins

Physicochemical properties of the screened druggable enzymes were

further assessed as these features have a critical role in the identification of potential druggable targets in the early stages of drug development. Physicochemical properties like isoelectric point (pI), molecular weight, aliphatic index, instability index, extinction co-efficient and residues accessibility were computed with the ProtParam tools (<https://www.expasy.org/resources/protparam>).

2.6. Collection of FDA approved/investigational drugs

For high-throughput virtual screening, we prepared a non-redundant dataset of FDA approved/investigational drugs. The Probes & Drugs portal (P&D) is a collection of high-quality bioactive compounds including drugs and probes with their targets and experimental values collected from different sources and updated on a daily basis [66]. We downloaded the approved/investigational drugs from the Probes & Drugs portal (P&D) on 13th July 2022. This dataset included molecules approved for clinical use, clinical testing, veterinary use or neurotraumatic applications.

2.7. In silico REOS and PAINS filtering

The SMILES of all the molecules were subjected to rapid elimination of swill (REOS) and pan-assay interferences (PAINS) filter using the RDKit [67], a python-based software (<https://rdkit.org>). Further, we implemented our in-house workflows using the Konstanz Information Miner (KNIME, v4.2.3; <http://knime.org>) followed by removal of redundancy.

2.8. Ligand preparation

All the non-redundant ligands were pre-processed using the OPLS_2005 force field in the LigPrep v5.3 module to obtain the accurate 3D energy minimized Lewis structures. Furthermore, EpiK v5.3 was used to generate the subsequent broad chemical, structurally diversified stereoisomers at pH 7.0 ± 2.0 .

2.9. Virtual screening of FDA approved/investigational small molecules in the highly druggable binding sites

The Glide program and its virtual screening workflow were applied to identify hits against the highly druggable target proteins by incorporating three docking protocols; high throughput virtual screening (HTVS), Standard Precision (SP) mode and Extra Precision (XP) mode [68,69]. A grid box of size 12 Å was assigned around the centroid of the binding sites predicted for each target structure. The van der Waals radii (1.00 Å) of the receptor along with partial atomic charge (0.25) was kept unchanged. Default docking parameters were used, and no constraints were included. For screening curated FDA library against all the targets, we used the xglide module of the Schrodinger suite at every step of the glide docking protocol. All the docked poses of each drug were transferred from HTVS to SP to filter out the false-positive results. The SP docked poses were then piped to XP docking where the false positives were further eliminated based upon ligand-receptor shape complementarity. These XP docked poses were then ranked according to a more stringent scoring function i.e., XP glide docking score (XP GScore).

2.10. Molecular dynamics simulations: system preparation and data generation

To evaluate the binding stability of the selected common top hits at the predicted highly druggable active sites of the non-host homologous core proteins (enzymes), we employed the molecular dynamics (MD) simulations of the protein-ligand complexes in the water environment.

Firstly, we prepared the conformational ensembles of all four apo proteins over the course of 400 ns MD simulation at a temperature of 300 K followed by simulation of the protein-ligand complexes over the course of 150 ns at the same temperature. The GROMACS suite v2021.1 [70] with charmm36-jul2021 force field was used to perform the MD simulation of apo proteins for the generation of ensembles and evaluation of the stability of bound ligands. The topology of proteins was generated by the in-built module of GROMACS and the topology of ligands was generated using the CHARMM General Force Field (CGenFF) web server (<https://cgenff.umaryland.edu>). Each system was solvated in a cubic periodic box with Tip3P charm-modified water model by keeping a 12 Å distance between the system and the water box. All the systems were neutralized by adding NaCl salt of concentration 0.15 M.

The structure of each system was minimized through a maximum of 50,000 steps using the steepest descent method of energy minimization. The equilibrations of each system were performed in two phases i.e. NVT (constant number of particles, volume and temperature) and NPT (constant number of particles, pressure and temperature) ensembles over the course of 5 ns each followed by MD production at constant temperature and pressure (1 bar and 300K). For ensemble simulation and MD production, each whole system was divided into two temperature-coupling groups of protein/protein-ligand and water-ions respectively. LINear Constraint Solver (LINCS) algorithm [71] was used to constraint all the bonds and isolated angles. Similarly, the velocity Verlet algorithm was used to integrate Newton's equations of motion. The temperature and pressure coupling on each ensemble was performed by V-rescale (modified Berendsen thermostat), Berendsen (NPT) and Parrinello-Rahman methods. However, no pressure coupling was used in NVT ensembles. For calculating the long-range (Electrostatic) interactions, the Particle Mesh Ewald (PME) method with a 1.2 nm cutoff was used. The output trajectories were recorded every 10 ps for subsequent analysis.

2.11. Molecular dynamics simulations: stability, clustering, essential dynamics and binding free energy analysis

Following MD data productions, the trajectories of all the systems was compared by analysis of root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), solvent accessible surface area (SASA) and amount of hydrogens bonds (HB) using the in-built module of the GROMACS. The analysis of positional changes of secondary structure elements (SSE) was accomplished by in-built module of biotite python library [72]. The dominant conformations of the trajectory data was inferred by clustering C α atoms RMSD of all the systems using TTClust [73]. The optimum cluster numbers was identified by elbow method. To understand the structural changes at C α atoms level over 150 ns time of MD simulation, MODE-TASK tool were used to study the principal component analysis (PCA) using singular value decomposition (SVD) of the cartesian coordinates as well as internal coordinates [74,75]. The first two eigenvectors (principal component 1 and 2) of PCA was plotted using the ggplot2 R package [76]. To compare the collective motions of C α atoms over the simulation time, the first essential principal component mode of PCA was represented as porcupine plots using Normal Mode Wizard (NMWiz), a ProDy python package based plugin in VMD [77,78]. The binding free energies and energies decomposition of key-interacting residues for all the complexes were calculated by AMBER's MMPBSA.py engine based gmx_MMPBSA tool [79,80] using molecular mechanics/Poisson-Boltzman surface area (MM/PBSA) model. The analysis module of the gmx_MMPBSA tool were used to infer and visualize the energies data. The average distance of bound ligands from the key-interacting residues, obtained from MM/PBSA analysis, were then calculated using the in-built gmx distance module of the GROMACS.

3. Result

3.1. The meaning of the open reading frame (ORF) and core proteins in this study

A clarification is necessary on what we mean by an ORF. This is because there is a slight discrepancy in how different authors like to define the term “Open Reading Frame” or “ORF”. In the bioinformatics community it is a standard practice to identify an ORF such that the evidence of translation is not required, a definition that virological community may be less acquainted with. For the purpose of this study, we defined an ORF having start codon, stop codon and no internal stop codon as a “protein-coding gene” which translates into a functional protein that contributes to viral replication, transmission, immune evasion or overall fitness. However, we also acknowledge that translation itself may not be a criterion significant enough for an ORF to be a protein-coding gene. We acknowledge that an act of translation may have a function even if the peptide produced is non-functional as in the case of regulatory uORFs [81]. Therefore, each predicted ORF can be important even if it is annotated as a hypothetical protein.

Likewise, the term “core proteins” used here frequently has been used to describe those proteins that are highly conserved across all the strains of mpox virus. This clarification is necessary as some members of virology community may be more accustomed to the term “core protein” as the proteins associated directly with the nucleocapsid.

3.2. A compendium of 125 mpox virus genomes was developed and explored

All the genome assemblies of mpox virus available as of 3rd June 2022, were retrieved from the NCBI. Our dataset included 24 newly assembled mpox virus genomes from the latest mpox outbreak described in non-endemic countries. In total, our final dataset comprised of 125 complete genome assemblies of mpox virus. The mean genome size of mpox virus was 196318 bp (~0.2 Mb). The highest number (275) of ORFs was identified in the strain mpox_FRA_2022_TLS67 (ON602722) whereas the strain mpox_Nig_2017_298464 (MG693724) harboured the least number of ORFs (193). The mean ORFs per genome was ~213. Due to lack of pre-existing annotations, approximately half of the total ORFs could be successfully annotated whereas the other remaining half remained unannotated or hypothetical. The detailed metadata regarding the sample isolation information and genome statistics like genome size, GC content (%), gene count and completeness etc have been summarised in [Supplementary Table S1](#).

3.3. Core proteome assessment identified sixty-nine core proteins from mpox virus genomes

Core proteins are major contributors to the replication and survival of a virus. Hence, developing novel therapeutics against such proteins could stop infectious pathways. Moreover, screening core proteins of an organism are advantageous as it helps to reduce the dependency on genome completeness [82]. In our core proteome assessment of 125 mpox virus strains, OrthoFinder assigned 26293 genes (99.4% of the total) to 294 ortho groups with 166 unassigned genes (0.6% of genes). Fifty percent of all genes were in orthogroups with 124 or more genes ($G_{50} = 124$) and were contained in the largest 101 ortho groups ($O_{50} = 101$). There were 69 ortho groups (core ortho groups) with all species present and 58 of these consisted entirely of single-copy genes. Two ortho groups were species-specific ortho groups with only 4 genes. The gene count in every orthogroup for each strain and strain-wise statistics is presented in [Supplementary Tables S2 and S3](#). Every set of the 69 core proteins was then aligned to create a consensus protein sequence of each ortholog. Thus, in doing so, we created a dataset comprising the core proteome of all the 125 mpox virus genomes. Unfortunately, many predicted core proteins remained unannotated for the reasons

mentioned previously. To overcome this problem, we reannotated every protein of our core proteomic dataset through a series of extensive manual curations. The detailed annotation of the core orthogroups dataset has been presented in [Table 1](#).

3.4. Subtractive proteomics successfully identifies four highly druggable, non-host homologous core proteins (enzymes)

The reannotated 69 core proteins of mpox virus were subjected to a set of multi-stage screenings to identify the druggable core proteins of mpox virus as shown in [Table 2](#). In the first stage, 56 globular proteins were identified based on their sub-cellular localization. These are most likely to be comprised of the proteins associated with the nucleocapsid, transcription factors and enzymes that may be critical to the virus life-cycle. In the second stage, we further screened out enzymes from the set of 56 globular core proteins. From the *in silico* assessment, we identified 23 core globular proteins which are most likely enzymes. Subsequently, in the third stage, the shared homology with human host proteins was assessed to limit the off-target issues. Eight enzymes namely, DNA-dependent RNA polymerase 132 kDa subunit (Rpo132), Phospholipase-D-like protein K4 (K4L), Ribonucleoside-diphosphate reductase small chain (OPG048), dual specificity protein phosphatase H1 (H1L), 3-beta-hydroxy-Delta(5)-steroid dehydrogenase (SALF7L), Cu-Zn superoxide dismutase-like protein (A46R), thymidylate kinase (TMK) and pseudokinase B12 (B12) shared 24.95%, 47.23%, 81.25%, 28.78%, 41.24%, 30%, 41.55% and 32.88% sequence identity with human proteins DNA-directed RNA polymerase II 140 kDa polypeptide (3J0K_B), 5'-3' exonuclease PLD3 (Q8IV08), Human ribonucleotide reductase subunit R2 (D6W4Z6), Dual specificity protein phosphatase 22 (6LVQ_A), 3 beta-hydroxysteroid dehydrogenase type 7 (Q9H2F3), Superoxide dismutase [Cu-Zn] (3GTV_A), Thymidylate Kinase (1E98_A) and Vaccinia-related kinase 1 (2LAV_A) respectively. These proteins were subsequently discarded as they exceeded the permitted threshold (e -value = 0.0001). The 15 remaining non-host homologous enzymes were processed for druggability studies. 12 out of these 15 enzymes shared excellent coverages and sequence similarities with the solved crystal structures of proteins from the PDB database. On the contrary, DNA polymerase processivity factor component A20 (A20R) shared 96.75% similarity with 6ZXP_A at very low coverage (28%) whereas crystal structures even remotely similar to core cysteine protease (I7L) and metalloendopeptidase G1 (G1L) were not available ([Table 1](#)). In the absence of experimentally derived high-resolution structures, modeled proteins present suitable alternatives to examine structural perturbations in proteins. Therefore, to overcome this problem we applied the hybrid, a top performer of CASP-5 and highly accurate deep learning-based modeling method the RoseTTAFold to model the 3D structures of all the 15 non-host homologous enzymes. The RoseTTAFold [55] employs a 3-track neural network to fold a protein sequence into a model. The accuracy and validation assessment of all the predicted structures were assessed through the MolProbity web server [56]. Out of these 15, the 7 enzymes successfully modeled by the RoseTTAFold were subjected to druggability assessments for the time being. Of these 7 core non-host homologous enzymes, only 4 enzymes namely the I7L, DNA topoisomerase 1B (Top1B), early transcription factor 70 kDa small subunit (VETFS) and A20R complied with the thresholds of druggability prediction tools. The druggability scores (DScore), simple score (SScore), binding affinity (pKd), volume and surface area associated with the predicted highly druggable binding site of the four proteins have been summarised in [Supplementary Table S4](#). The model quality statistics of these four druggable proteins are presented in [Table 3](#). Unfortunately, this study was impelled to abandon the druggability assessment of 8 out of these 15 enzymes (Group 5, [Table 2](#)). These 8 proteins will be assessed for their druggability in our future endeavors. Hence, by combining genomics and an extensive subtractive proteomic screening on the core proteomic dataset of mpox virus we finally identified 4 highly druggable, globular, non-host homologous core proteins

Table 1

Table summarising the reannotation of the 69 core proteins identified in 125 genomes of mpox virus.

Core_Og	Length (aa)	Similar to Uniprot ID	Protein name	Gene	% identity	alignment length	Mismatches	E-Value	Closest crystal structure (% Identity, % Coverage)
OG000	344	P42926	Serine proteinase inhibitor 2 (Serp-2) (Serp-2)	SPI-2	65.70	344	117	2.7E-149	1F0C_A (60.07, 89)
OG010	492	Q80DV6	DNA helicase; Transcript termination protein A18	A19R	97.16	493	13	0.0E+00	6JDE_A (21.76, 68)
OG011	1880	Q8QPZ7	Poxvirus B22R protein, putative transmembrane glycoprotein	B22R	86.3	1829	18	0.0E+00	
OG012	479	Q8QMZ9	Poly(A) polymerase catalytic subunit (EC 2.7.7.19)	PAPL	99.17	479	4	0.0E+00	3ER8_C (98.75, 100)
OG013	1164	Q8V4V3	DNA-dependent RNA polymerase 132 kDa subunit,	Rpo132	99.66	1164	4	0.0E+00	6RFL_B (98.97, 100)
OG014	167	P0DSW8	truncated interferon antagonist D8 (Host range protein 2)	D8L	98.67	150	2	1.8E-105	5CYW_B (98, 89)
OG015	155	P0DSX4	Protein C6	C6L	92.21	154	12	3.0E-101	
OG016	316	P17370	Protein C4	C4L	94.94	316	16	0.0E+00	8AG3_C (44.04, 99)
OG018	273	P23372	Protein E8	E8R	98.90	273	3	0.0E+00	
OG019	665	P21093	R1L protein, Protein O1	O1L	97.30	666	17	0.0E+00	
OG020	152	P33004	Protein J1	J1R	98.01	151	3	2.6E-104	
OG024	206	P21039	Protein C5	C5L	94.82	193	10	7.8E-131	4HXI_A (23.81, 40)
OG025	43	P20639	Protein K3	K3L	95.24	42	2	1.6E-23	1LUZ_A (92.86, 97)
OG026	424	P20537	Phospholipase-D-like protein K4	K4L	98.11	424	8	0.0E+00	7E0M_A (24.28, 88)
OG027	214	Q6RZR2	Apoptosis regulator OPG045	OPG045	87.10	222	21	2.9E-134	2VTY_A (89.94, 78)
OG028	486	P21013	Kelch repeat protein F3	F3L	97.08	479	14	0.0E+00	6N3H_A (25.46, 47)
OG029	319	O57175	Ribonucleoside-diphosphate reductase small chain (EC 1.17.4.1)	OPG048	99.06	319	3	0.0E+00	1H0N_A (80.94, 99)
OG030	64	P24360	Protein F8	F8L	96.92	65	1	2.4E-39	
OG031	354	P21052	Protein F11	F11L	97.18	354	10	0.0E+00	
OG032	49	P0DTM8	Protein OPG059	OPG059	95.8	48	2	4.4E-40	
OG033	101	P68455	Phosphoprotein F17	TF17R	97.03	101	3	6.6E-70	
OG034	153	P21605	RNA-binding protein E3 (p25)	E3L	88.89	153	17	9.7E-96	1OYI_A (73.33, 29)
OG035	259	P21603	DNA-directed RNA polymerase 30 kDa polypeptide (EC 2.7.7.6)	RPO30	97.68	259	6	0.0E+00	6RIC_S (97.68, 100)
OG036	166	P68447	Protein E7	E7R	93.98	166	10	2.3E-113	
OG037	95	P21050	Probable FAD-linked sulfhydryl oxidase E10 (EC 1.8.3.2)	E10R	98.95	95	1	2.6E-66	
OG038	129	P68448	viral core protein E11	E11L	97.67	129	3	1.6E-91	6RFL_Q (97.67, 100)
OG039	269	P12923	ssDNA-binding phosphoprotein, Protein I3	I3L	98.51	269	4	0.0E+00	
OG042	423	P20501	Core protease I7	I7L	99.05	423	4	0.0E+00	
OG043	591	P21022	Metalloendopeptidase G1	G1L	98.64	590	8	0.0E+00	
OG044	220	P68456	Late transcription elongation factor G2, Protein G2	G2R	98.64	220	3	5.5E-161	
OG045	124	Q8V507	Glutaredoxin-2	EVM065	100.00	124	0	4.5E-91	2G2Q_A (97.56, 99)
OG046	340	P07611	Myristoylated protein G9, Protein F1	G9R	98.53	340	5	0.0E+00	
OG047	250	P0DOT7	Protein L1; Virion membrane protein M25	L1R	99.20	250	2	0.0E+00	4U6H_E (98.91, 73)
OG048	251	P20981	core protein VP8	L4R	98.81	251	3	0.0E+00	
OG049	129	P68608	DNA-dependent RNA polymerase 22 kDa subunit rpo22	Rpo22	100.00	129	0	2.7E-92	6RFL_E (100, 100)
OG051	171	P20495	Dual specificity protein phosphatase H1	H1L	98.83	171	2	1.0E-118	2RF6_A (97.66, 100)
OG052	189	P0DSY9	IMV membrane protein, Late protein H2	H2R	100.00	189	0	3.7E-145	
OG053	324	P20497	Envelope protein H3	H3L	94.14	324	19	0.0E+00	5EJ0_A (93.67, 73)
OG054	314	P68697	DNA topoisomerase 1B	Top1B	99.36	314	2	0.0E+00	2H7G_X (97.77, 100)
OG055	146	O57208	Late protein H7	H7R	97.95	146	3	1.7E-105	4W60_A (96.58, 100)
OG056	845	Q80DX6	mRNA capping enzyme catalytic 97 kDa subunit	E1R	99.05	845	7	0.0E+00	4CKB_A (98.94, 100)
OG057	637	P04308	Early transcription factor 70 kDa small subunit	VETFS	99.84	637	1	0.0E+00	7AMV_W (99.69, 100)
OG058	161	P04310	DNA-dependent RNA polymerase 18 kDa subunit rpo18	Rpo18	97.52	161	4	5.0E-117	6RFL_G (96.89, 100)
OG059	304	Q8V4Y0	Cell Surface Binding Protein, carbonic anhydrase homolog	E8L	99.67	304	1	0.0E+00	4E90_X (93.89, 86)
OG060	287	P20980	mRNA-capping enzyme 33 kDa small subunit	D12L	98.96	287	3	0.0E+00	2VDW_B (98.61, 100)

(continued on next page)

Table 1 (continued)

Core_Og	Length (aa)	Similar to Uniprot ID	Protein name	Gene	% identity	alignment length	Mismatches	E-Value	Closest crystal structure (% Identity, % Coverage)
OG061	551	P68440	Scaffold protein D13, Rifampicin resistance protein	D13L	99.09	551	5	0.0E+00	6BED_A (99.09, 100)
OG062	150	P0DSV3	Viral late gene transcription factor 2	VLTF2	100.00	150	0	1.3E-110	
OG063	644	P20643	Virion core protein 4b, p4b	A3L	99.07	644	6	0.0E+00	
OG064	281	P20983	39 kDa core protein	A4L	95.37	281	13	5.5E-135	
OG065	161	P68610	DNA-dependent RNA polymerase 19 kDa subunit	Rpo19	97.56	164	1	8.1E-94	6RFL_F (97.56, 100)
OG066	891	P0DOL1	Virion core protein 4a precursor	A10L	97.87	892	18	0.0E+00	
OG067	318	P20988	Protein A11	A11R	99.37	318	2	0.0E+00	
OG068	190	P0DOK9	25 kDa core protein	A12L	98.42	190	15	5.7E-99	
OG069	196	P0DOR5	Virion membrane protein A17 precursor, 23 kDa late protein	A17L	97.45	196	4	1.3E-96	
OG070	77	P20994	Zinc finger-like protein, Protein A19	A19L	97.40	77	2	3.5E-51	2DFY_C (42.86, 54)
OG071	426	P68709	DNA polymerase processivity factor component A20	A20R	97.18	426	12	0.0E+00	6ZXP_A (96.75, 28)
OG072	696	P24759	A-type inclusion protein A25 (ATI)	A25	95.40	696	28	0.0E+00	
OG073	146	Q8V4U9	Envelope protein A28 homolog (Protein A30)	A30L	100.00	146	0	5.4E-106	
OG074	305	Q6RZF4	DNA-directed RNA polymerase 35 kDa subunit (EC 2.7.7.6)	RPO35	98.36	305	5	0.0E+00	6RFL_C (97.71, 100)
OG075	181	P68616	Protein A33	A33R	96.67	180	6	3.0E-130	4LQF_A (92.31, 50)
OG077	227	P68618	Protein A36	A36R	95.02	221	9	5.9E-134	
OG078	213	P21064	Protein A41	A41L	95.24	210	8	6.5E-147	2VGA_A (97.48, 92)
OG079	346	P26670	3-beta-hydroxy-Delta(5)-steroid dehydrogenase (EC 1.1.1.145)	SALF7L	98.84	346	4	0.0E+00	6JKG_A (27.02, 69)
OG080	125	Q8V4T3	Cu-Zn superoxide dismutase-like protein	A46R	100.00	125	0	9.8E-85	1P1V_A (30.67, 86)
OG081	204	Q80DS7	Thymidylate kinase (EC 2.7.4.9) (dTTP kinase)	TMK	98.53	204	3	3.9E-151	2V54_A (98.53, 100)
OG083	334	P21069	Protein A51	A51R	96.11	334	13	0.0E+00	
OG085	176	P68443	Protein B6	B6R	88.14	177	16	6.9E-109	
OG086	282	P21098	Pseudokinase B12	B12	97.17	283	7	0.0E+00	2LAV_A (32.88, 95)
OG089	190	P17365	Protein C13 (Protein B23R)	C13L	95.79	190	8	3.5E-131	

(Table 2). The modeled proteins and their predicted druggable sites have been illustrated in Fig. 2.

3.5. The druggable proteins are soluble and highly stable over a wide range of temperatures

The physicochemical properties of the 4 highly druggable proteins were predicted (Table 3). Proteins with instability index above 40 are considered unstable whereas proteins with an instability index below 40 are considered stable. The instability value of Top1B was slightly above 40. Therefore, this enzyme could be slightly unstable. The instability values of the remaining 3 proteins were below 40 and reflect their highly stable nature. This prediction was also supported by the high aliphatic index of these proteins which suggests their stability over a wide range of temperatures. Similarly, the hydrophilic features and solubility of a protein are reflected in its grand averages of hydropathy (GRAVY) score. The GRAVY scores of all 4 enzymes were below 0 which reflects upon their globular and soluble nature.

3.6. A curated library of 5893 FDA approved/investigational drugs was prepared

For high-throughput virtual screening (HTVS), we collected ~19 thousand FDA-approved/investigational drugs of approved, investigational and experimental categories from the Probes and Drugs Portal, a highly curated chemical space portal comprising data from ~50 different sources [66]. Discontinued FDA-approved/investigational

drugs were excluded from our selection. For filtering problematic functional groups and false positives, we implemented the REOS and PAINS filters on the collected library. The PAINS and REOS filter are both based on the RDKit substructure counter and compare the substructures present in the input database with a list of problematic functional groups. After filtering, ~32% (5893) of drugs remained in the dataset and were subjected to the ligand preparation stage. Finally, a set of 12,048 stereoisomers corresponding to 5893 drugs were obtained following library preparation with the LigPrep module of the Schrodinger suite.

3.7. Burixafor, batefenterol and eluxadoline are the top hits common to the four highly druggable enzymes of mpox virus

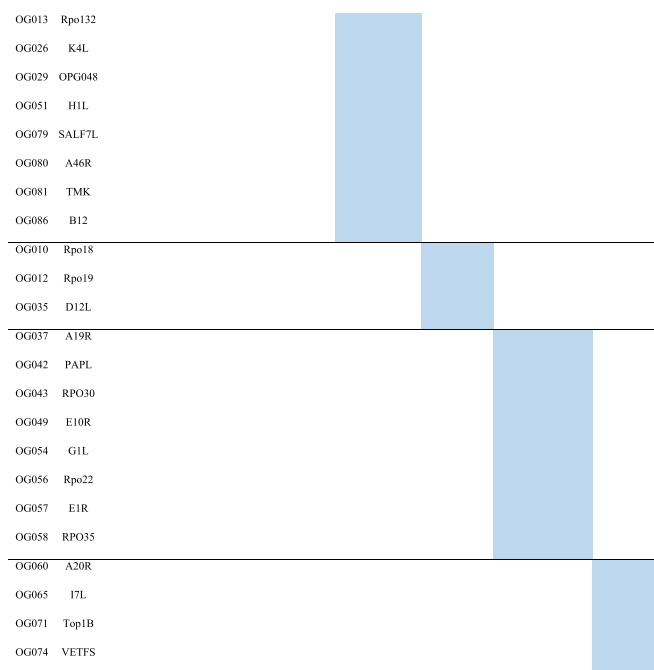
We performed the multi-step (HTVS, SP and XP) docking of 12,048 stereoisomers corresponding to 5893 drugs against all the identified target proteins by using the Schrodinger suite's Glide module. Numerous binding modes of all the drugs across the 4 targets were predicted. The docking scores varied between -13.36 kcal/mol to -0.50 kcal/mol for core protease I7L, -10.74 kcal/mol to 2.19 kcal/mol for Top1B, -11.55 kcal/mol to 0.75 kcal/mol for VETFS and -12.30 kcal/mol to 0.35 kcal/mol for A20R as shown in Fig. 3. Interestingly, the drug batefenterol (BAT) was among the top 30 hits for all the proteins namely; A20R, I7L, Top1B and VETFS whereas burixafor (BUR) and eluxadoline (ELU) were amongst the top 30 hits for I7L, Top1B and VETFS. The binding affinities of BAT with A20R, I7L, VETFS and Top1B ranged between -12.07 kcal/mol to -8.89 kcal/mol. The binding affinities of BUR with I7L, VETFS

Table 2

Table summarising the subtractive proteomics strategy and its outcomes.

The sixty-nine core proteins were screened through four stages to pinpoint therapeutic targets with highly desirable features. Four proteins namely; I7L, Top1B, VETFS, and A20R were identified as highly druggable, globular, non-host homologous, enzymatic proteins and were subjected to virtual screening in this study. Eight proteins namely, A19R, PAPL, Rpo30, E10R, G1L, Rpo22, E1R and Rpo35 passed the first three stages of screening and remain under investigation for druggability assessment. Proteins that fit under a particular group of are shaded in blue

Core OG	Gene Name	Membrane	Globular			
			Non-enzyme	Enzyme		
				Host homologous	Non-host homologous	
					Undruggable	Under Investigation
Group 1 (n=13)	Group 2 (n=33)	Group 3 (n=8)	Group 4 (n=3)	Group 5 (n=8)	Group 6 (n=4)	
OG011	B22R					
OG018	E8R					
OG019	O1L					
OG046	G9R					
OG047	L1R					
OG052	H2R					
OG053	H3L					
OG059	E8L					
OG067	A11R					
OG069	A17L					
OG073	A30L					
OG075	A33R					
OG077	A36R					
OG000	SPE2					
OG014	D8L					
OG015	C6L					
OG016	C4L					
OG020	J1R					
OG024	C5L					
OG025	K3L					
OG027	OPG045					
OG028	F3L					
OG030	F8L					
OG031	F11L					
OG032	OPG059					
OG033	TF17R					
OG034	E3L					
OG036	E7R					
OG038	E11L					
OG039	I3L					
OG044	G2R					
OG045	EVM065					
OG048	L4R					
OG055	H7R					
OG061	D13L					
OG062	VLTF2					
OG063	A3L					
OG064	A4L					
OG066	A10L					
OG068	A12L					
OG070	A19L					
OG072	A25					
OG078	A41L					
OG083	A51R					
OG085	B6R					
OG089	C13L					



and Top1B ranged between -11.84 kcal/mol to -8.29 kcal/mol. Similarly, the binding affinities of ELU with I7L, VETFS and Top1B ranged between -12.73 kcal/mol to -8.14 kcal/mol respectively. Likewise, tobramycin (PD001728), dibekacin (PD001070) and GLPG-0187 (PD058153) were identified as hits common only to I7L and VETFS and were among their top 30 hits. The XP GScore, XP HBond energy and ligand efficiency of the top 30 hits for each target is presented in Supplementary Tables S5–S8. The primary indications of the top five hits against all the targets is summarised in Table 4.

3.8. RMSD, RMSF, Rg and SASA calculations suggests stability of the ligand-bound complexes

Two independent sets of MD simulations were undertaken in this study by using the GROMACS package. The first set involved the model refinement of four proteins A20R, I7L, Top1B and VETFS to reduce steric clashes among their residues. This was achieved by simulating every protein for 400 ns. The trajectories were subjected to clustering using the TtClust. The centroid of the cluster with the least mean rmsd between the frames were selected for protein-ligand complex MD simulation. The summary of cluster analysis of the trajectories after 400 ns of simulation has been presented in Supplementary Fig. S1. Thereafter, to access the conformational stabilities of the 12 protein-ligand complexes during the 150 ns of MD production, various quality control parameters like RMSD, Rg, SASA, RMSF, H-Bonds, secondary structure, essential dynamics, MM/GBSA binding free energies and distance of ligand from active site were examined.

RMSD characterises a protein-ligand complex's conformational stability in its dynamic state during simulation. A low difference in RMSD indicates low and consistent fluctuation between the ensembles which suggests a protein-ligand complex is stable. To ensure the sampling method's reliability, the RMSDs of the proteins' alpha carbon atoms (C- α) were analysed by plotting them against the time scale of 150 ns from the starting structure (Fig. 4). None of the A20R ligand associated complexes could achieve equilibrium after 150 ns of simulation (Fig. 4A). The RMSD plot of the I7L-BAT complex (Fig. 4B) depicted an increasing trend for the first 45 ns. However, the ensembles that followed gradually traced a constant trajectory for the next 105 ns i.e.,

from 45 ns to 150 ns with consistent minor fluctuations within a permissible window of 3 Å to 4.5 Å and an average RMSD of 0.63 ± 0.26 Å. The trajectory of the I7L-BUR complex (Fig. 4B) followed a stable trend for the first 55 ns while fluctuating between a small window of 2 Å to 3 Å. The complex then encountered a minor fluctuation before finally settling into a relatively flat trajectory with an average RMSD of 0.9 ± 0.3 Å for the next 95 ns i.e., from 55 ns to 150 ns within a permissible window of 2.5 Å to 4.5 Å. Similarly, the trajectory of I7L-ELU complex (Fig. 4B) projected an increasing trend for the first 35 ns. Thereafter, the ensembles adopted a relatively flat trajectory with an average RMSD of 0.72 ± 0.28 Å for the last 115 ns i.e., from 35 ns to 150 ns within a small window of 3 Å to 5 Å. These results indicate the protein I7L has successfully achieved conformational stability with the three ligands. The ensembles of Top1B-BAT complex (Fig. 4C) traced out a relatively flat trajectory with consistent minor fluctuations for the first 85 ns within a confined permissible space between 2 Å to 4 Å. Minor blips within very narrow time frames of (25–30) ns and (45–48) ns were also encountered. This was followed by high fluctuations for the next 30 ns i.e., from (85–115) ns following which the complex again reverted back to its original 2 Å to 4 Å window for the last 35 ns. All the stable ensembles of Top1B-BAT complex that fluctuated within the 2 Å to 4 Å permissible window had an average RMSD of 1.16 ± 0.85 Å. Strangely, this pattern repeated itself again in Top1B-BUR and Top1B-ELU complexes respectively (Fig. 4C). The Top1B-BUR complex retained a consistent trajectory for the first 40 ns within an allowable window of 2 Å to 4 Å. This complex then navigated through a similar violent RMSD fluctuations for the next 55 ns i.e., from (40–95) ns before finally returning back to its initial 2 Å to 4 Å window for the last 55 ns with consistent minor fluctuations. All the stable ensembles of Top1B-BUR complex that fluctuated within the 2 Å to 4 Å permissible window had an average RMSD of 0.88 ± 0.59 Å. The ensembles of Top1B-ELU complex followed an increasing trend for the first 20 ns. Thereafter, the ensembles retained a stable trajectory throughout the entire 150 ns between a 2 Å to 4 Å window with an average RMSD of 1.17 ± 0.77 Å with the exception of two small time frames between 75 ns to 85 ns and 115 ns–120 ns respectively where major fluctuations in RMSD were clearly distinguishable (Fig. 4C). The ensembles of VETFS-BAT complex pursued an upward trajectory for the first 10 ns (Fig. 4D). Thereafter, the ensembles retained

Table 3

Physicochemical properties and model quality statistics of the four proteins of mpox virus identified through subtractive proteomics.

Features	Core protease I7	DNA topoisomerase 1B	Early transcription factor 70 kDa small subunit	DNA polymerase processivity factor
Gene Name	I7L	Top1B	VETFS	A20R
Enzyme Commission (EC)	3.2.22.-	5.99.1.2	3.6.4.13	3.-
KEGG Orthology (KO) Identifier	-	-	-	K21082
Closest Crystal Structure	-	2H7G (97.77%, 100%)	7AMV (99.69%, 100%)	6ZXP (96.75%, 28%)
Length (Amino Acid)	423 aa	314 aa	637 aa	426 aa
Mol. Weight (Dalton)	49023.57	36665.56	73844.91	49147.14
pI	7.85	9.50	6.93	5.62
Instability Index	33.81	40.22	34.45	36.35
Aliphatic Index	83.14	88.03	97.85	92.77
GRAVY	-0.22	-0.32	-0.17	-0.26
Predicted IDDT Score	0.64	0.87	0.70	0.80
Overall Quality Score	95.42	95.75	95.66	91.87
MolProbity Score	1.44 (96th percentile)	0.90 (100th percentile)	1.40 (97th percentile)	1.37 (98th percentile)
Clash Score	3.49 (97th percentile)	1.53 (99th percentile)	2.78 (98th percentile)	2.88 (98th percentile)
Poor Rotamers (%)	0	0	0	0
RamaPlot Most Favoured (%)	95.72 (403 aa)	100.00 (312 aa)	95.12 (604 aa)	95.80 (406 aa)
RamaPlot Allowed (%)	3.8 (16 aa)	0	3.48 (22 aa)	4 (17 aa)
RamaPlot Disallowed (%)	0.48 (2 aa)	0	1.42 (9 aa)	0.24 (1 aa)

a stable trajectory for the next 140 ns i.e., from 10 ns to 150 ns with constant but minor fluctuations within an allowable window between 3 Å to 5 Å with an average RMSD of 0.57 ± 0.36 Å. The ensembles of VETFS-BUR complex traced an ascending step ladder trajectory for the first 70 ns before gradually settling into a stable trajectory for the remaining 80 ns i.e., from 70 ns to 150 ns (Fig. 4D). The stable ensembles during this time frame constantly fluctuated within a narrow and acceptable window between 4 Å to 6 Å with an average RMSD of 0.88 ± 0.39 Å. A rising trend was observed in the trajectory of VETFS-ELU complex for the first 40 ns (Fig. 4D). The trajectory thereafter declined which was followed by a steady rise in the next 45 ns before finally achieving stability for the last 55 ns i.e., from 85 ns to 150 ns. The stable ensembles during this time frame fluctuated within a 6 Å to 8 Å window with an average RMSD of 0.88 ± 0.44 Å. It is also interesting to note that the VETFS-BAT system achieved equilibrium within a very short duration in comparison to its BUR and ELU bound counterparts. This suggests, in comparison to VETFS-BAT complex, VETFS-BUR and VETFS-ELU complexes had to navigate through several conformational transitions to accommodate their respective ligands before finally achieving stable conformations. Overall, these results indicate that the enzyme VETFS has successfully acquired conformational stability with all the three ligands. Taken together, only nine complexes that showed conformational stability were considered for further examination. The density distribution of RMSDs of these nine stable complexes have also been provided in [Supplementary Fig. S2](#).

Subsequently, the radius of gyration (Rg) of the 9 stable protein-ligand complexes were also assessed (Fig. 5). Rg examines a protein's compactness where a constant Rg suggests that the system has achieved a relatively constant shape and size during the entire simulation. The average radius of gyration of I7L-BAT, I7L-BUR and I7L-ELU complexes were 23.16 ± 0.19 Å, 22.74 ± 0.14 Å and 23.16 ± 0.24 Å respectively. These complexes successfully retained their compactness without undergoing any major structural re-arrangements during the entire 150 ns simulation as demonstrated by the consistency of their time dependent Rg plot (Fig. 5A). Similarly, the Rg of the Top1B-BUR and Top1B-ELU complexes also remained fairly consistent during the entire simulation with average Rg's of 25.51 ± 0.51 Å and 25.34 ± 0.52 Å respectively (Fig. 5B). The average Rg of Top1B-BAT complex was 24.71 ± 0.69 Å during the entire simulation with some minor fluctuations. The Rg of Top1B-BAT complex remained consistent for the first 75 ns following which it experienced a sharp decline and a subsequent upturn in the next 55 ns i.e., between 75 ns and 130 ns before retracing its original trajectory in the last 20 ns (Fig. 5B). Overall, the consistent Rg of the three Top1B-ligand bound complexes suggest they have retained their structural integrity and are relatively compact after 150 ns of simulation. The VETFS-BAT complex also remained compact with an average Rg of

30.05 ± 0.27 Å (Fig. 5C). This was evident from its time-evolution Rg plot which remained fairly consistent during the entire simulation. The Rg of VETFS-BUR and VETFS-ELU complexes remained steady for the first 70 ns (Fig. 5C). Thereafter, the Rg of VETFS-BUR declined linearly for the next 30 ns before retaining its steadiness for the last 50 ns. On the contrary, the Rg of VETFS-ELU complex witnessed a linear increment for the next 10 ns before becoming relatively constant for the last 70 ns. This suggests that the presence of BUR in the active site of VETFS has enhanced its compactness. It also suggests that BUR is probably better adapted to VETFS in comparison to ELU. The Rg of VETFS-BUR and VETFS-ELU complex was 29.59 ± 0.32 Å and 30.64 ± 0.39 Å respectively. Taken together trends of Rg of all the nine complexes are in congruence with their corresponding RMSD trends that reflect the relative stabilities and compactness of the 9 ligand bound enzyme complexes.

SASA is another frequently studied quality control parameter that is examined to scrutinize a protein-ligand complex's conformational stability. The solvent environment around the protein plays a crucial role in retaining a protein's folding and governs the protein-ligand interaction processes, orientation, and stability. The time evolution SASA plots clearly indicate that all the nine complexes are relatively well equilibrated during the entire 150 ns simulation (Fig. 5D, E and 5F). The SASA values varied within a narrow range between 233.4 ± 4.43 nm² - 242.11 ± 6.67 nm² for I7L-ligand complexes, 191.76 ± 3.32 nm² - 192.50 ± 3.04 nm² for Top1B-ligand complexes and 339.96 ± 4.35 nm² - 348.41 ± 5.27 nm² for VETFS-ligand bound complexes respectively. These evidences are in agreement with their corresponding RMSDs and suggest that all the proteins have successfully attained conformational stabilities with their respective ligands without undergoing any major alterations in their available surface area.

Root Mean Square Fluctuation (RMSF) was computed to examine the positional fluctuation of each amino acid residue around its mean position. The Cα atoms of all the I7L complexes shared very similar RMSF trends (Fig. 6A). This case was also similar for the Cα atoms of Top1B (Fig. 6B) and VETFS (Fig. 6C) ligand bound complexes. The Cα atoms associated with the loops experienced high fluctuations whereas the Cα atoms associated with α-helix and β-sheets were comparatively rigid in all the complexes. High fluctuations were also associated with the C terminal residues in all the nine complexes. Amongst the I7L complexes, the highest average atomic fluctuation of 1.83 ± 0.97 nm was recorded in the residues of the ELU bound complex. Similarly, amongst the Top1B and VETFS complexes, highest average atomic fluctuations of 3.19 ± 1.6 nm and 2.71 ± 1.27 nm was recorded in the residues of BUR and ELU bound complexes respectively. From Fig. 6B it is quite evident that high atomic fluctuations of all the Top1B complexes are mostly restricted to the residues from 1 to 70 which encompass the protein's N-

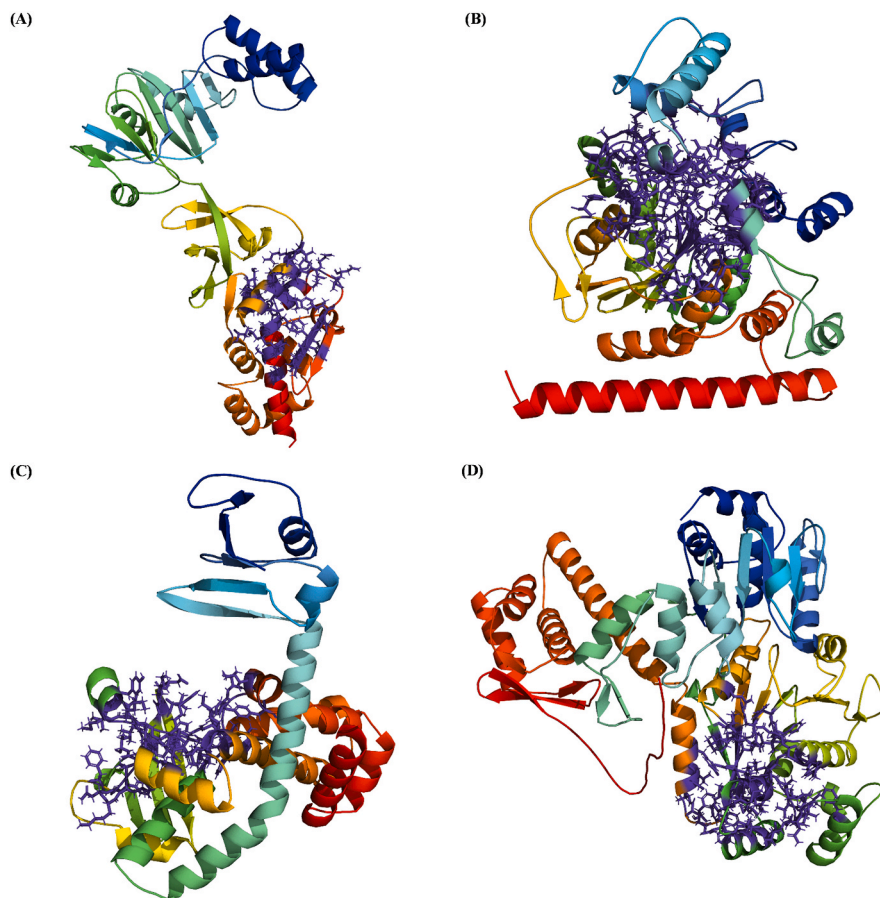


Fig. 2. Homology models of four highly druggable core proteins of mpox virus. (A) Model of DNA polymerase processivity factor component A20 (A20R). (B) Model of Core cysteine proteinase (I7L). (C) Model of DNA topoisomerases type 1B (Top1B). (D) Model of Early transcription factor 70 kDa small subunit (VETFS). The figures of the models are represented as spectrum coloured cartoons. The predicted active site residues are styled as sticks with purple blue colour. The active site residue information has been provided in [Supplementary Table S4](#). The figures have been illustrated with Pymol v2.5.3.

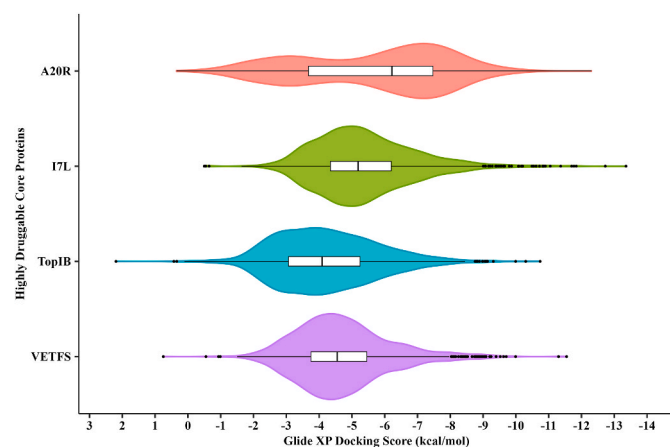


Fig. 3. Distribution of glide XP docking score (kcal/mol) against DNA polymerase processivity factor component A20 (A20R), Core cysteine proteinase (I7L), DNA topoisomerases type 1B (Top1B) and Early transcription factor 70 kDa small subunit (VETFS).

terminal domain. The fluctuations in the residues that encompass the ligand bound C-terminal DNA binding domain (76–314) are relative restricted which hints at the conformational rigidity of this domain. This suggests that the C-terminal DNA binding domain of Top1B has been successfully stabilized by the three ligands and the wild fluctuations in Top1B's RMSD trajectories are most probably related to the flexibility of its N-terminal domain.

The molecular dialogues between a ligand and the flexible active site amino acid residues are largely driven by H-bond interactions. We

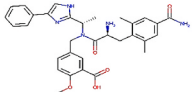
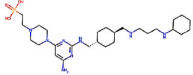
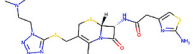
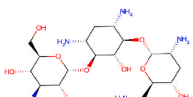
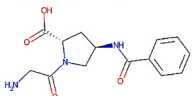
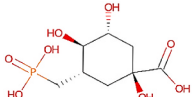
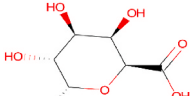
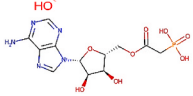
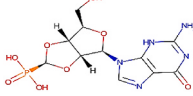
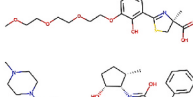
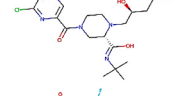
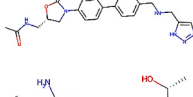
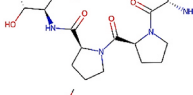
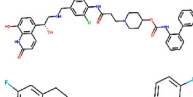
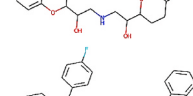
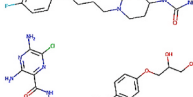
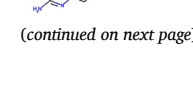
further examined the time evolution plots of H-bonds involved in the molecular interaction of the three ligands BAT, BUR and ELU with each of the three proteins namely, I7L, Top1B and VETFS. The nine protein-ligand complexes depicted differential intermolecular H-bonding patterns during the 150 ns of MD run (Fig. 6D, E and 6F). The equilibrated complexes of protein I7L formed an average of 2.31 ± 1.36 , 2.1 ± 1.19 and 1.5 ± 1.03 H-bonds per frame with the ligands BAT, BUR and ELU respectively. The equilibrated complexes of Top1B formed an average of 1.28 ± 1.03 , 1.79 ± 1.17 and 1.55 ± 1.06 H-bonds per frame with the ligands BAT, BUR and ELU respectively. Similarly, the fully equilibrated complexes of VETFS formed an average of 1.14 ± 0.86 , 1.705 ± 1.22 and 1.59 ± 0.65 H-bonds per frame with the ligands BAT, BUR and ELU respectively. H-bond interactions govern the correct orientation and stability of a ligand in a protein's active site which subsequently drives the protein's overall compactness and structural conformity. Therefore, the compactness and structural rigidity of I7L, Top1B and VETFS as inferred from their stable time evolution RMSD, Rg and SASA plots are most probably driven by these H-bond interactions.

3.9. The stabilities of the protein-ligand complexes are largely driven by switching between secondary structural elements

Changes in secondary structure as a function of time were also mapped for every complex using the Biolite, a Python-based molecular biology suite. Helices and β -sheets are stable amongst the secondary structure elements whereas coils, loops and turns are highly flexible. Constant switchings between coils, bends and turns were clearly visible in all the complexes during their simulations. These switchings probably contribute in parts to the molecular motions of all the complexes. The integrity of all the major α -helices and β -sheets remained intact in all the complexes. These α -helices and β -sheets are most likely responsible for

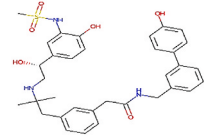
Table 4

Mode of actions, application and 2D structure of shortlisted top five docked drugs with high binding affinity for each target. The mode of actions, applications was collected from DrugBank, Inxight Drugs portal and PubChem database.

Target	Drug	PDID	Drug Status	Primary Indications (1. Mode of Action; 2. Treatment)	2D Structure
I7L	Eluxadoline	PD008978	Approved	1. An agonist of mixed mu-opioid receptor. 2. Used for the treatment of irritable bowel syndrome with diarrhea.	
I7L, VETFS	Burixafor	PD058763	Investigational	1. Inhibitor of CXC chemokine receptor 4 (CXCR4) 2. Used in trials studying treatment of Hodgkin's disease, Non-hodgkin's Lymphoma and Multiple Myeloma.	
I7L	Cefotiam	PD010157	Approved	1. Inhibits the bacterial cell wall biosynthesis. 2. Used for the treating a myriad of bacterial infections.	
I7L	Tobramycin	PD001728	Approved	1. Inhibits the synthesis of protein by binding to ribosome 30S subunit in bacterial cells. 2. Used for the treating a myriad of bacterial infections.	
I7L	Danegaptide	PD058923	Investigational	1. A selective 2nd generation gap junction modifier in adjacent cardiomyocytes. 2. Used in studying treatment of chronic atrial fibrillation (AF) and postoperative AF in large animal models.	
Top1B	Carbaphosphonate	PD059915	Investigational	1. 3-dehydroquinate synthase inhibitor (<i>Staphylococcus aureus</i>)	
Top1B	α -D-galacturonic acid	PD006851	Investigational	1. Backbone of pectin, cellular binders in the peel of many different fruits and vegetables. 2. Used as: antidiarrheal drug, food emulsifier, food stabilizer, food thickening agent and food gelling agent.	
Top1B	Adenosine phosphonoacetic acid	PD059913	Investigational	1. ndp kinase (human) inhibitor.	
Top1B	Guanosine-2',3'-O-methylidene phosphonate	PD006419	Investigational	1. catalyze the phosphorylytic breakdown of the N-glycosidic bond in the beta-(deoxy) ribonucleoside molecules	
Top1B	Deferitazole	PD058219	Investigational	1. An iron chelator 2. Used in trials studying the treatment and basic science of Beta-thalassemia.	
VETFS	DB02785	PD059991	Investigational	1. Gag-Pol polyprotein inhibitor 2. Used as an antiviral (antiHIV-1)	
VETFS	Radezolid	PD012727	Investigational	1. A 2nd generation's oxazolidinone antibiotic, inhibits the RRBP1. 2. Used in trials for studying the treatment of Abscess, Infectious Skin Diseases.	
VETFS	Rapastinel	PD070040	Investigational	1. NMDA receptor modulator with glycine-site partial agonist properties. 2. Used in trials studying the treatment of Major and Obsessive-Compulsive Disorder (OCD).	
I7L, A20R	Batefenterol	PD058591	Investigational	1. A bifunctional muscarinic (M2 and M3 receptors) antagonist and β 2-agonist 2. Used in trials for studying the treatment of Chronic Obstructive Pulmonary Disorder (COPD).	
A20R	Nebivolol	PD009358	Approved	1. β -1 adrenergic receptor antagonist (beta blockers). 2. Used to treat high blood pressure either alone or in combination with other medications.	
A20R	Pimozide	PD001879	Approved	1. Dopamine type 2 receptors blocker 2. Used to control motor or verbal tics caused by Tourette's disorder.	
A20R	BAZ2-ICR	PD046767	Investigational	1. Epithelial sodium channel blocker 2. Used in trials for studying cystic fibrosis and chronic bronchitis.	

(continued on next page)

Table 4 (continued)

Target	Drug	PDID	Drug Status	Primary Indications (1. Mode of Action; 2. Treatment)	2D Structure
A2OR	PF-610355	PD058896	Investigational	1. A novel Ultra-Long-Acting β 2-Adrenoreceptor agonist. 2. Used for studying the treatment of Asthma and Chronic Obstructive Pulmonary Disease.	

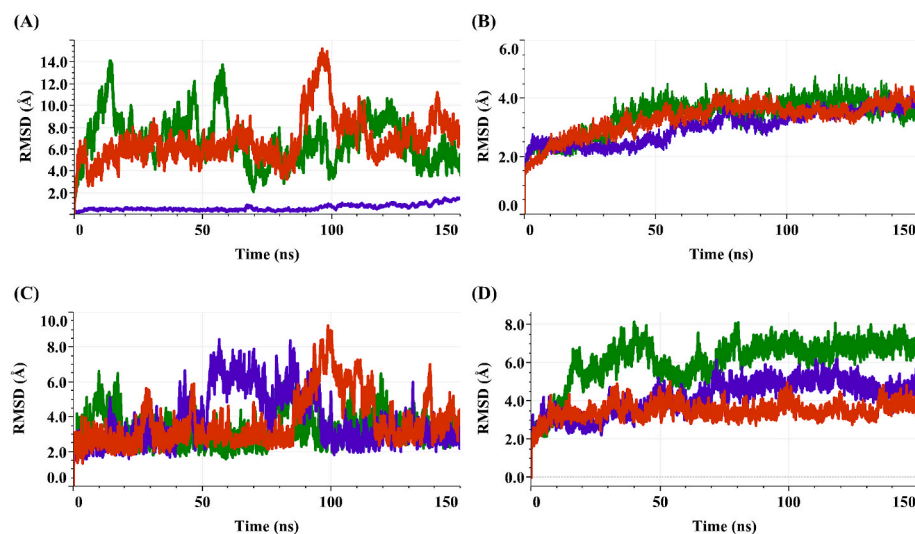


Fig. 4. Dynamics of RMSD of the mpxv virus ligand bound protein complexes during 150 ns of simulation. (A) RMSD of A2OR ligand bound complexes. (B) RMSD of I7L ligand bound complexes. (C) RMSD of Top1B ligand bound complexes. (D) RMSD of VETFS ligand bound complexes. The colours red, blue and green represent the colours of the three ligands batafenterol, burixafor and eluxadoline respectively in complex with the respective proteins. The illustrations were created with Tableau.

the global conformational rigidity of the proteins. Analysis of the time evolution map of the secondary structure elements resulted in some interesting observations. The flexible turn between the residues 40 and 50 in the I7L-ELU complex gradually transformed into a more rigid α -helix after 10 ns and remained constant for the remaining duration of the simulation (Supplementary Fig. S3). This correlates well with its RMSD plot where the stability of the complex was achieved immediately after 10 ns. This observation suggests that ELU mediated stabilization of the enzyme I7L was most likely driven by the metamorphosis of a turn between the residues 40 and 50 into an α -helix. The turn at the residues between 70 and 75 of Top1B-BUR complex was replaced by a more flexible coil between 50 ns and 95 ns which probably explains the corresponding sharp RMSD and RMSF fluctuations (Supplementary Fig. S4). The residues between 420 and 430 in VETFS-BUR complex continued to switch between a turn and α -helix for the first 85 ns before smoothly transitioning into an α -helical structure for the remainder of the simulation (Supplementary Fig. S5). This correlates well with the RMSD trajectory of this complex and suggests that ELU driven stabilization of VETFS is regulated by the formation of an α -helix. Overall, these results suggest that the transition of unstable protein-ligand complexes into their equilibrated states and vice versa are largely driven by switching between secondary structural elements.

3.10. The equilibrated ensembles of the protein-ligand complexes are grouped into many clusters

The ensembles of I7L, Top1B and VETFS ligand bound complexes generated after 150 ns of the simulation were clustered with the TtClust to identify the clusters with stable ensembles and the corresponding centroid structures that represent these clusters. The TtClust analyses the trajectories to generate graphical imagery like distribution within the clusters, the relative distance between clusters and hierarchical cluster dendrograms. Fig. 7 and Supplementary Fig. S6 depict the results of the cluster analysis of I7L complexes during the 150 ns of MD

simulation. The equilibrated ensembles of I7L-BAT complex was subgrouped into 3 clusters (C2, C3, C4), whereas those of I7L-BUR and I7L-ELU complex were further subgrouped into 2 clusters (C2, C3) each. The equilibrated ensembles of Top1B-BAT, Top1B-BUR and Top1B-ELU complexes were also subgrouped into 2 clusters (C1, C2) each (Fig. 8 and Supplementary Fig. S6). Similarly, the equilibrated ensembles of VETFS-BAT, VETFS-BUR and VETFS-ELU complexes were subgrouped into 3 (C1, C2, C3), 2 (C3, C4) and 1 (C3) clusters respectively (Fig. 9 and Supplementary Fig. S6). The distribution of these equilibrated ensembles into different clusters suggests that they probably sample small conformational subspaces with minor structural differences. The protein-ligand interactions experienced by the centroid frames of these stable sub-clusters are summarised in Table 5.

3.11. PCA suggests the motions of the protein ligand complexes are relatively constricted

Protein function is controlled by shifting between various conformations. This ability to flexibly switch between multiple conformations is regulated by a set of their collective motions. These motions are critical to different biological processes and have a key contribution in the transmission of biological signals. A fair amount of flexibility as well as rigidity is necessary, especially of the binding site residues for any protein to be functional. A tighter ligand interaction would essentially restrict a protein's motion, thereby preventing it from sampling the critical conformations necessary for its functionality. Therefore, essential dynamics (ED) analysis was attempted to better understand the collective motions of all the nine ligand bound protein complexes in the 2D conformational phase space during the 150 ns simulation. As a multivariate statistical method, ED is reliant upon diagonal covariance matrix of a protein's C α atom to trace its global motion through Eigen vectors (EVs) popularly called principal components (PCs) and eigenvalues [83]. The EVs describe the global direction of motion of the atoms and the associated eigenvalues depict the atomic contribution of motion in

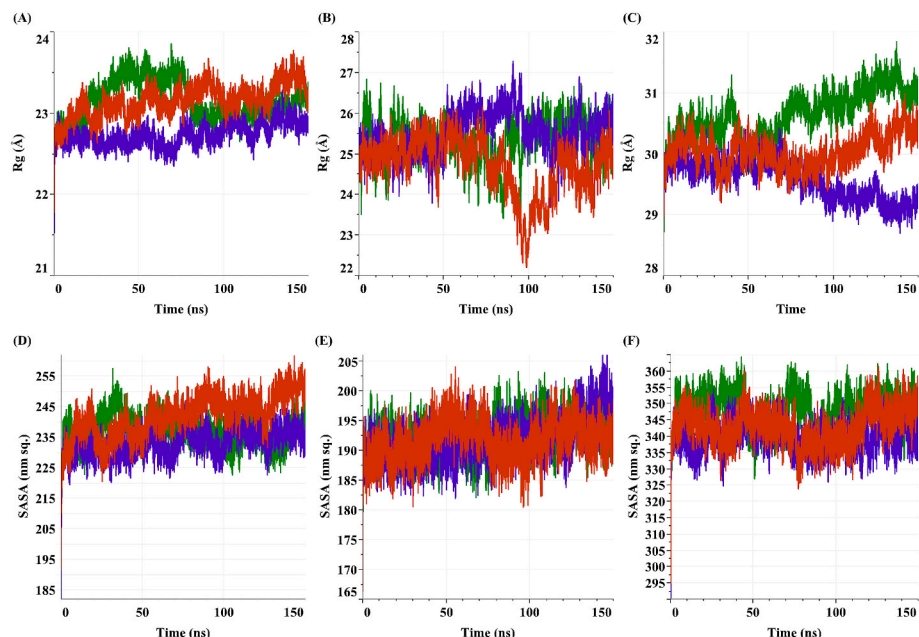


Fig. 5. Dynamics of radius of gyration (Rg) and solvent accessible surface area (SASA) of the mpox virus ligand bound protein complexes during 150 ns of simulation. (A) Rg of I7L ligand bound complexes. (B) Rg of Top1B ligand bound complexes. (C) Rg of VETFS ligand bound complexes. (D) SASA of I7L ligand bound complexes. (E) SASA of Top1B ligand bound complexes. (F) SASA of VETFS ligand bound complexes. The colours red, blue and green represent the colours of the three ligands batafenterol, burixafor and eluxadoline respectively in complex with the three proteins. The illustrations were created with Tableau.

MD trajectories of a ligand bound system which is essentially controlled by a proteins' secondary structure. The collective dynamic motions of the proteins captured through the projections of EVs (PC1 and PC2) are shown in Fig. 10. The PCA statistics and the percentage of variance covered by the first three eigenvectors of I7L, Top1B and VETFS complexes are summarised in Supplementary Table S9. As visible from the figure, the first two PCs accounted for the bulk of all the internal motions. This indicates that the vectors mapped within this conformational subspace elucidate the essential subspace of the system. The trace values calculated from the covariance matrices of I7L-BAT, I7L-BUR, I7L-ELU, Top1B-BAT, Top1B-BUR, Top1B-ELU, VETFS-BAT, VETFS-BUR and VETFS-ELU complexes were 17.09 nm², 16.99 nm², 18.22 nm², 30.52 nm², 40.01 nm², 23.69 nm², 44.98 nm², 42.12 nm² and 57.12 nm² respectively. The trace values of Top1B-BUR, VETFS-BAT, VETFS-BUR and VETFS-ELU complexes were relatively higher compared to the

other five complexes. These results are consistent with higher occupation of the conformational subspace by these complexes which is clearly visible in the form of dispersed spectral dots and further hints at greater flexibility of these proteins. Therefore, as a consequence of higher fluctuations these complexes had to sample a much wider conformational space most probably to accommodate the ligands before accomplishing the ensembles in their dynamically equilibrated states. It is interesting to note that the molecular motions of the 3 Top1B complexes are relatively constricted (as supported by their low phase space occupancy). This further strengthen the idea that ensembles of all the Top1B complexes have achieved as state of dynamic equilibrium with their respective ligands. Overall, ED analysis suggests the collective molecular motions of all the 9 complexes are restricted to a small and localised conformational space. These results are consistent with their corresponding RMSD, SASA and Rg plots which further strengthen the notion

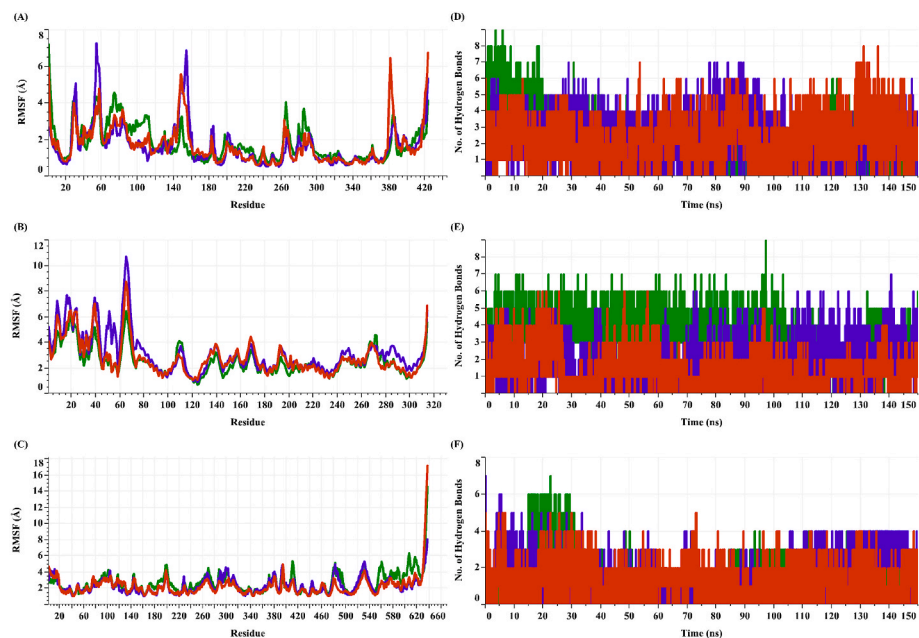


Fig. 6. Dynamics of RMSF and intermolecular hydrogen bonds of the mpox virus ligand bound protein complexes during 150 ns of simulation. (A) RMSF of I7L ligand bound complexes. (B) RMSF of Top1B ligand bound complexes. (C) RMSF of VETFS ligand bound complexes. (D) H-Bonds of I7L ligand bound complexes. (E) H-Bonds of Top1B ligand bound complexes. (F) H-Bonds of VETFS ligand bound complexes. The three colours red, blue and green represent the colours of the three ligands batafenterol, burixafor and eluxadoline respectively in complex with the three proteins. The illustrations were created with Tableau.

that the three proteins namely, I7L, Top1B and VETFS have been stabilized by the ligands BAT, BUR and ELU.

3.12. Porcupine plot assessment suggest stability of Top1B's ligand bound DNA binding domain

Porcupine plots were developed to gain quantitative insights into the molecular motion of the 9 protein-ligand complexes. To picture the molecular motions, a cone was drawn along the C-alpha atom of each residue in the direction traced by its extreme projection on PC1 (Fig. 11) where the length of the cone signifies the amplitude of motion of the C α atom. It shows that a vast majority of all the fluctuations are closely associated with the regions representing the loops/coils in all the protein ligand complexes. These results are completely in tune with their corresponding RMSF plots. Similarly, cones were also observed on the residues neighbouring the ligands in all the complexes. These fluctuations have most likely occurred to accommodate the ligands in the active site cavity. Interestingly, large protruding cones were clearly visible in the N-terminal domains of all the Top1B complexes which was not visible in their ligand bound C-terminal DNA binding domain. This clearly suggests the large fluctuations in Top1B complexes are particularly associated with the N-terminal domain. This is perfectly in sync with their RMSF plots where large fluctuations were specifically confined to the residues of the N-terminal domain. Combined together, the rmsf plots and porcupine plots of the Top1B complexes evidently explain the large fluctuations in RMSDs of all the Top1B complexes. These correlations beyond doubt clarify that the target C-terminal DNA binding domain of Top1B has successfully established an equilibrium with the ligands BAT, BUR and ELU.

3.13. MM/PBSA assessment suggests stable binding of the ligands BAT, BUR and ELU with the three proteins I7L, Top1B and VETFS

The assessment of RMSD trajectories suggest all the nine protein-ligand systems stabilized in the last 50 ns. Therefore, MM/PBSA was used to estimate the binding free energy and understand the associated molecular interactions between the proteins I7L, Top1B, VETFS and the ligands BAT, BUR and ELU for the last 50 ns of MD run. Supplementary Fig. S7 suggests the trajectories of binding free energy of all the nine protein-ligand complexes remained almost constant in the last 50 ns. Fig. 12 suggests that BAT, BUR and ELU bind favourably with all the three proteins. The three proteins, I7L, Top1B, VETFS, had highest binding affinity (ΔG_{mmpbsa}) of -29.63 ± 7.53 kcal/mol, -32.72 ± 5.17 kcal/mol and -24.96 ± 4.48 kcal/mol respectively with BAT suggesting BAT might be the most suitable ligand for their inhibition. The decomposition of the average binding free energy of all the nine complexes into energy terms suggests the binding of the ligands BAT, BUR and ELU in the active site of majority of the proteins is stabilized by Van der Waal's (vdW) energies (Fig. 12). However, another critical energy term average electrostatic (ΔE_{ele}) seems to be the major contributor towards the stabilization of Top1B-ELU ($\Delta E_{\text{ele}} = -59.94 \pm 11.95$ kcal/mol) and VETFS-ELU ($\Delta E_{\text{ele}} = -42.85 \pm 10.55$ kcal/mol) complex respectively.

We further employed per residue binding free energy decomposition to determine the molecular interactions of every ligand with the binding pocket residues of the proteins BAT, BUR and ELU (Fig. 13). Per residue decomposition analysis suggests the active site residues energetically favour the stable binding of BAT with protein I7L. The free energy decomposition analysis plot suggests the active site residues Arg99, Ile298, Ile316, Leu109, Phe256, Pro110 and Pro264 favour the stable binding of BAT with I7L. The residues Met67, Pro64, Pro110, Pro114, Ser112, Thr111, Tyr46, Tyr238, Val47 favour the binding of BUR to the protein I7L. Similarly, the active site residues Asp258, Leu239, Pro264, Ser112 and Tyr238 favour the binding of ELU with I7L. The residues Pro 110, Ser112 and Pro264 of I7L protein were common to two of the three ligands. This suggests these residues might be important to I7L. The residues Met126 and Phe131 favoured the binding of BAT to the active

site of Top1B. The residues Arg223, Gly264, Ile129, Phe127, Phe128, Phe174, Tyr233, Val 262, and Val 263 favoured the stable binding of BUR in the active site of Top1B whereas the pocket residues Arg130, Ile129, Phe131, Val 262 and Val 263 contributed towards the favourable binding of ELU in Top1B's active site. The residues Ile129, Phe 131, Val262 and Val263 were common to two of the three ligands bound to Top1B. Therefore, these residues might be important to Top1B. The residues Lys273, Met 275, Tyr276 and Tyr477 favoured the binding of BAT in the active site of protein VETFS whereas the active site residues Arg262, Met275, Phe271, Tyr276 guided the favourable binding of ELU with VETFS. Residues favouring the binding of BUR with VETFS could not be identified. The residues Met275 and Tyr276 were common to both the ligands. As such, these residues might be critical to VETFS.

All the ligand bound complexes maintained a favourable distance below 10 Å between the centre of mass of the interacting active site residues and centre of mass of their associated ligands with the exception of VETFS-BUR complex in the last 50 ns of simulation (Supplementary Fig. S8). This suggests stable binding modes of these ligands with the active site residues of the three proteins. The distance between the centre of mass of the ligand BUR and the centre of mass of the active site pocket residues of VETFS was far greater than 10 Å. This suggests BUR has been displaced from the active site of VETFS and therefore is no longer bound to it. Therefore, taken together only eight complexes namely, I7L-BAT, I7L-BUR, I7L-ELU, Top1B-BAT, Top1B-BUR, Top1B-ELU, VETFS-BAT and VETFS-ELU have successfully attained a state of dynamic equilibrium. These findings suggest the ligands BAT and ELU have the potential to function as inhibitors of all the three enzymes whereas BUR can potentially inhibit the enzymes I7L and Top1B.

4. Discussion

In this study, 125 genomes of mpox virus were collected from the public repository. Mining of these genomes led to the identification of four highly druggable non-host homologous enzymes namely, I7L, Top1B, VETFS and the A20R. Further, screening of a non-redundant library of 5893 compounds yielded BAT, BUR and ELU as potential inhibitors with an affinity for multiple targets. Interestingly, these molecules were amongst our top 30 hits for every target. Following MD simulation, the interactions of these ligands with three targets namely I7L, Top1B, VETFS were demonstrated to be stable. We fully acknowledge that our findings currently lack *in vivo* and *in vitro* antiviral experimental validations. However, we wanted to make our findings available to the scientific community at the earliest with the hope that our findings can be suitably exploited by others in the fight against the global mpox epidemic.

The presently ongoing mpox outbreak is unique as it is the most dispersed and the largest to be ever documented in non-endemic countries. Its rapid rise and dissemination across the globe are a cause of great concern. However, the only probable saving grace is the fact that the current strains under circulation are closely related to the less infectious strains of clade II [84,85]. Nevertheless, a global outbreak of its more nefarious and highly virulent cousin from clade I (with case fatality rate >10%) [86] can never be ruled out in the future. Currently, there are no specific antivirals against mpox. Therefore, it is highly desirable to quickly identify potential therapeutics against mpox. Identifying 'drug repurposing' potential through *in silico* analysis is an approach to fasten antiviral development. However, the availability of valid, druggable targets is key to the success of this strategy. Unfortunately, no information is available about the druggable targets of mpox virus.

Genes that are conserved across diverse phylogenetic lineages are most likely to be essential and their subsequent protein products are likely to have a critical role in a pathogen's lifecycle and virulence [87, 88]. In the current study, 69 highly conserved proteins were identified in all genomes of mpox virus. The ease of access to the therapeutic targets under investigation is paramount for an experimental biologist to

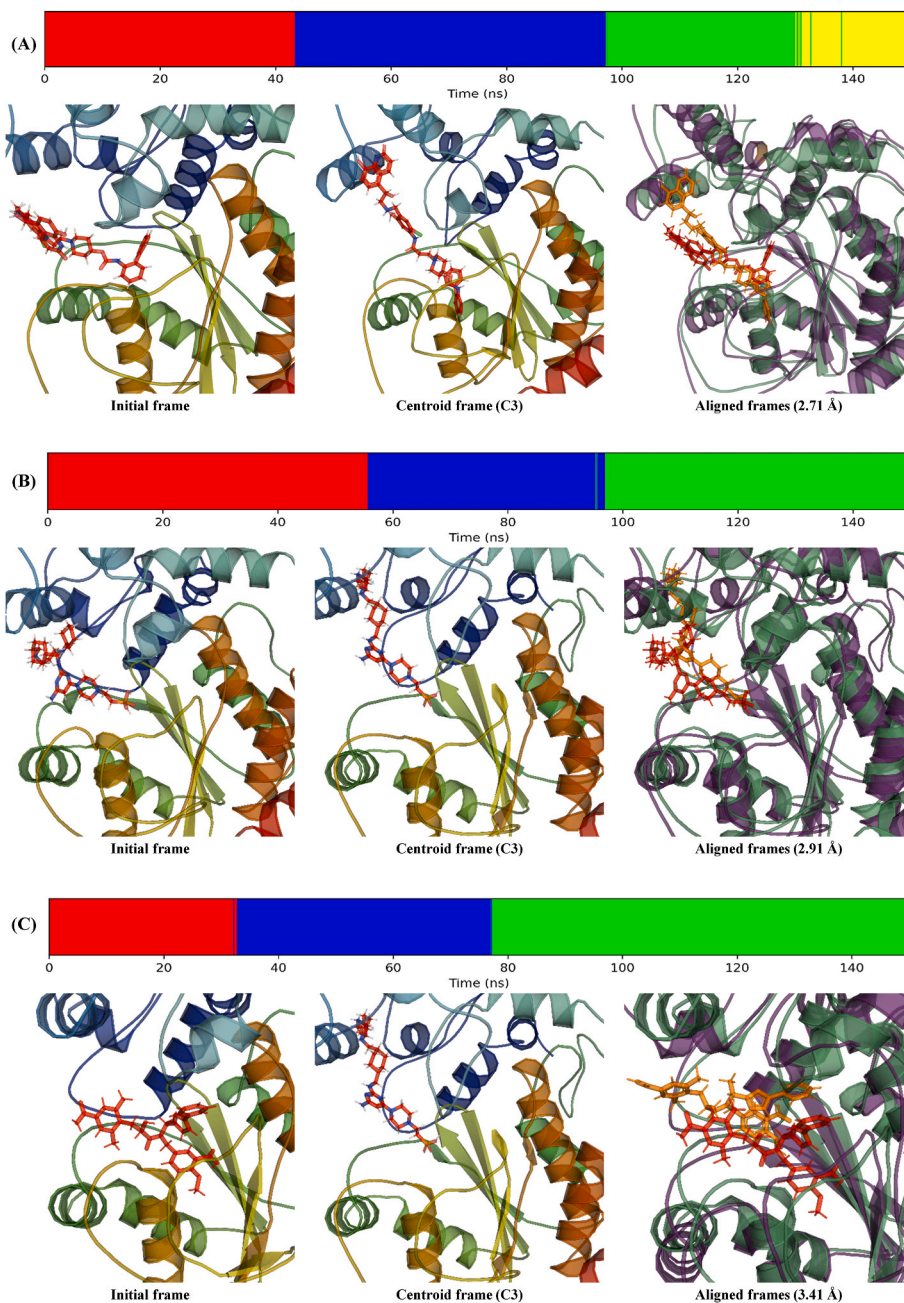


Fig. 7. Cluster analysis of the I7L ligand bound complexes. (A) I7L-BAT. (B) I7L-BUR. (C) I7L-ELU. The linear plot depicting the distribution of conformations (frames) over 150 ns of simulation. Close view of initial frame, centroid frame of the most populated cluster and 3D alignment of initial frame with centroid frame are presented below each linear plot from left to right. The total number of clusters and frames per cluster has been detailed in [Supplementary Fig. S6](#). The colours in the time evolution linear plots represent the cluster to which a frame belongs at a given point of time. The cartoons of initial (far left) and centroid frames are illustrated as spectrum. Close view of aligned frames (far right) where the initial and centroid frames are illustrated as limegreen and violetpurple coloured cartoons respectively. The bound ligands with initial and centroid frames are illustrated in red and orange colours respectively.

rapidly investigate and verify the outcomes of *in silico* studies. Globular proteins provide a significant advantage over membrane-associated proteins in this regard. Thus, a pool of 56 globular proteins was selected for further downstream investigations.

Since the identified possible core transcription factors associated with mpox virus remain functionally uncharacterised, targeting them without prior scientific evidence might not yield desirable results. Hence, the possibility of transcription factors as therapeutic targets was ruled out by subtractive proteomic approach from the list of identified globular proteins for the time being. On the contrary, the identified core enzymes shared sequence similarities and structural identities with the enzymes of its nearest neighbour, the vaccinia virus, whose functionalities are fairly well characterised. Therefore, only the 23 enzymes from the pool of 56 globular proteins were pursued further to identify therapeutic targets.

An ideal drug target in a pathogen must always be non-homologous to the host's proteome to minimise cross-reactivity and related side

effects [89,90]. However, as clearly visible from our study, 8 enzymes shared significant sequence homology with the critical human proteins. Therefore, these enzymes were eliminated from this study. The remaining 15 non-host homologous enzymes were prioritised for further examination.

To be considered a therapeutic target, a protein should be critically involved in the signalling or metabolic pathway of a virus. Secondly, the biological function of this protein must be tuneable by binding of the drug candidates with high affinity. These combined features were together termed “druggability” by Hopkins and Groom [91]. Therefore, druggability is a critical step in the selection, categorisation, and validation of suitable drug targets in the early stages of drug discovery. Unfortunately, we were impelled to temporarily halt our pursuit of assessing the druggability of 8 enzymes because of some unavoidable circumstances mentioned previously. Accordingly, the 7 remaining host non-homologous enzymes were subjected to druggability assessment using CavityPlus [62], DoGSiteScorer [63] and DeepSite [64]. The

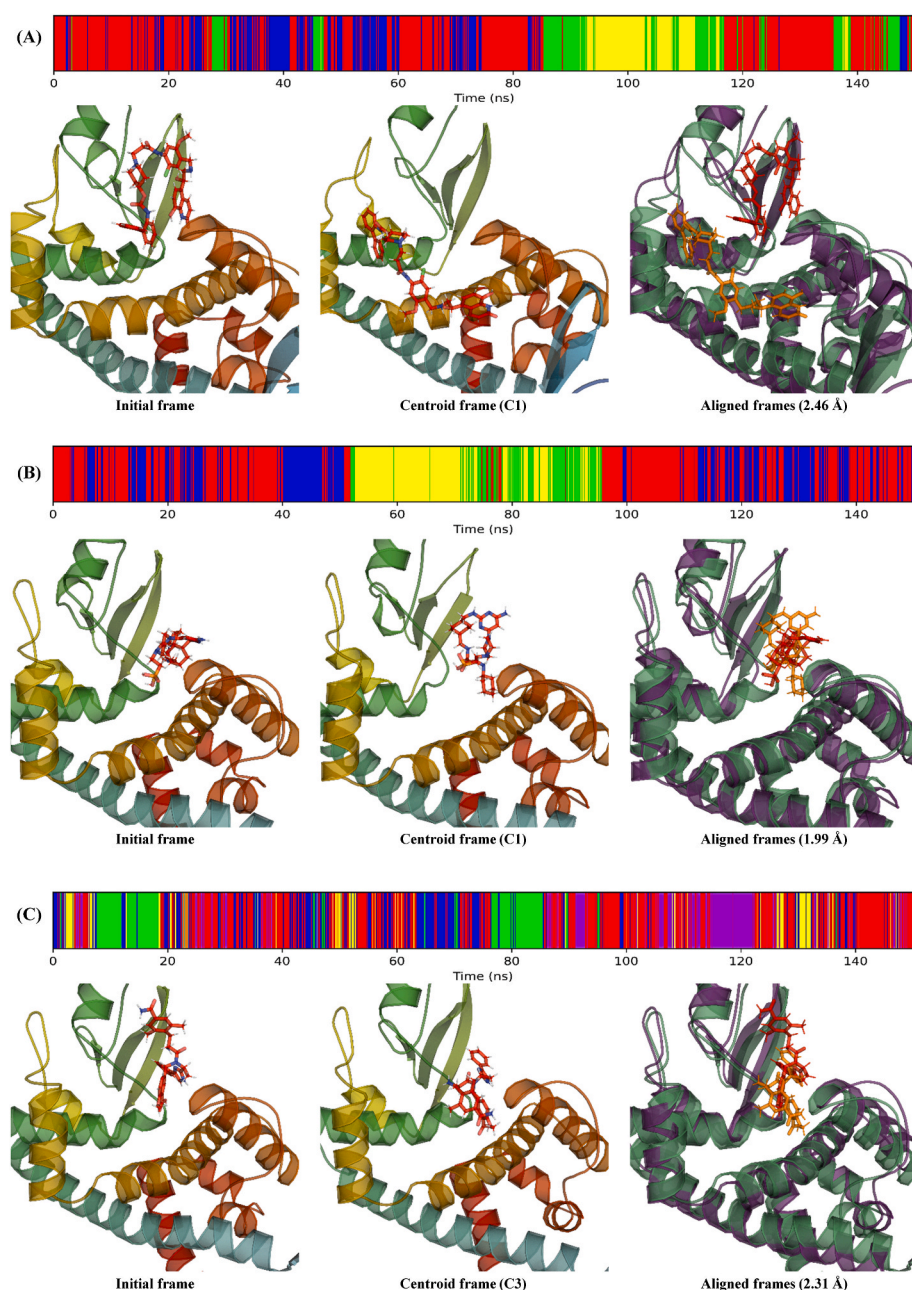


Fig. 8. Cluster analysis of the Top1B ligand bound complexes. (A) Top1B-BAT. (B) Top1B-BUR. (C) Top1B-ELU. The linear plot depicting the distribution of conformations (frames) over 150 ns of simulation. Close view of initial frame, centroid frame of the most populated cluster and 3D alignment of initial frame with centroid frame are presented below each linear plot from left to right. The total number of clusters and frames per cluster has been detailed in [Supplementary Fig. S6](#). The colours in the time evolution linear plots represent the cluster to which a frame belongs at a given point of time. The representation of colouring scheme are same as described in [Fig. 7](#).

outcomes were interesting and, in a manner, similar to Cheng et al. (2007); only 4 key enzymes of mpox virus were predicted as “druggable”. The remaining 3 enzymes namely, DNA-dependent RNA polymerase 18 kDa subunit (Rpo18), mRNA-capping enzyme 33 kDa small subunit (D12L) and DNA-dependent RNA polymerase 19 kDa subunit (Rpo19) were predicted as “undruggable”. Taken together, 4 highly druggable, host non-homologous enzymes of mpox virus were identified. Their physicochemical properties supported their stability and suitability for further analysis.

The I7L gene product i.e., core cysteine protease was identified as one of the four therapeutic targets. This is the main candidate protein for viral core protein proteinase (vCPP) activity in the viruses which is essential for the production of other non-structural proteins. Like other virus core proteases, the mpox virus I7L protein also contains putative catalytic residues (His241 and Cys328) in its highly druggable binding site [92]. Top1B DNA topoisomerase relaxes DNA supercoils by iteratively cleaving and rejoining one strand of the DNA duplex through a

covalent DNA-(3'-phosphotyrosyl)-enzyme intermediate [93]. This makes it a critical chokepoint of the mpox virus replication machinery and an ideal therapeutic target. VETFS was identified as the third target. This protein is expressed late in the infectious cycle of orthopox virus and regulates the cascade mechanism of gene transcription in the virus. VETFS is recruited by RNA polymerase associated protein (RAP94, 94 kDa) to form the early transcription complex including core RNA polymerase and binds to early gene promoters [94]. Inhibition of this can potentially prevent the assembly of the functional early transcription complex necessary for the successful expression of genes pertaining to the regulation of its lifecycle, virulence, and immunomodulation. Previous studies [95,96] suggest A20R in vaccinia virus as a processivity factor on the essentially distributive viral DNA polymerase along with D4R, D5R and H5R proteins [95]. Thus, targeting these core enzymes can choke the replication, translation and assembly of the mpox virus into the mature virion particles. Therefore, these 4 proteins can be considered potential therapeutic targets for mpox. However, the actual

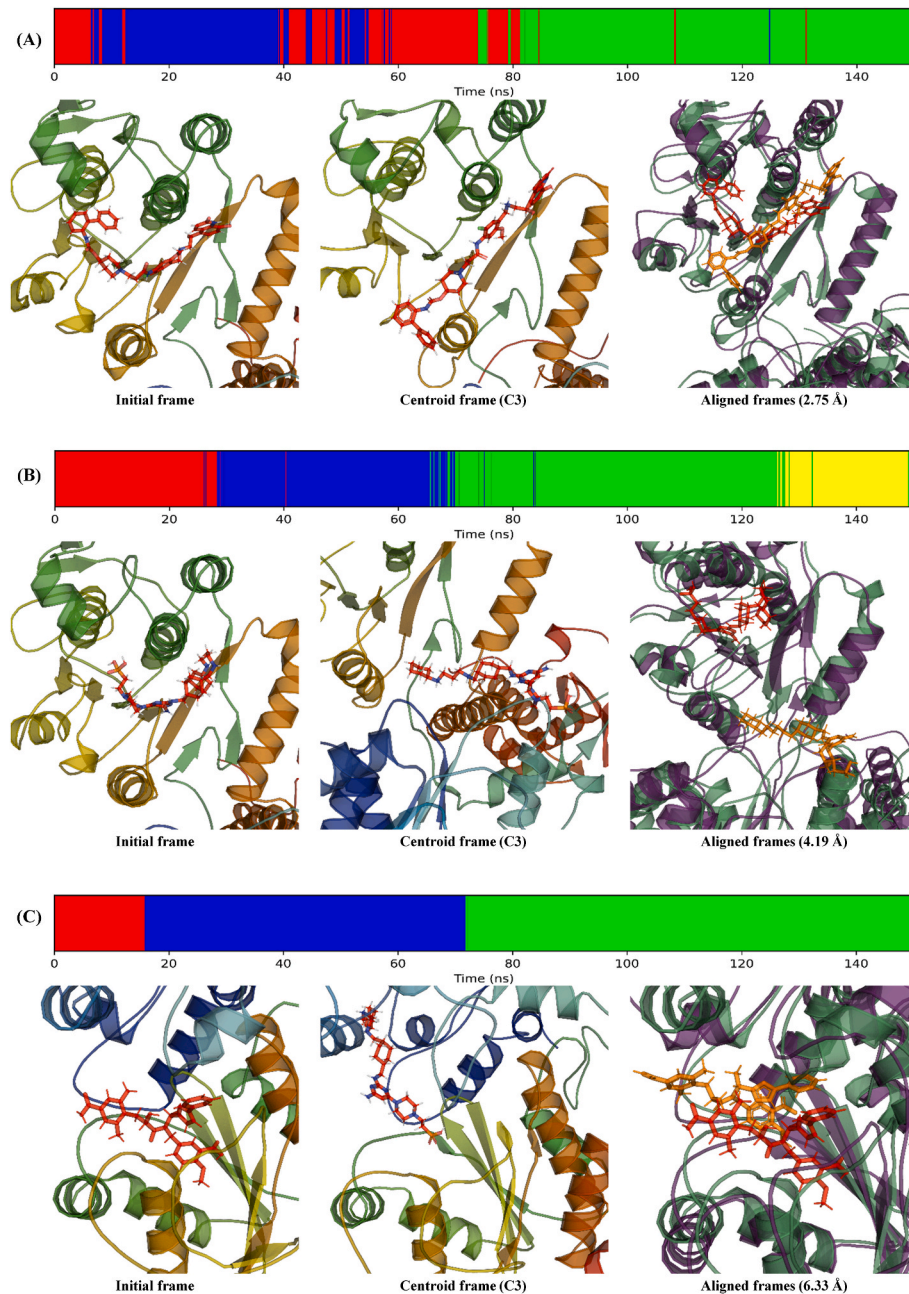


Fig. 9. Cluster analysis of the VETFS ligand bound complexes. (A) VETFS-BAT. (B) VETFS-BUR. (C) VETFS-ELU. The linear plot depicting the distribution of conformations (frames) over 150 ns of simulation. Close view of initial frame, centroid frame of the most populated cluster and 3D alignment of initial frame with centroid frame are presented below each linear plot from left to right. The total number of clusters and frames per cluster has been detailed in [Supplementary Fig. S6](#). The colours in the time evolution linear plots represent the cluster to which a frame belongs at a given point of time. The representation of colouring scheme are same as described in [Fig. 7](#).

Table 5

Protein-ligand interactions of the representative centroid complex of each cluster that falls within the stable region (RMSD) of the trajectories in every complex.

Target	Ligand	Cluster	Hydrogen Bond				Hydrophobic Interaction		Halogen Bonds			Salt Bridge		
			Residue	Sidechain	Distance (H-A, Å)	Distance (D-A, Å)	Residue	Distance (Å)	Residue	Sidechain	Distance (Å)	Residue	Distance (Å)	
I7L	Batefenterol	C2	GLU108	FALSE	2.09	2.97	LEU109	3.70						
			ASN295	TRUE	3.51	3.88	TYR238	3.63						
							LYS243	3.78						
							ILE298	3.73						
			C3	SER078	FALSE	2.52	3.46	LEU077	3.87					
		SER078		TRUE	2.41	3.21	ARG099	3.73	GLU108	TRUE	2.87			
		SER078		TRUE	2.96	3.81	PRO264	3.66						
		ARG099		TRUE	3.40	4.10	PHE267	3.64						
		TYR100		TRUE	2.47	3.35	PHE275	3.70						
		GLU108		FALSE	1.79	2.76	PHE278	3.83						
								ILE298	3.77					
								ILE316	3.62					
		C4	HIS076	FALSE	2.89	3.41	ARG099	3.81	GLU108	TRUE	2.62			
	SER078		FALSE	3.25	3.69	LEU109	3.53							
	SER078		TRUE	2.37	3.21	ASN262	3.97							
	SER078		TRUE	3.13	4.06	PHE267	3.89							
	ARG099		TRUE	3.02	3.44	PHE275	3.89							
	ARG099		TRUE	3.08	4.06	ILE316	3.76							
		Burixafor	C2	THR111	TRUE	2.74	3.55	VAL047	3.55				ASP297	4.61
	SER200			FALSE	3.03	4.01	PRO064	3.91						
	LYS206			TRUE	2.65	3.38								
	LYS243			TRUE	3.23	3.76								
	ASN293		FALSE	2.13	2.96									
	THR294		TRUE	3.03	3.94									
	ASN295		TRUE	1.81	2.77									
			C3	THR111	TRUE	2.86	3.59	TYR046	3.49					
THR111	TRUE			2.99	3.59	LYS050	3.55							
LYS243	TRUE			3.63	4.08	TYR051	3.58							
						VAL060	3.97							
						PRO064	3.84							
	Eluxadoline	C2	THR111	TRUE	1.99	2.89						LYS243	5.00	
SER240			FALSE	2.20	3.18									
		C3	SER240	FALSE	2.04	3.02	TYR046	3.81				LYS243	4.37	
HIS241			FALSE	3.16	3.48	ILE113	3.64							
						TYR238	3.76							
						LEU239	3.70							
						LEU239	3.83							
Top1B	Batefenterol	C1	ARG223	TRUE	3.42	3.87	LEU122	3.94						
			ARG223	FALSE	3.06	3.85	LEU122	3.67						
			HIS265	TRUE	3.02	3.48	PHE131	3.40						
			TYR274	FALSE	1.78	2.76	PHE131	3.90						
							LEU146	3.90						
							GLU205	3.66						
						VAL208	3.78							
						VAL208	3.79							
						ARG223	3.32							
						VAL227	3.96							
			C2	MET126	FALSE	2.30	3.26	LEU122	3.84					
		ILE129		FALSE	2.03	2.90	ARG130	3.88						
	ARG130	TRUE		3.17	3.76	PHE131	3.58							
	ARG130	TRUE		3.18	3.76	VAL208	3.47							
	PHE131	FALSE		2.46	3.39	ILE212	3.63							
	GLY132	FALSE		2.33	3.09	LEU222	3.53							
		Burixafor	C1	PHE128	FALSE	3.18	3.85	PHE128	3.57					
	ARG130			TRUE	2.88	3.44	PHE128	3.85						
	HIS172			TRUE	2.31	3.06	ILE129	3.79						
						PHE164	3.29							
					HIS172	3.97								
					PHE174	3.99								
					TYR233	3.74								
	C2		MET126	FALSE	2.02	2.97	ILE129	3.92						
ARG130		FALSE	2.75	3.39	PHE164	3.62								

(continued on next page)

Table 5 (continued)

Target	Ligand	Cluster	Hydrogen Bond				Hydrophobic Interaction		Halogen Bonds			Salt Bridge	
			Residue	Sidechain	Distance (H-A, Å)	Distance (D-A, Å)	Residue	Distance (Å)	Residue	Sidechain	Distance (Å)	Residue	Distance (Å)
Eluxadoline	C1	GLU173	FALSE	2.02	2.94	VAL175	3.60						
		ARG223	TRUE	1.80	2.75								
		ARG223	TRUE	1.89	2.80								
		ARG130	FALSE	3.06	3.56	HIS172	3.87						
		PHE131	FALSE	2.02	3.02								
		ARG223	TRUE	3.17	4.03								
		THR230	TRUE	1.86	2.84								
		VAL262	FALSE	1.86	2.73								
		ARG130	FALSE	2.81	3.30								
		PHE131	FALSE	1.71	2.69							ARG223	4.80
		ARG223	TRUE	2.22	3.01								
		VAL262	FALSE	2.22	2.90								
		GLY264	FALSE	2.49	3.15								
		ARG130	FALSE	2.24	2.97	ARG130	3.80						
		PHE131	FALSE	2.22	3.20	PHE164	3.88						
		VAL262	FALSE	1.93	2.91	HIS172	3.72						
		VAL263	FALSE	2.80	3.79								
		VETFS	Batefenterol	ARG262	TRUE	1.65	2.65	PHE271	3.47				ASP443
ASP272	FALSE			2.38	3.36	TYR276	3.90						
						TYR276	3.92						
						THR444	3.78						
TYR258	TRUE			1.97	2.88	ASP272	3.72				ASP443	5.07	
ARG262	TRUE			3.14	3.84	MET275	3.78						
ARG262	TRUE			2.38	3.30	TYR477	3.61						
ASP272	FALSE			2.10	3.06								
ARG262	TRUE			1.86	2.82	PHE271	3.60	LYS273	TRUE	3.59	ASP443	5.25	
ASP272	FALSE			2.10	3.09	TYR276	3.53						
						TYR276	3.62						
						ASP443	3.91						
						TYR477	3.54						
ASN170	TRUE			2.93	3.63	PHE445	3.88						
ASN497	FALSE			3.12	4.05								
SER166	FALSE			1.82	2.79								
GLU494	TRUE			2.69	3.52	PHE445	3.88						
GLU494	TRUE			2.62	3.52								
ASN497	FALSE	2.25	3.09										
Eluxadoline	C3	ARG262	TRUE	1.95	2.88	PHE271	3.83						
		ARG262	TRUE	3.34	3.93	PHE271	3.68						
						MET275	3.67						
						TYR477	3.70						

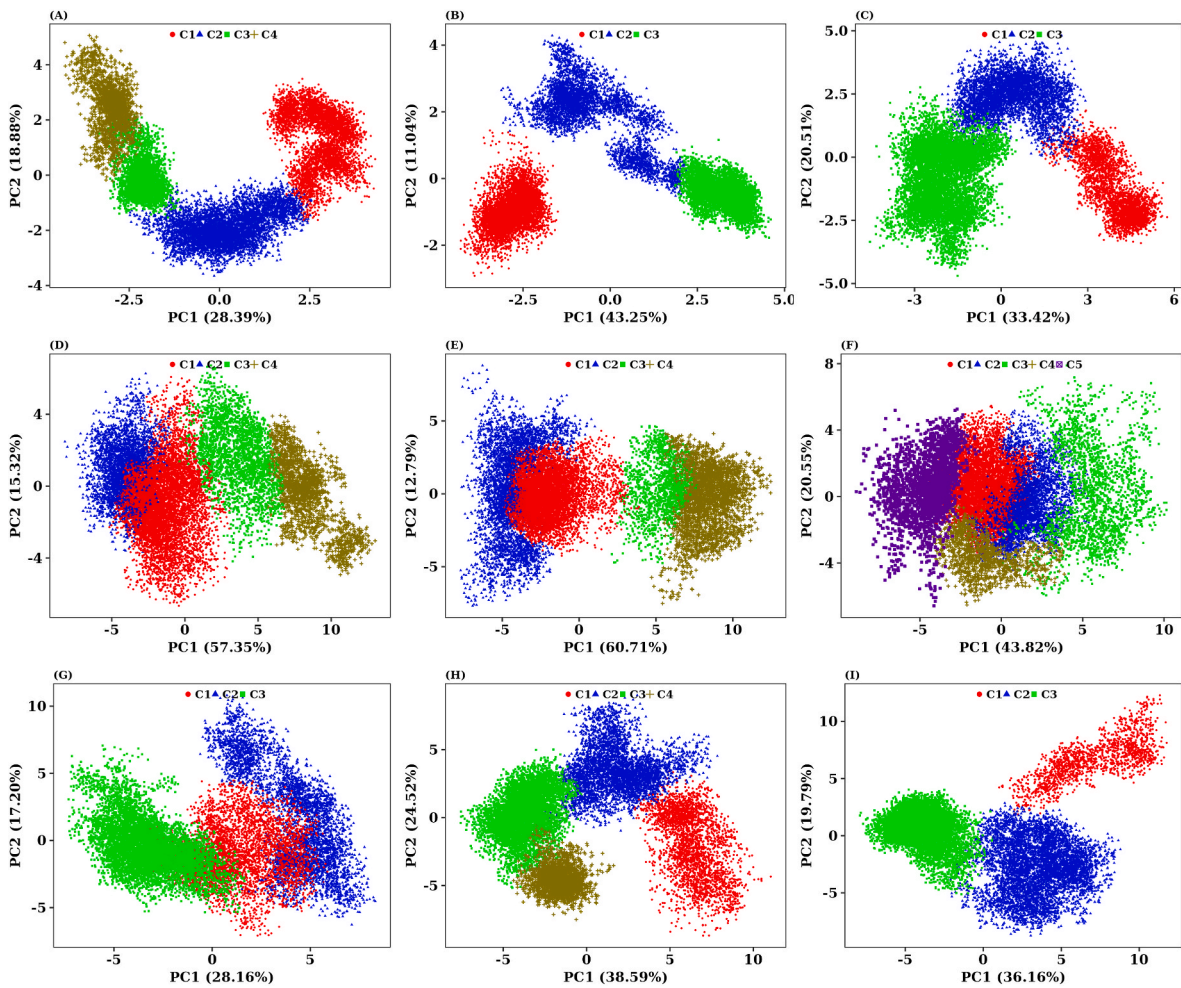


Fig. 10. Principal component analysis revealing the dynamics of motion for all nine mapped ligand-bound complexes. Projection of motion in phase space along the PC1 and PC2 is drawn as scatter plot. (A) PCA plot of I7L-BAT complex. (B) PCA plot of I7L-BUR complex. (C) PCA plot of I7L-ELU complex. (D) (A) PCA plot of Top1B-BAT complex. (E) PCA plot of Top1B-BUR complex. (F) PCA plot of Top1B-ELU complex. (G) PCA plot of VETFS-BAT complex. (H) PCA plot of VETFS-BUR complex. (I) PCA plot of VETFS-ELU complex. The coloured dots in every PCA plot correspond to the sub clusters formed by each protein in complex with its respective ligand during 150 ns of simulation. The explained variance (%) of each principal component are presented in brackets.

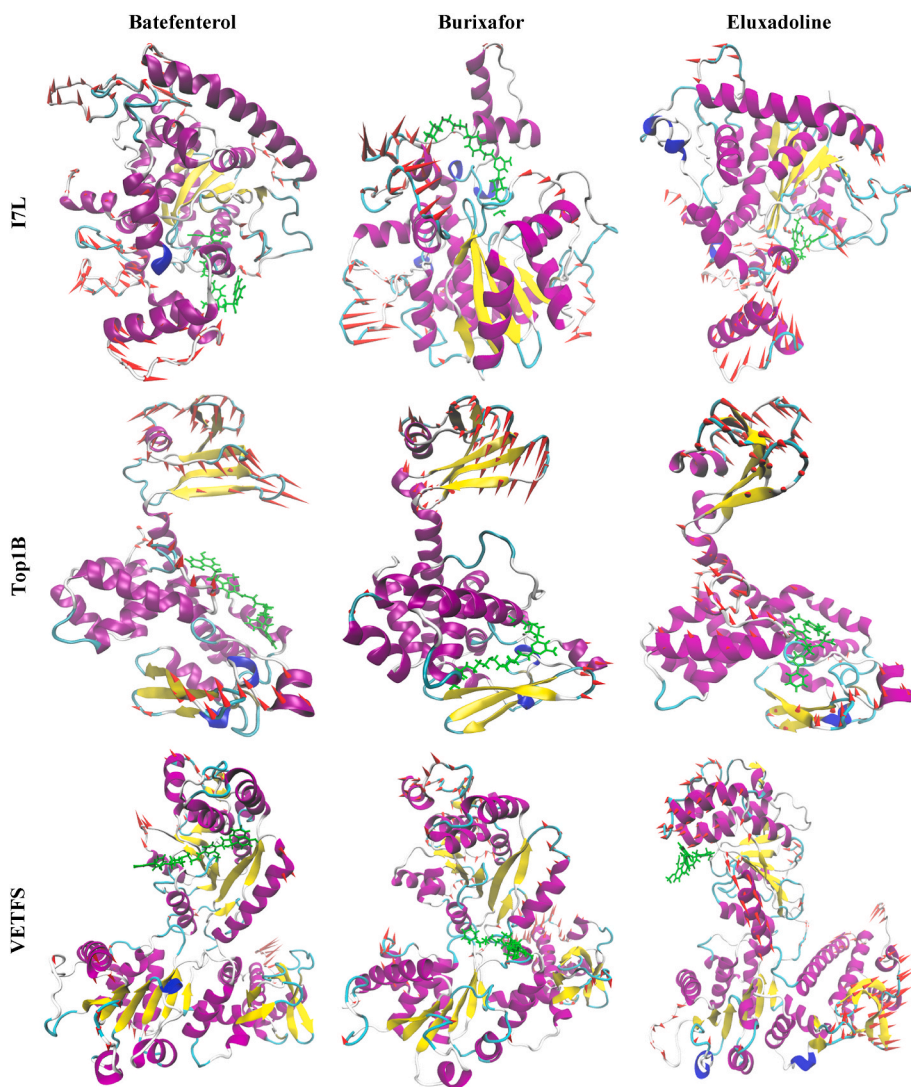


Fig. 11. Porcupine plot displaying the motion of mpx virus proteins I7L, Top1B and VETFS with the ligands BAT, BUR and ELU obtained from PC1 during the 150 ns of simulation. The colours pink, blue, yellow and white represent α -helix, 3–10 helix, β -sheet and loops/coils of the proteins respectively. The ligands bound to the proteins are depicted in green colour. The fluctuations in $C\alpha$ atoms above 2 Å are presented in the form of red coloured cones where the length of the cones represent amplitude of the $C\alpha$ atoms from their mean positions. The illustrations were created with VMD.

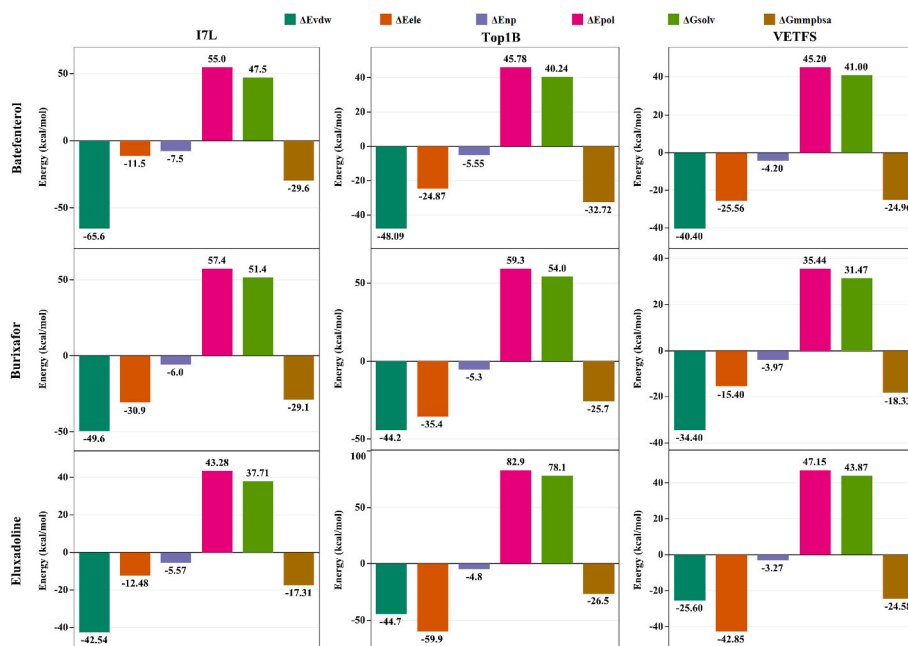


Fig. 12. The binding free energy terms derived from MM/PBSA calculations relative to the binding of BAT, BUR and ELU with the three proteins I7L, Top1B and VETFS. The colour coding schemes for different energy terms are shown at the top of the figure.

expression of these targets must be verified, and their cellular functions must be characterised to validate them.

With drug repurposing as the objective, a highly curated library of 5893 FDA approved/investigational drugs was screened through HTVS against these targets. We employed the Schrodinger suite's Glide module for this purpose. The Glide module follows a three-stage screening protocol namely, HTVS, SP and XP where the sampling and scoring gradually become more extensive and stringent [69,97]. Moreover, the XGlide module allowed us to screen the library against multiple target proteins simultaneously. A list of the top 30 hits corresponding to 0.5% of the drugs contained in the FDA approved/investigational library was also compiled for every target protein. BAT, BUR and ELU were the common top-ranked hits against every protein target as visible from the list of top 30 hits. Targeting multiple non-structural proteins is expected to enhance antiviral potency [98,99]. It is also expected to overcome resistance due to point mutations in targets. Since some of these targets are also conserved in other strains/viruses, multi-targeting can lead to a broad spectrum of action [100]. BAT is an approved drug for treating pulmonary diseases [101,102]. BUR is used for treating multiple myeloma, Hodgkin's disease, and non-Hodgkin's lymphoma [103]. ELU is primarily prescribed for bowel syndrome with diarrhea [104]. Although the primary indication of these drugs do not overlap with viral pathology, any overlap of the gene involved in these indications with that of viral disease cannot be ruled out at this stage.

Molecular dynamics (MD) simulation mimics the flexible behaviour of biomolecules, conformational changes in proteins and protein-ligand interactions to paint a realistic picture with atomic resolution with reference to time. Therefore, to gain further insights into the structural dynamics and stabilities of BAT, BUR and ELU with all the enzymes namely, A20R, I7L, Top1B and VETFS, MD simulations were executed for 150 ns for every complex.

A set of quality control parameters like RMSD, RMSF, Rg, SASA and H-Bond were applied on all the twelve protein-ligand complexes to infer their stability. Unfortunately, none of the ligands could stabilize the enzyme A20R. Although the RMSD trajectory of A20R-BUR appeared to be stable for the first 90 ns, it continued its upward ascension for the remaining time duration. Hence, it was not possible to draw a reliable conclusion about the stability of A20R-BUR complex. On the contrary, RMSD trajectories of the remaining 9 complexes suggested that BAT,

BUR and ELU successfully stabilized the enzymes I7L, Top1B and VETFS. Their RMSD trajectories were also coherent with their Rg and SASA trajectories. Such findings suggest the stable ligand bound complexes have successfully retained their compactness throughout the simulations. Prominent transformations in the secondary structures of some stable ligand bound proteins were also identified after examining their time dependent secondary structure evolution graph. These transformations in the secondary structural elements might be key drivers of ligand dependent stabilization of these complexes. Therefore, we believe a better understanding of these structural changes may be fundamental in designing antiviral therapeutics against these proteins of mpox virus.

The complexes of the enzyme Top1B presented a very interesting case in itself. The RMSDs of the Top1B ligand bound complexes fluctuated irregularly while transitioning between their stable conformations. However, upon further examining their Rg, SASA, H-Bond and PCA plots, a more cleaner picture of their relative stability was obtained. The assessment of RMSF plots and porcupine plots further vindicated the stabilities of these complexes. Top1B is a multidomain protein comprising of a N-terminal DNA binding domain and a C-terminal domain which is separated by a hinge comprising of three residues Gly73, Lys71 and Met75. Our assessments suggest this hinge has added to the flexibility of Top1B's overall structure by granting a higher degree of flexibility to the N-terminal domain whereas the desired target DNA binding C-terminal domain was relatively stable upon ligand binding. Therefore, in future drug discovery endeavours researchers are expected to be mindful of studying the two domains of Top1B enzyme independently to avoid false negative predictions. Similarly, the RMSD trajectories and all the quality control parameters clearly indicated towards the stable binding of the VETFS-BUR complex initially. However, further assessment of the distance between the centre of mass of the ligand BUR and the interacting active site residues of VETFS suggests that the ligand has moved away from the active site pocket and is therefore no longer bound to it. This somewhat explains the increased compactness (Rg) of VETFS-BUR complex which was wrongly interpreted for a stable protein-ligand complex earlier.

Clinically, BAT does not show any general or cardiovascular adverse events when inhaled even after 6 continuous weeks of use for chronic obstructive pulmonary disease [105]. Burixafor is also well tolerated in clinical applications against myeloma and lymphoma [103]. Besides,

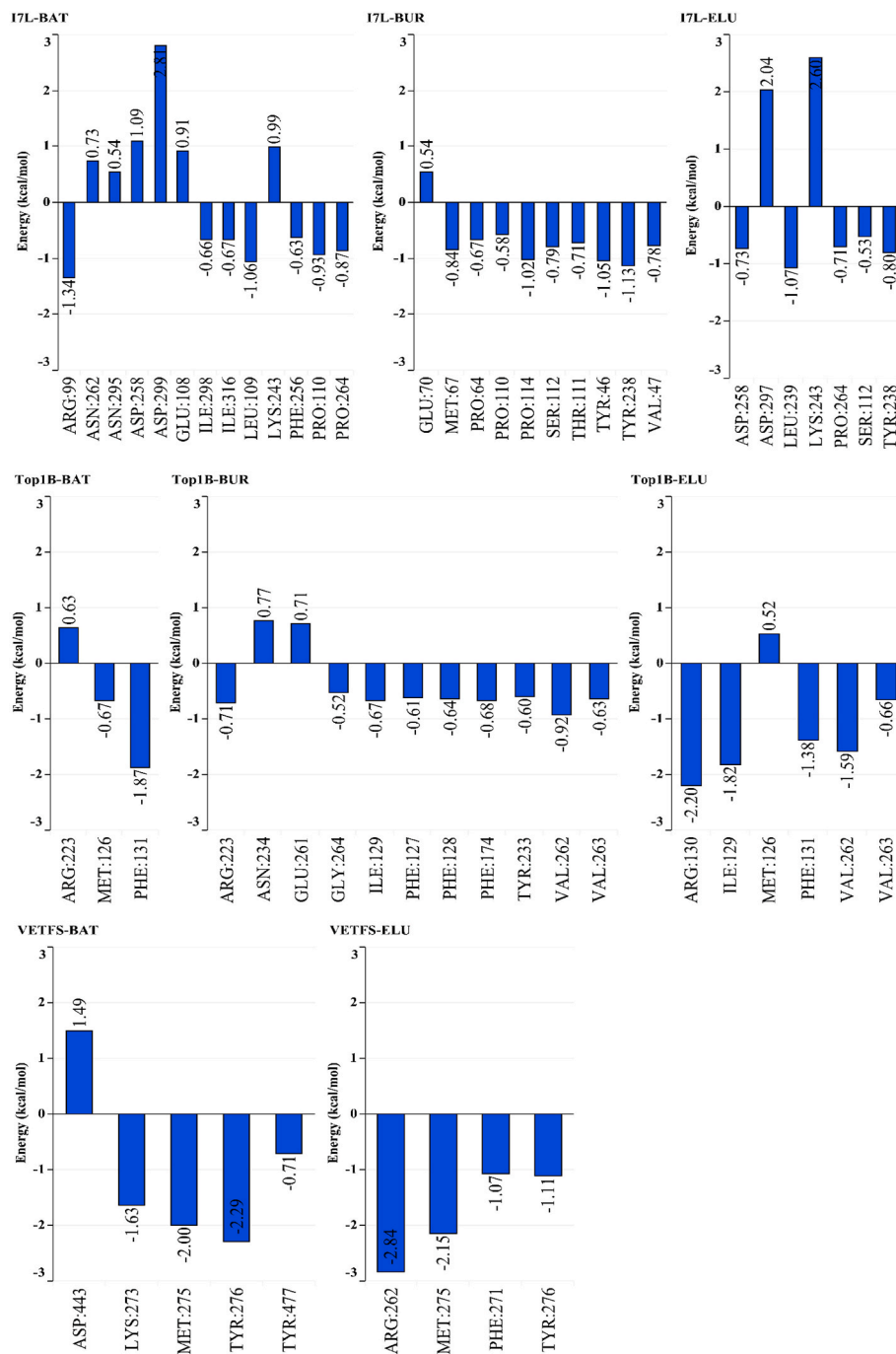


Fig. 13. The per residue decomposition plots (MM/PBSA) representing the binding free energy contributed by the residues of the active sites of the three proteins I7L, Top1B and VETFS which energetically contribute towards the stabilization of the three ligands BAT, BUR and ELU.

ELU has been demonstrated to be safe in a 12 week long treatment against irritable bowel syndrome with diarrhea [106]. Taken together BAT, BUR and ELU show potential for repurposing against mpox. However, it requires further experimental, pre-clinical as well as clinical validations. Nonetheless, this is an attempt to leverage multi-omics mining strategies to identify therapeutic targets of mpox virus and their potential inhibitors. As newly sequenced genomes of mpox virus gradually get deposited in the public repositories, our findings can easily be reauthenticating by incorporating newer genomes. This can encourage research for repurposing drugs against mpox.

5. Conclusion

This study demonstrated the integration of genomics and subtractive proteomics to identify therapeutic targets for mpox virus. Although this requires further validation, our *in silico* analysis and prior literature indicate their druggability and relevance as therapeutic targets. This can encourage structure-based analysis for finding antivirals for mpox. The virtual screening followed by MD simulation suggested batdefenterol, burixafor and eluxadolone as potential inhibitors of multiple mpox virus targets. The clinical safety of these drugs also supports their suitability for repurposing. This can encourage experimental validations to repurpose these as antivirals for the therapeutic management of mpox virus infection.

Author contributions

Conceptualization: AS, BBS; Formal Analysis: AS, MG; Funding Acquisition: BBS; Investigation: AS, MG; Methodology: AS, MG, NCM, BBS; Project administration: BBS; Resources: BBS; Supervision: BBS; Visualization: AS, MG; Writing – original draft: AS, MG; Writing – review & editing: AS, MG, NCM, ES, RPS, BBS.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that can influence the work reported in this paper.

Acknowledgements

We thank Dean, School of Pharmaceutical Sciences, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India for the facilities. MG is supported by a fellowship funded from by Department of Biotechnology (DBT), Ministry of Science and Technology, New Delhi, India (Grant Id: BT/INF/22/SP45078/2022). The authors would like to acknowledge the Indian Council of Medical Research (ICMR), Ministry of Health & Family Welfare, New Delhi, India (Grant ID: AMR/DHR/GIA/4/ECD-II-2020) for providing high-performance computational resources for this study. We acknowledge the Bioinformatics Resources and Applications Facility (BRAAF), C-DAC, Pune, India for providing timely access to their high-performance computing clusters, which allowed us to perform complex simulations and data analysis for this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compbmed.2023.106971>.

References

- [1] World Health Organization (16 May 2022), Monkeypox - United Kingdom of Great Britain and Northern Ireland. <https://www.who.int/emergencies/disease-outbreak-news/item/2022-DON381>, 2022.
- [2] World Health Organization, WHO Director-General's statement at the press conference following IHR Emergency Committee regarding the multi-country

- outbreak of monkeypox - 23 July 2022. <https://www.who.int/director-general/speeches/detail/who-director-general-s-statement-on-the-press-conference-following-IHR-emergency-committee-regarding-the-multi-country-outbreak-of-monkeypox-23-july-2022>, 2022.
- [3] M. Kozlov, Monkeypox goes global: why scientists are on alert, *Nature* 606 (2022) 15–16, <https://doi.org/10.1038/d41586-022-01421-8>.
- [4] (ICTV) ICoToV. Virus Taxonomy, Release. <https://talk.ictvonline.org/taxonomy>, 2020.
- [5] P. von Magnus, E.K. Andersen, K.B. Petersen, A. Birch-Andersen, A pox-like disease in cynomolgus monkeys, *acta pathol. Microbiol. Scand* 46 (1959) 156–176, <https://doi.org/10.1111/j.1699-0463.1959.tb00328.x>.
- [6] J.G. Breman, Kalisa-ruti, M. V. Steniowski, E. Zanotto, A.I. Gromyko, I. Arita, human monkeypox, 1970-79, *Bull. World Health Organ.* 58 (1980) 165–182. <http://www.ncbi.nlm.nih.gov/pubmed/6249508>.
- [7] E.M. Bunge, B. Hoet, L. Chen, F. Lienert, H. Weidenthaler, L.R. Baer, R. Steffen, The changing epidemiology of human monkeypox—a potential threat? A systematic review, *PLoS Negl. Trop. Dis.* 16 (2022), <https://doi.org/10.1371/journal.pntd.0010141>.
- [8] Centers for Disease Control and Prevention (CDC), Multistate outbreak of monkeypox—Illinois, Indiana, and Wisconsin, 2003, *MMWR, Morb. Mortal. Wkly. Rep.* 52 (2003) 537–540. <http://www.ncbi.nlm.nih.gov/pubmed/12803191>.
- [9] Centers for Disease Control and Prevention (CDC), Update: multistate outbreak of monkeypox—Illinois, Indiana, Kansas, Missouri, Ohio, and Wisconsin, 2003., *MMWR, Morb. Mortal. Wkly. Rep.* 52 (2003) 561–564. <http://www.ncbi.nlm.nih.gov/pubmed/12816106>.
- [10] Centers for Disease Control and Prevention (CDC), Update: multistate outbreak of monkeypox—Illinois, Indiana, Kansas, Missouri, Ohio, and Wisconsin, 2003., *MMWR, Morb. Mortal. Wkly. Rep.* 52 (2003) 642–646. <http://www.ncbi.nlm.nih.gov/pubmed/12855947>.
- [11] K.D. Reed, J.W. Melski, M.B. Graham, R.L. Regnery, M.J. Sotir, M. V. Wegner, J. J. Kazmierczak, E.J. Stratman, Y. Li, J.A. Fairley, G.R. Swain, V.A. Olson, E. K. Sargent, S.C. Kehl, M.A. France, R. Kline, S.L. Foldy, J.P. Davis, I.K. Damon, The detection of monkeypox in humans in the Western Hemisphere., *N. Engl. J. Med.* 350 (2004) 342–350, <https://doi.org/10.1056/NEJMoa032299>.
- [12] European Centre for Disease Prevention and Control, Risk assessment: monkeypox multi-country outbreak. <https://www.ecdc.europa.eu/en/publications-data/risk-assessment-monkeypox-multi-country-outbreak>, 2022.
- [13] E.F. Alakunle, M.I. Okeke, Monkeypox virus: a neglected zoonotic pathogen spreads globally, *Nat. Rev. Microbiol.* (2022), <https://doi.org/10.1038/s41579-022-00776-z>.
- [14] I.K. Damon, Poxviruses, *Manual Clin. Microbiol.* 2 (2006) 1631–1640, n.d.
- [15] ramazan azim okay, Another epidemic in the shadow of Covid 19 pandemic: a review of monkeypox, *Eurasian J. Med. Oncol.* (2022), <https://doi.org/10.14744/ejmo.2022.2022>.
- [16] S.N. Shchelkunov, A.V. Totmenin, I.V. Babkin, P.F. Safronov, O.I. Ryazankina, N. A. Petrov, V.V. Gutorov, E.A. Uvarova, M.V. Mikheev, J.R. Sisler, J.J. Esposito, P. B. Jahrling, B. Moss, L.S. Sandakchiev, Human monkeypox and smallpox viruses: genomic comparison, *FEBS Lett.* 509 (2001) 66–70, [https://doi.org/10.1016/S0014-5793\(01\)03144-1](https://doi.org/10.1016/S0014-5793(01)03144-1).
- [17] J.R. Kugelman, S.C. Johnston, P.M. Mulembakani, N. Kialu, M.S. Lee, G. Koroleva, S.E. McCarthy, M.C. Gestole, N.D. Wolfe, J.N. Fair, B.S. Schneider, L. L. Wright, J. Huggins, C.A. Whitehouse, E.O. Wemakoy, J.J. Muyembe-Tamfum, L.E. Hensley, G.F. Palacios, A.W. Rimoin, Genomic variability of monkeypox virus among humans, Democratic Republic of the Congo, *Emerg. Infect. Dis.* 20 (2014) 232–239, <https://doi.org/10.3201/eid2002.130118>.
- [18] D. Kmiec, F. Kirchhoff, Monkeypox: a new threat? *Int. J. Mol. Sci.* 23 (2022) <https://doi.org/10.3390/ijms23147866>.
- [19] B.T. Seet, J.B. Johnston, C.R. Brunetti, J.W. Barrett, H. Everett, C. Cameron, J. Sypula, S.H. Nazarian, A. Lucas, G. McFadden, Poxviruses and immune evasion, *Annu. Rev. Immunol.* 21 (2003) 377–423, <https://doi.org/10.1146/annurev.immunol.21.120601.141049>.
- [20] T.G. Senkevich, N. Yutin, Y.I. Wolf, E.V. Koonin, B. Moss, Ancient gene capture and recent gene loss shape the evolution of orthopoxvirus-host interaction genes, *mBio* 12 (2021), <https://doi.org/10.1128/mBio.01495-21>.
- [21] S.S. Marennikova, E.M. Seluhina, N.N. Mal'ceva, K.L. Cimiskjan, G.R. Macevic, Isolation and properties of the causal agent of a new variola-like disease (monkeypox) in man, *Bull. World Health Organ.* 46 (1972) 599–611.
- [22] G. McFadden, Poxvirus tropism, *Nat. Rev. Microbiol.* 3 (2005) 201–213, <https://doi.org/10.1038/nrmicro1099>.
- [23] S. Realegeno, A.S. Puschnik, A. Kumar, C. Goldsmith, J. Burgado, S. Sambhara, V. A. Olson, D. Carroll, I. Damon, T. Hirata, T. Kinoshita, J.E. Carette, P. S. Satheshkumar, Monkeypox virus host factor screen using haploid cells identifies essential role of GARP complex in extracellular virus formation, *J. Virol.* 91 (2017), <https://doi.org/10.1128/jvi.00011-17>.
- [24] B. Moss, Membrane fusion during poxvirus entry, *Semin. Cell Dev. Biol.* 60 (2016) 89–96, <https://doi.org/10.1016/j.semcdb.2016.07.015>.
- [25] M.E. Wilson, J.M. Hughes, A.M. McCollum, I.K. Damon, Human monkeypox, *Clin. Infect. Dis.* 58 (2014) 260–267, <https://doi.org/10.1093/cid/cit703>.
- [26] A. Yinka-Ogunleye, O. Aruna, M. Dalhat, D. Ogoina, A. McCollum, Y. Disu, I. Mamadu, A. Akinpelu, A. Ahmad, J. Burga, A. Ndorero, E. Nkunzimana, L. Manneh, A. Mohammed, O. Adeoye, D. Tom-Aba, B. Silenou, O. Ipadeola, M. Saleh, A. Adeyemo, I. Nwadiutor, N. Aworabhi, P. Uke, D. John, P. Wakama, M. Reynolds, M.R. Mauldin, J. Doty, K. Wilkins, J. Musa, A. Khalakdina, A. Adedeji, N. Mba, O. Ojo, G. Krause, C. Ihekweazu, A. Mandra, W. Davidson, V. Olson, Y. Li, K. Radford, H. Zhao, M. Townsend, J. Burgado, P. S. Satheshkumar, Outbreak of human monkeypox in Nigeria in 2017–18: a

- clinical and epidemiological report, *Lancet Infect. Dis.* 19 (2019) 872–879, [https://doi.org/10.1016/S1473-3099\(19\)30294-4](https://doi.org/10.1016/S1473-3099(19)30294-4).
- [27] K. Simpson, D. Heymann, C.S. Brown, W.J. Edmunds, J. Elsgaard, P. Fine, H. Hochrein, N.A. Hoff, A. Green, C. Ihekweazu, T.C. Jones, S. Lule, J. MacLennan, A. McCollum, B. Mühlemann, E. Nightingale, D. Ogoina, A. Ogunleye, B. Petersen, J. Powell, O. Quantick, A.W. Rimoin, D. Ulaeato, A. Wapling, Human monkeypox – after 40 years, an unintended consequence of smallpox eradication, *Vaccine* 38 (2020) 5077–5081, <https://doi.org/10.1016/j.vaccine.2020.04.062>.
- [28] C. Chakraborty, M. Bhattacharya, S.S. Nandi, R.K. Mohapatra, K. Dhama, G. Agoramorthy, Appearance and re-appearance of zoonotic disease during the pandemic period: long-term monitoring and analysis of zoonosis is crucial to confirm the animal origin of SARS-CoV-2 and monkeypox virus, *Vet. Q.* 42 (2022) 119–124, <https://doi.org/10.1080/01652176.2022.2086718>.
- [29] M.G. Reynolds, A.M. McCollum, B. Nguete, R.S. Lushima, B.W. Petersen, Improving the care and treatment of monkeypox patients in low-resource settings: applying evidence from contemporary biomedical and smallpox biodefense research, *Viruses* 9 (2017), <https://doi.org/10.3390/v9120380>.
- [30] N. Sklenovská, M. Van Ranst, Emergence of monkeypox as the most important orthopoxvirus infection in humans, *front. Publ. Health* 6 (2018), <https://doi.org/10.3389/fpubh.2018.00241>.
- [31] J.P. Thornhill, S. Barkati, S. Walmsley, J. Rockstroh, A. Antinori, L.B. Harrison, R. Palich, A. Nori, I. Reeves, M.S. Habibi, V. Apea, C. Boesecke, L. Vandekerckhove, M. Yakubovsky, E. Sendagorta, J.L. Blanco, E. Florence, D. Moschese, F.M. Maltez, A. Goorhuis, V. Pouchier, P. Migaud, S. Noe, C. Pintado, F. Maggi, A.-B.E. Hansen, C. Hoffmann, J.I. Lezama, C. Mussini, A. Cattelan, K. Makofane, D. Tan, S. Nozza, J. Nemeth, M.B. Klein, C.M. Orkin, SHARE-Net clinical group, monkeypox virus infection in humans across 16 countries - april-june 2022, *N. Engl. J. Med.* (2022), <https://doi.org/10.1056/NEJMoa2207323>.
- [32] P.E.M. Fine, Z. Jezek, B. Grab, H. Dixon, The transmission potential of monkeypox virus in human populations, *Int. J. Epidemiol.* 17 (1988) 643–650, <https://doi.org/10.1093/ije/17.3.643>.
- [33] A.K. Rao, B.W. Petersen, F. Whitehill, J.H. Razeq, S.N. Isaacs, M.J. Merchlinsky, D. Campos-Outcalt, R.L. Morgan, I. Damon, P.J. Sánchez, B.P. Bell, Use of JYNNEOS (smallpox and monkeypox vaccine, live, nonreplicating) for preexposure vaccination of persons at risk for occupational exposure to orthopoxviruses: recommendations of the advisory committee on immunization practices — United States, 2022, *MMWR, Morb. Mortal. Wkly. Rep.* 71 (2022) 734–742, <https://doi.org/10.15585/mmwr.mm7122e1>.
- [34] Y. Xiang, A. White, Monkeypox Virus Emerges from The Shadow of Its More Infamous Cousin: Family Biology Matters, *Emerg. Microbes Infect.* (2022) 1–14, <https://doi.org/10.1080/22221751.2022.2095309>.
- [35] Y. Xiao, S.N. Isaacs, Therapeutic vaccines and antibodies for treatment of orthopoxvirus infections, *Viruses* 2 (2010) 2381–2403, <https://doi.org/10.3390/v2102381>.
- [36] G. Yang, D.C. Pevear, M.H. Davies, M.S. Collett, T. Bailey, S. Rippen, L. Barone, C. Burns, G. Rhodes, S. Tohan, J.W. Huggins, R.O. Baker, R.L.M. Buller, E. Touchette, K. Waller, J. Schriewer, J. Neyts, E. DeClercq, K. Jones, D. Hrubby, R. Jordan, An orally bioavailable antipoxvirus compound (ST-246) inhibits extracellular virus formation and protects mice from lethal orthopoxvirus challenge, *J. Virol.* 79 (2005) 13139–13149, <https://doi.org/10.1128/jvi.79.20.13139-13149.2005>.
- [37] G. Andrei, P. Fiten, M. Krečmerová, G. Opendakker, D. Topalis, R. Snoeck, Poxviruses bearing DNA polymerase mutations show complex patterns of cross-resistance, *Biomedicines* 580 (10) (2022), <https://doi.org/10.3390/biomedicines10030580>.
- [38] M. Karim, M. Nazrul Islam, G.M.N.A. Jewel, Silico identification of potential drug targets by subtractive genome analysis of *Enterococcus faecium* DO, *bioRxiv* (2020), <https://doi.org/10.1101/2020.02.14.948232>.
- [39] V. Solanki, M. Tiwari, V. Tiwari, Subtractive proteomic analysis of antigenic extracellular proteins and design a multi-epitope vaccine against *Staphylococcus aureus*, *Microbiol. Immunol.* 65 (2021) 302–316, <https://doi.org/10.1111/1348-0421.12870>.
- [40] S. Ahmad, A. Navid, A.S. Akhtar, S.S. Azam, A. Wadood, H. Pérez-Sánchez, Subtractive Genomics, Molecular Docking, Molecular Dynamics, Simulation revealed LpxC as a potential drug target against multi-drug resistant *Klebsiella pneumoniae*, *interdiscip. Sci. Comput. Life Sci.* 11 (2019) 508–526, <https://doi.org/10.1007/s12539-018-0299-y>.
- [41] V. Solanki, V. Tiwari, Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against *Acinetobacter baumannii*, *Sci. Rep.* 8 (2018), <https://doi.org/10.1038/s41598-018-26689-7>.
- [42] R. Uddin, F. Jamil, Prioritization of potential drug targets against *P. aeruginosa* by core proteomic analysis using computational subtractive genomics and Protein-Protein interaction network, *Comput. Biol. Chem.* 74 (2018) 115–122, <https://doi.org/10.1016/j.compbiolchem.2018.02.017>.
- [43] T. Seemann, Prokka: rapid prokaryotic genome annotation, *Bioinformatics* 30 (2014) 2068–2069, <https://doi.org/10.1093/bioinformatics/btu153>.
- [44] D.M. Kristensen, A.S. Waller, T. Yamada, P. Bork, A.R. Mushegian, E.V. Koonin, Orthologous Gene Clusters, Taxon Signature, Genes for viruses of prokaryotes, *J. Bacteriol.* 195 (2013) 941–950, <https://doi.org/10.1128/JB.01801-12>.
- [45] W.G. Nogueira, A.K. Jaiswal, S. Tiwari, R.T.J. Ramos, P. Ghosh, D. Barh, V. Azevedo, S.C. Soares, Computational identification of putative common genomic drug and vaccine targets in *Mycoplasma genitalium*, *Genomics* 113 (2021) 2730–2743, <https://doi.org/10.1016/j.ygeno.2021.06.011>.
- [46] D.M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics, *Genome Biol.* 238 (20) (2019), <https://doi.org/10.1186/s13059-019-1832-y>.
- [47] D.M. Emms, S. Kelly, OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy, *Genome Biol.* 16 (2015), <https://doi.org/10.1186/s13059-015-0721-2>.
- [48] M. Acebrón-García-de-Eulate, T.L. Lundell, S.C. Vedditi, Strategies for drug target identification in *Mycobacterium leprae*, *Drug Discov. Today Off.* 26 (2021) 1569–1573, <https://doi.org/10.1016/j.drudis.2021.03.026>.
- [49] J. Hallgren, K.D. Tsirigos, M. Damgaard Pedersen, J. Juan, A. Armenteros, P. Marcattili, H. Nielsen, A. Krogh, O. Winther, DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks, *bioRxiv* (2022), <https://doi.org/10.1101/2022.04.08.487609>, 2022.04.08.487609.
- [50] L. Dobson, I. Reményi, G.E. Tusnady, CCTOP: a Consensus Constrained TOPology prediction web server, *Nucleic Acids Res.* 43 (2015) W408–W412, <https://doi.org/10.1093/nar/gkv451>.
- [51] T. Nugent, D.T. Jones, Detecting pore-lining regions in transmembrane protein sequences, *BMC Bioinformatics* 169 (13) (2012), <https://doi.org/10.1186/1471-2105-13-169>.
- [52] K.D. Tsirigos, C. Peters, N. Shu, L. Käll, A. Elofsson, The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides, *Nucleic Acids Res.* 43 (2015) W401–W407, <https://doi.org/10.1093/nar/gkv485>.
- [53] Y. Li, S. Wang, R. Umarov, B. Xie, M. Fan, L. Li, X. Gao, DEEPre: sequence-based enzyme EC number prediction by deep learning, *Bioinformatics* 34 (2018) 760–769, <https://doi.org/10.1093/bioinformatics/btx680>.
- [54] M. Tahir ul Qamar, M.U. Mirza, J.-M. Song, M.J. Rao, X. Zhu, L.-L. Chen, Probing the structural basis of Citrus phytochrome B using computational modelling and molecular dynamics simulation approaches, *J. Mol. Liq.* 340 (2021), 116895, <https://doi.org/10.1016/j.molliq.2021.116895>.
- [55] M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G.R. Lee, J. Wang, Q. Cong, L.N. Kinch, R. Dustin Schaeffer, C. Millán, H. Park, C. Adams, C.R. Glassman, A. DeGiovanni, J.H. Pereira, A.V. Rodrigues, A.A. Van Dijk, A. C. Ebrecht, D.J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C.K. Yip, J.E. Burke, K. Christopher Garcia, N. V. Grishin, P.D. Adams, R.J. Read, D. Baker, Accurate prediction of protein structures and interactions using a three-track neural network, *Science* 373 (2021) 871–876, https://doi.org/10.1126/SCIENCE.ABJ8754/SUPPL_FILE/ABJ8754_MDAR_REPRODUCIBILITY_CHECKLIST.PDF.
- [56] C.J. Williams, J.J. Headd, N.W. Moriarty, M.G. Prisant, L.L. Videau, L.N. Deis, V. Verma, D.A. Keedy, B.J. Hintze, V.B. Chen, S. Jain, S.M. Lewis, W.B. Arendall, J. Snoeyink, P.D. Adams, S.C. Lovell, J.S. Richardson, D.C. Richardson, MolProbity: more and better reference data for improved all-atom structure validation, *Protein Sci.* 27 (2018) 293–315, <https://doi.org/10.1002/pro.3330>.
- [57] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr.* 26 (1993) 283–291, <https://doi.org/10.1107/s0021889892009944>.
- [58] D. Eisenberg, R. Lüthy, J.U. Bowie, VERIFY3D: assessment of protein models with three-dimensional profiles, *Methods Enzymol.* 97 (1997) 77022–77028, <https://doi.org/10.1016/S0076-6879>.
- [59] C. Colovos, T.O. Yeates, Verification of protein structures: patterns of nonbonded atomic interactions, *Protein Sci.* 2 (1993) 1511–1519, <https://doi.org/10.1002/PRO.5560020916>.
- [60] T. Liu, R. Altman, Identifying druggable targets by protein microenvironments matching: application to transcription factors, CPT pharmacometrics syst, *Pharmacol.* 3 (2014) 93, <https://doi.org/10.1038/psp.2013.66>.
- [61] S.B. Jamal, S.S. Hassan, S. Tiwari, M.V. Viana, L. De Jesus Benedites, A. Ullah, A. G. Turjanski, D. Barh, P. Ghosh, D.A. Costa, A. Silva, R. Röttger, J. Baumbach, V. A.C. Azevedo, An integrative in-silico approach for therapeutic target identification in the human pathogen *Corynebacterium diphtheriae*, *PLoS One* 12 (2017), e0186401, <https://doi.org/10.1371/journal.pone.0186401>.
- [62] Y. Xu, S. Wang, Q. Hu, S. Gao, X. Ma, W. Zhang, Y. Shen, F. Chen, L. Lai, J. Pei, CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction, *Nucleic Acids Res.* 46 (2018) W374–W379, <https://doi.org/10.1093/nar/gky380>.
- [63] A. Volkamer, D. Kuhn, F. Rippmann, M. Rarey, Dogsitercror: a web server for automatic binding site prediction, analysis and druggability assessment, *Bioinformatics* 28 (2012) 2074–2075, <https://doi.org/10.1093/bioinformatics/bts310>.
- [64] J. Jiménez, S. Doerr, G. Martínez-Rosell, A.S. Rose, G. De Fabritiis, DeepSite: protein-binding site predictor using 3D-convolutional neural networks, *Bioinformatics* 33 (2017) 3036–3042, <https://doi.org/10.1093/bioinformatics/btx350>.
- [65] J. Zhang, Y. Gan, H. Li, J. Yin, X. He, L. Lin, S. Xu, Z. Fang, B. wook Kim, L. Gao, L. Ding, E. Zhang, X. Ma, J. Li, L. Li, Y. Xu, D. Horne, R. Xu, H. Yu, Y. Gu, W. Huang, Inhibition of the CDK2 and Cyclin A complex leads to autophagic degradation of CDK2 in cancer cells, *Nat. Commun.* 13 (2022), <https://doi.org/10.1038/s41467-022-30264-0>.
- [66] C. Skuta, M. Popr, T. Muller, J. Jindrich, M. Kahle, D. Sedlak, D. Svozil, P. Bartunek, Probes & Drugs portal: an interactive, open data resource for chemical biology, *Nat. Methods* 14 (2017) 759–760, <https://doi.org/10.1038/nmeth.4365>.
- [67] RDKit, Open-Source cheminformatics software (RRID: SCR_014274), (n.d.), <https://www.rdkit.org>.
- [68] M.P. Repasky, M. Shelley, R.A. Friesner, Flexible ligand docking with glide, *Curr. Protoc. Bioinforma* (2007), <https://doi.org/10.1002/0471250953.bi0812s18>.

- [69] R.A. Friesner, R.B. Murphy, M.P. Repasky, L.L. Frye, J.R. Greenwood, T. A. Halgren, P.C. Sanschagrin, D.T. Mainz, Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes, *J. Med. Chem.* 49 (2006) 6177–6196, <https://doi.org/10.1021/jm051256o>.
- [70] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX*. 1–2 (2015) 19–25, <https://doi.org/10.1016/j.softx.2015.06.001>.
- [71] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINC: a linear constraint solver for molecular simulations, *J. Comput. Chem.* 18 (1997) 1463–1472, [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H).
- [72] P. Kunzmann, K. Hamacher, Biotite: a unifying open source computational biology framework in Python, *BMC Bioinformatics* 346 (19) (2018), <https://doi.org/10.1186/s12859-018-2367-z>.
- [73] T. Tubiana, J.-C. Carvaillo, Y. Boulard, S. Bressanelli, TTClust: a versatile molecular simulation trajectory clustering program with graphical summaries, *J. Chem. Inf. Model.* 58 (2018) 2178–2182, <https://doi.org/10.1021/acs.jcim.8b00512>.
- [74] C.C. David, D.J. Jacobs, Principal component analysis: a method for determining the essential dynamics of proteins, *Methods Mol. Biol.* (2014), https://doi.org/10.1007/978-1-62703-658-0_11.
- [75] C. Ross, B. Nizami, M. Glenister, O.S. Amamuddy, A.R. Atilgan, C. Atilgan, Ö. T. Bishop, MODE-TASK: large-scale protein motion tools, *Bioinformatics* (2018), <https://doi.org/10.1093/bioinformatics/bty427>.
- [76] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, 2016. <https://ggplot2.tidyverse.org>.
- [77] H. William, D. Andrew, S. Klaus, VMD: visual molecular dynamics, *J. Mol. Graph.* 14 (1996) 33–38.
- [78] S. Zhang, J.M. Krieger, Y. Zhang, C. Kaya, B. Kaynak, K. Mikulska-Ruminska, P. Doruker, H. Li, I. Bahar, ProDy 2.0: increased scale and scope after 10 years of protein dynamics modelling with Python, *Bioinformatics* 37 (2021) 3657–3659, <https://doi.org/10.1093/bioinformatics/btab187>.
- [79] M.S. Valdés-Tresanco, M.E. Valdés-Tresanco, P.A. Valiente, E. Moreno, gmx_MMPBSA: a new tool to perform end-state free energy calculations with GROMACS, *J. Chem. Theor. Comput.* 17 (2021) 6281–6291, <https://doi.org/10.1021/acs.jctc.1c00645>.
- [80] B.R. Miller, T.D. McGee, J.M. Swails, N. Homeyer, H. Gohlke, A.E. Roitberg, MMPBSA.py: an efficient program for end-state free energy calculations, *J. Chem. Theor. Comput.* 8 (2012) 3314–3321, <https://doi.org/10.1021/ct300418h>.
- [81] Upstream, ORFs are prevalent translational repressors in vertebrates | the EMBO Journal, (n.d. <https://www.embopress.org/doi/full/10.15252/embj.201592759>.
- [82] T.C. Vilela Rodrigues, A.K. Jaiswal, A. de Sarom, L. de Castro Oliveira, C.J. Freire Oliveira, P. Ghosh, S. Tiwari, F.M. Miranda, L. de Jesus Benevides, V. Ariston de Carvalho Azevedo, S. de Castro Soares, Reverse vaccinology and subtractive genomics reveal new therapeutic targets against *Mycoplasma pneumoniae*: a causative agent of pneumonia, *R. Soc. Open Sci.* 6 (2019), 190907, <https://doi.org/10.1098/rsos.190907>.
- [83] B. Dehury, S.K. Behera, N. Mahapatra, Structural dynamics of Casein Kinase I (CKI) from malarial parasite *Plasmodium falciparum* (Isolate 3D7): insights from theoretical modelling and molecular simulations, *J. Mol. Graph. Model.* 71 (2017) 154–166, <https://doi.org/10.1016/j.jmglm.2016.11.012>.
- [84] J. Isidro, V. Borges, M. Pinto, D. Sobral, J.D. Santos, A. Nunes, V. Mixão, R. Ferreira, D. Santos, S. Duarte, L. Vieira, M.J. Borrego, S. Nuncio, I.L. de Carvalho, A. Pelerito, R. Cordeiro, J.P. Gomes, Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus, *Nat. Med.* 28 (2022) 1569–1572, <https://doi.org/10.1038/s41591-022-01907-y>.
- [85] C. Happi, I. Adetifa, P. Mbala, R. Njoum, E. Nakoune, A. Happi, N. Ndodo, O. Ayansola, G. Mboowa, T. Bedford, R.A. Neher, C. Roemer, E. Hodcroft, H. Tegally, A. O’Toole, A. Rambaut, O. Pybus, M.U.G. Kraemer, E. Wilkinson, J. Isidro, V. Borges, M. Pinto, J.P. Gomes, L. Freitas, P.C. Resende, R.T.C. Lee, S. Maurer-Stroh, C. Baxter, R. Lessells, A.E. O’Gwell, Y. Kebede, S.K. Tessema, T. de Oliveira, Urgent need for a non-discriminatory and non-stigmatizing nomenclature for monkeypox virus, *PLoS Biol.* 20 (2022), e3001769, <https://doi.org/10.1371/journal.pbio.3001769>.
- [86] M. Kozlov, Monkeypox goes global: why scientists are on alert, *Nature* 606 (2022) 15–16, <https://doi.org/10.1038/d41586-022-01421-8>.
- [87] K.E. Mushegian AR, A minimal gene set for cellular life derived by comparison of complete bacterial genomes, in: *Proc Natl Acad Sci USA*, 1996.
- [88] M. Elhefnawi, O. Alaidi, N. Mohamed, M. Kamar, I. El-Azab, S. Zada, R. Siam, Identification of novel conserved functional motifs across most Influenza A viral strains, *Virology*. J. 8 (2011), <https://doi.org/10.1186/1743-422X-8-44>.
- [89] F. Yan, F. Gao, A systematic strategy for the investigation of vaccines and drugs targeting bacteria, *Comput. Struct. Biotechnol. J.* 18 (2020) 1525–1538, <https://doi.org/10.1016/j.csbj.2020.06.008>.
- [90] S.S. Shiragannavar, A.K. Shettar, S.B. Madagi, S. Sarawad, Subtractive genomics approach in identifying polysaccharide biosynthesis protein as novel drug target against *Eubacterium nodatum*, *Asian J. Pharm. Pharmacol.* 5 (2019) 382–392, <https://doi.org/10.31024/ajpp.2019.5.2.24>.
- [91] A.L. Hopkins, C.R. Groom, The druggable genome, *Nat. Rev. Drug Discov.* 1 (2002) 727–730, <https://doi.org/10.1038/nrd892>.
- [92] C.M. Byrd, T.C. Bolken, D.E. Hrubby, The vaccinia virus I7L gene product is the core protein proteinase, *J. Virol.* 76 (2002) 8973–8976, <https://doi.org/10.1128/JVI.76.17.8973-8976.2002>.
- [93] B. Reed, L. Yakovleva, S. Shuman, R. Ghose, Characterization of DNA binding by the isolated N-terminal domain of vaccinia virus DNA topoisomerase IB, *Biochemistry* 56 (2017) 3307–3317, <https://doi.org/10.1021/acs.biochem.7b00042>.
- [94] P.D. Gershon, B. Moss, Early transcription factor subunits are encoded by vaccinia virus late genes, *Proc. Natl. Acad. Sci.* 87 (1990) 4401–4405, <https://doi.org/10.1073/pnas.87.11.4401>.
- [95] K. Ishii, B. Moss, Role of vaccinia virus A20R protein in DNA replication: construction and characterization of temperature-sensitive mutants, *J. Virol.* 75 (2001) 1656–1663, <https://doi.org/10.1128/JVI.75.4.1656-1663.2001>.
- [96] N. Klemperer, W. McDonald, K. Boyle, B. Unger, P. Traktman, The A20R protein is a stoichiometric component of the processive form of vaccinia virus DNA polymerase, *J. Virol.* 75 (2001) 12298–12307, <https://doi.org/10.1128/JVI.75.24.12298-12307.2001>.
- [97] A.J. Clark, P. Tiwary, K. Borrelli, S. Feng, E.B. Miller, R. Abel, R.A. Friesner, B. J. Berne, Prediction of protein-ligand binding poses via a combination of induced fit docking and metadynamics simulations, *J. Chem. Theor. Comput.* 12 (2016) 2990–2998, <https://doi.org/10.1021/acs.jctc.6b00201>.
- [98] A. Levitzki, S. Klein, My journey from tyrosine phosphorylation inhibitors to targeted immune therapy as strategies to combat cancer, *Proc. Natl. Acad. Sci.* 116 (2019) 11579–11586, <https://doi.org/10.1073/pnas.1816012116>.
- [99] H.M. Ismail, V. Barton, M. Phanchana, S. Charoensuthivarakul, M.H.L. Wong, J. Hemingway, G.A. Biagini, P.M. O’Neill, S.A. Ward, Artemisinin activity-based probes identify multiple molecular targets within the asexual stage of the malaria parasites *Plasmodium falciparum* 3D7, *Proc. Natl. Acad. Sci. U. S. A.* 113 (2016) 2080–2085, <https://doi.org/10.1073/pnas.1600459113>.
- [100] S. Tassini, L. Sun, K. Lanko, E. Crespan, E. Langron, F. Falchi, M. Kissova, J. I. Armijos-Rivera, L. Delang, C. Mirabelli, J. Neyts, M. Pieroni, A. Cavalli, G. Costantino, G. Maga, P. Vergani, P. Leyssen, M. Radi, Discovery of multitarget agents active as broad-spectrum antivirals and correctors of cystic fibrosis transmembrane conductance regulator for associated pulmonary diseases, *J. Med. Chem.* 60 (2017) 1400–1416, <https://doi.org/10.1021/acs.jmedchem.6b01521>.
- [101] C. Ambery, K. Riddell, P. Daley-Yates, Open-label, randomized, 6-way crossover, single-dose study to determine the pharmacokinetics of batefenterol (GSK961081) and fluticasone furoate when administered alone or in combination, *Clin. Pharmacol. Drug Dev.* 5 (2016) 399–407, <https://doi.org/10.1002/cpdd.274>.
- [102] A.D. Hughes, L.H. Jones, Dual-pharmacology muscarinic antagonist and β 2 agonist molecules for the treatment of chronic obstructive pulmonary disease, *Future Med. Chem.* 3 (2011) 1585–1605, <https://doi.org/10.4155/fmc.11.106>.
- [103] G. Setia, N. Hagog, B. Jalilizeinali, S. Funkhouser, L. Pierzchanowski, F. Lan, T. G. Gabig, B. Kiner-Strachan, K. Kelleher, M.-C. Hsu, L.-W. Chang, M.W. Schuster, A Phase II, Open-label pilot study to evaluate the hematopoietic stem cell mobilization of TG-0054 combined with G-CSF in 12 patients with multiple myeloma, non-hodgkin lymphoma or Hodgkin lymphoma - an interim analysis, *Blood* 126 (2015), <https://doi.org/10.1182/blood.V126.23.515.515>, 515–515.
- [104] A.J. Lembo, B.E. Lacy, M.J. Zuckerman, R. Schey, L.S. Dove, D.A. Andrae, J. M. Davenport, G. McIntyre, R. Lopez, L. Turner, P.S. Covington, Eluxadoline for irritable bowel syndrome with diarrhea, *N. Engl. J. Med.* 374 (2016) 242–253, <https://doi.org/10.1056/NEJMoa1505180>.
- [105] C. Crim, M. Gotfried, S. Spangenthal, M. Watkins, A. Emmett, C. Crawford, C. Baidoo, R. Castro-Santamaria, A randomized, controlled, repeat-dose study of batefenterol/fluticasone furoate compared with placebo in the treatment of COPD, *BMC Pulm. Med.* 119 (20) (2020), <https://doi.org/10.1186/s12890-020-1153-7>.
- [106] D.M. Brenner, G.S. Sayuk, C.R. Gutman, E. Jo, S.J.R. Elmes, L.W.C. Liu, B.D. Cash, Efficacy and safety of eluxadoline in patients with irritable bowel syndrome with diarrhea who report inadequate symptom control with loperamide: RELIEF phase 4 study, *Am. J. Gastroenterol.* 114 (2019) 1502–1511, <https://doi.org/10.14309/ajg.0000000000000327>.