Journal of the Egyptian Society of Parasitology, Vol. 55, No. 2, August 2025
J. Egypt. Soc. Parasitol. (JESP), 55(2), 2025: 151 – 160
Online: 2090-2549

IDENTIFICATION OF A UNIQUE CYP450 STRUCTURE IN THE SAUDI MOMOUVIRUS AMOEBOPHAGE USING ARTIFICIAL INTELLIGENCE AND BIOINFORMATICS

By

YASIR MOHAMMED A. ALQURASHI

Department of Medical Laboratory, College of Applied Medical Sciences in Yanbu Governorate, Taibah University, Yanbu, 46522, Saudi Arabia
(Correspondence: yqurashi@taibahu.edu.sa, Mobile: 00966-533230246 ORCID: http:// orcid.org/ 0000-0003-3398-1557)

Abstract

The Saudi moumouvirus amoebophage was detected in 2016 in *Acanthamoeba polyphaga*. Although the whole genome of the virus has been sequenced, most of the virus genes and gene products are not characterized to date. The virus's genome contains a gene encoding a cytochrome P450 enzyme which has been shown to produce confliction annotation and classification. The classification of CYP450 enzymes is based on the homology of the protein sequence while protein characterization is based normally on experimentation. Computational approaches have been shown to be a practical and highly accurate approach for characterizing proteins especially in the context of unusual proteins to overcome the flaws of experimental protein structure. Using artificial intelligence and bioinformatic tools, the current study identified this CYP450 (named CYP503 fused to 20G-fe (II) oxygenase domain) as a unique self-sufficient CYP450 enzyme, a class of biocatalysts renowned for their significant potential in biotechnological applications.

Key words: Saudi Arabia, *Acanthamoeba polyphaga,* Artificial intelligence, Bioinformatics

Introduction

Cytochrome P450 (CYP450) enzymes are classified among a superfamily containing 2250 families of organism from all kingdoms of life (NCBI 2023). CYP450 are classified according to the amino acid sequence identity where enzymes with 40% or higher identity are classified in the same family and enzymes with 55% or higher identity within the family are classified in the same subfamily (Nelson *et al,* 1996). The recently discovered viruses of order, *Imitervirales* or the giant viruses are known to possess unprecedented properties including unusual genes and protein products (Alqurashi, 2024). The amoebophage Saudi moumouvirus (SDMV) of *Acanthamoeba polyphaga* was first isolated in Saudi Arabia in 2016 and its genome was fully sequenced, and virus was classified in the *Mimivirus* genus within *Imitervi rales* order based on the virus genome sequence, a CYP450 predicted enzyme sequence (GenBank accession: AQN67958.1) detected in the R96 gene. Automated annotation classified enzyme as a putative CYP503A1-like protein; Conserved Domain Database entry cd1104 (Bajrai *et al,* 2016). Conversely, Lambe *et al,* (2019) classified vurus as CYP5253A1, noting high sequence identity to the *Acanthamoeba polyphaga* mimivirus (APMV) CYP450 (gene MIMI_L808). Experimental recombinant protein studies to classify the CYP450 gene product based on the substrate range of sterols including lanosterol yielded no activity under multiple conditions. This lack of activity of the enzyme on typical sterol substrate suggests its true classification to be different form initial prediction. Both viral CYP450 enzymes were not experimentally classified to date because of their unusual structure (Lamb *et al,* 2019).

The previous utilised effectively variable artificial intelligence and bioinformatics tools analysed genes (Al Qurashi *et al,* 2025), variants (Al Qurashi *et al,* 2024) and even new species (Alissa Alkhalaf *et al,* 2014). Building on this experience, various computational approaches were employed in this study to resolve the discrepancy in the classification of the viral CYP450.

This study aimed to analyse CYP450 enzyme (named CYP503 fused to 20G-fe (II) oxygenase domain) CYP503A1-like of the Saudi moumouvirus amoebaphage R96 gene to classify it according to its amino acid sequence identity, function and predicted protein structure using computational analytical approach utilizing a suite of artificial intelligence and bioinformatic tools.

Material and Methods

The methods used included artificial intelligence and bioinformatic methods. Artificial intelligence methods included the use of the deep learning and machine learning methods. Deep learning method tool used for protein structure prediction was Alphafold2 in addition to machine learning tools Uniprot, Interpro and ProtNLM. The various bioinformatic tools were utilised this study include Expasy-ProtParam protein identification and analysis tool, NCBI SPARCLE protein family search tool, CD and Batch CD-search on conserved domain database search (CDD) and phylogenetic analysis tools (Fig. 1).

Data access: Amino acid sequences were retrieved from National Centre of Biotechnology Information (NCBI) protein database includes cytochrome P450 fused to 2OG-f (II) oxygenase domain (Saudi moumouvirus; accession: AQN67958.1), cytochrome P450 [*Postia placenta*] (accession: BAK09443.1) & lanosterol 14-alpha-dem-ethylase [*Acanthamoeba polyphaga* mimivirus] (accession: AAV51068.1 updated Uni-ProtKB/Swiss-Prot accession: Q5UQI3.1). The three-dimensional structure (3D) structure of cytochrome P450 from fungus *Postia placenta* (accession: F1SY99) was retrieved from Alphafold2 protein database and used as a representative of conserved protein domain family compared with SDMV CYP450 enzyme.

Protein sequence analysis: Expasy-ProtParam protein analysis tool was accessed on the Expert Protein Analysis System server. It was used to determine the protein physical and chemical characteristics based on amino acid sequence of cytochrome P450 fused to 2OG-fe(II) oxygenase domain [Saudi moumouvirus; accession: AQN67958.1] (Gasteiger *et al*, 2007).

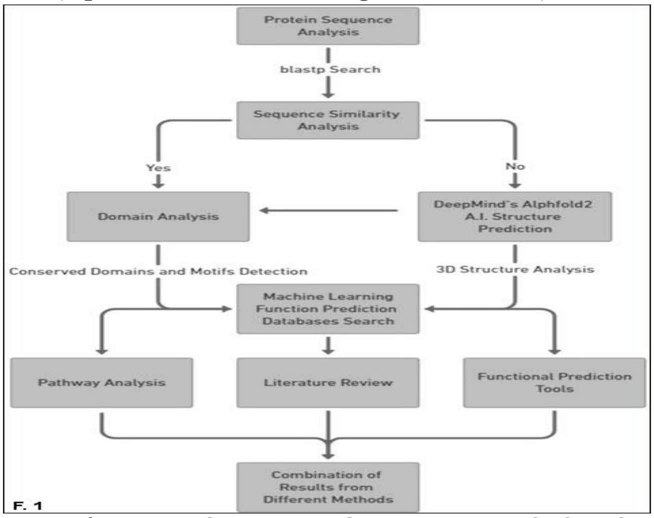

Fig. 1: Flowchart to identify architecture of CYP450 Architecture in Saudi moumouvirus amoebophage showed flow and interaction of data.

Sequence similarity analysis: Identification of similar protein sequences to SDMV CYP450 was conducted by using the Basic Local Alignment Search Tool (BLAST) BLASTp suite. First, the tool was used to search the non-redundant protein sequences database (nr) based on the BLOSUM62 Scoring matrix with scoring parameters including existence and extension gap costs of 11 & 1 respectively and conditional compositional score matrix adjustment (Altschul *et al*, 1997). Similar sequences were analysed using the neighbour joining method of the tool (Altschul *et al*, 2005). Phylogenetic classification among similar sequences in the tree was retrieved from the NCBI taxonomy browser and tree visualisation was done by using iTol tool (Letunic and Bork, 2019).

Conserved domain analysis: Similar protein domains and architecture search and data collection were done using CD, and Batch CD-search on the conserved domain database search too (CDD). All 105 sequences with similarity on conserved protein domain family CYP503A1_like (entry: cd 11041) were retrieved. Multiple sequence alignment was conducted using the NCBI Constrain-based Multiple Alignment Tool (COBALT). The tool generated pairwise constraints firmed on homology in the conserved domains and motifs databases in addition to similarity inferred using RPS-BLAST, PHI-BLAST and BLASTP (Job ID: CD sequences and AQN67958 cobalt-Cobalt RID P26T1TG0 212 or 106 sequences (Papadopoulos and Agarwala, 2007). Multiple sequence alignment was performed using Clustal Omega v. 1.2.2 (Sievers and Higgins, 2018). Phylogenetic trees were constructed using the NCBI Genome Work Bench package, of gbench, v3.9.1 (Kuznetsov and Bollin 2021). Further visualisation, processing and phylogenetic trees editing were done using the iTOL tool (Letunic and Bork, 2019). InterProScan tool provided by Inter-Prot centurium was used to define protein family membership & variable domain architectures (Blum *et al,* 2020)

Three-dimensional structure prediction: The R96 gene deduced amino acid sequence of the SDMV CYP503 fused to 20G-fe (II) oxygenase domain (accession: AQN67958. 1) was uploaded to Alphafold2 to predict the 3D structure (Jumper *et al,* 2021). Relaxed run was done using web browser interface via the GPU node of the Google Collaborators platform ColabFold v1.5.3. Plat-form uses the MMseqs2 (Many-against-Many-sequence searching) sequences search & clustering software suite (Mirdita *et al,* 2022). Protein structure visualization was carried out using PyMol™ molecular visualization system, v.2.5.2 Schrödinger, LLC. NetSurf-P-3.0 webserver tool was used to analyse the secondary protein structure (Høie *et al,* 2022). DeepTMHMM v1.0.24 tool by the Technical University of Denmark was used to predict the transmembrane protein and membrane topology of the analysed enzyme (Hallgren *et al,* 2022).

Machine learning functional prediction & metabolic pathway database search: Functional annotation and metabolic pathway analysis were performed using database integrating machine-learning-based predictions. Databases used were the UniProt (UniProt Consortium, 2022) and InterPro (Paysan-Lafosse *et al*, 2022). Also, the AI-driven annotational tool, ProtNLM via the Neuro-snap interface was used to predict the enzyme's function using a natural language processing models (Gane *et al*, 2022).

Results

Protein sequence analysis: Saudi moumouviurs CYP450 enzyme physical and chemical calculated parameters included a molecular formula of the protein as $C_{3698}H_{5730}N_{942}O_{1055}S_{37}$ from an 11462 total number of atoms. The 703 amino acid enzyme calculated molecular weight was 81.45 kDa and isoelectric point (pI) of 6.06. The total number of negatively and positively charged amino acids of the protein is 95 and 86 respectively. The aliphatic index is 89.8 units indicating a significant stability and grand average of hydropathicity of -0.22 units suggested slight overall hydrophilicity.

Sequence similarity analysis: The BLASTp analysis showed that close homologs of the SDMV CYP450 are extremely rare outside giant viruses. Top hundred BLASTp similarity sequences are summarised in where similarity on whole length of SDMV CYP450 was observed only with CYP450 sequences from the *Mimiviridae* virus family. Phylogenetic tree of SDMV CYP450 and similar sequences detected with balstp search (Fig. 2) showed similarity to CYP450 sequences from all domains and kingdoms of life. Significantly similar sequences with sequence coverage exceeding 95% were all from *Mimiviridae* family except for two organisms from phylum Arthropoda, namely *Folsomia candida* (accession MW004169.1 & XP_ 02 1953323), and *Dinothrombium tinctorium* (accession RWS0 11 98.1). The calculated similarity was 45.32 & 42.28% on 98 & 99% coverage with the putative lanosterol 14-alpha demethylase CYP 450 of *F. can-*

dida & hypothetical unknown CYP450 of *D. tinctorium* respectively.

The highest calculated similarity to SDMV CYP450 was detected with CYP450 fused to 2OG-Fe(II) oxygenase domain in *A. polyph aga* moumouvirus (accession YP_0073540 63.1). Calculated similarity between sequences was 99.57% on 100% coverage with 4 amino acid variations between the two sequences confirmed that the enzyme of both viruses shares the same lineage. The only other similarity calculated exceeding 99% was detected with a putative CYP450 from Moum ouvirus Monve. Putative CYP450 sequence was annotated automatically as CYP503A1-like (accession AEX63235.1) and showed three amino acid variations from the query sequence, but the multiple sequence alignment between this Moumouvirus Monve sequence and the query sequence showed a variation of 23 amino acid stretches on the N-terminal of Moumouvirus Monve, which was not taking into count by the matrix. This close sequence relationship was reflected in the phylogenetic tree as these two sequences were the only ones shared final branching on tree with the query sequence of Saudi moumouvirus in a node of the tree (Fig. 3).

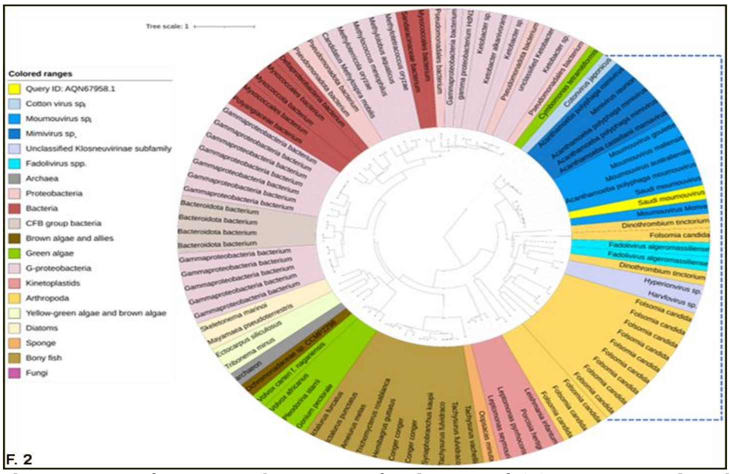

Fig. 2: Neighbor joining phylogenetic tree of SMV cytochrome P450 fused to 2OG-fe(II) oxygenase and top hundred BLASTp similarity sequences.

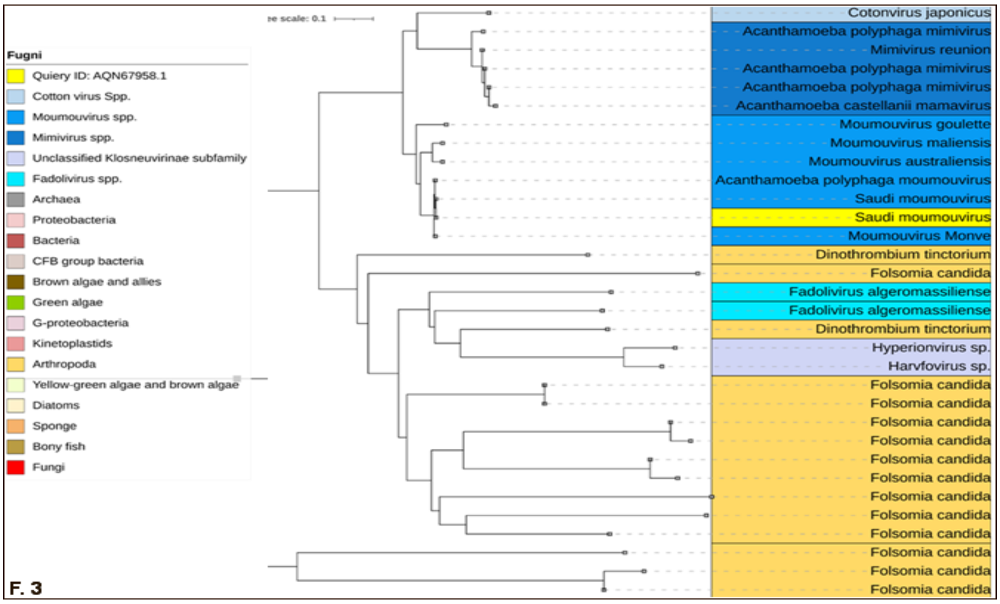

Fig. 3: Neighbor joining phylogenetic subtree of SMV cytochrome P450 fused to 2OG-fe(II) oxygenase and BLASTp similarity sequences showed sequences with significant similarity exceeding 95% amino acid similarity.

Domain sequence analysis: Conserved domain database homology (CDD) search for the amino acid sequence of the Saudi moumouviurs CYP503 fused to 20G-fe (II) oxygenase domain (accession AQN67958.1) indicated significant homology with CYP503 A1-like (entrty:cd11041) and cytochrome CYP450 superfamily in the CDD. The CDD entry: cd11041 (CYP503A1-like) contains members of subfamily CYP503 A and related enzymes containing the polypeptide A in the CYP450 superfamily. The cd11041 sequence cluster contains 105 sequences of the subfamily A and related sequences which all did not cluster on the phylogenetic tree except for the sequences of the *Acanthamoeba polyphaga* CYP450 Fig. 4 (Genbank accession: AAV51068.1, UniPortyKB/Swiss-Prot accession: Q5UQI3.1). The sequence of CYP450 of SDMV and *A. polyphaga mimivirus* is seen branching from a node containing CYP450 from *Pyricularia oryzae* 70-15 (Genbank accession: EAQ71726.1) and Postia placenta Mad-698-R (GenBank accession: BAK09443.1, (Alphafold accession: Fq SY99.CY110_POSM) (Fig. 5). Among the three similar proteins, *Postia placenta* Mad-698-R is the only sequence with three-dimensional structure reviewed by UniProtKB in all protein databases. The pairwise alignment of the SDMV CYP450 of CYP450 from *Pyricularia oryzae* displayed indicated 22.72% similarity to the sequences on 62% coverage of the Saudi moumouvirus sequence with an E-value of 8.0X10-17 and similarity between the Saudi moumouviurs CYP503 first domain after (AA:1-477) and CYP450 from *Pyricularia oryzae* indicated 23.04% similarity between sequences on 90% of the SDMV CYP450 with an E-value of 4.0X10-17 by using the NCBI pairwise alignment in BLASTp suite.

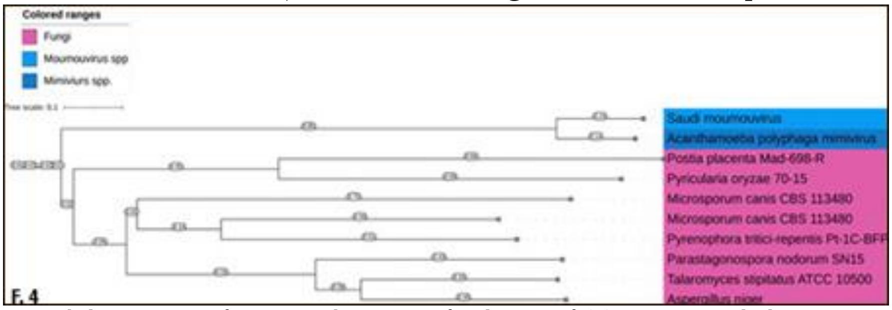

Fig. 4: Neighbor joining phylogenetic tree of SMV cytochrome P450 fused to 2OG-fe(II) oxygenase and cd11041 sequence cluster contains 105 sequences of subfamily A.

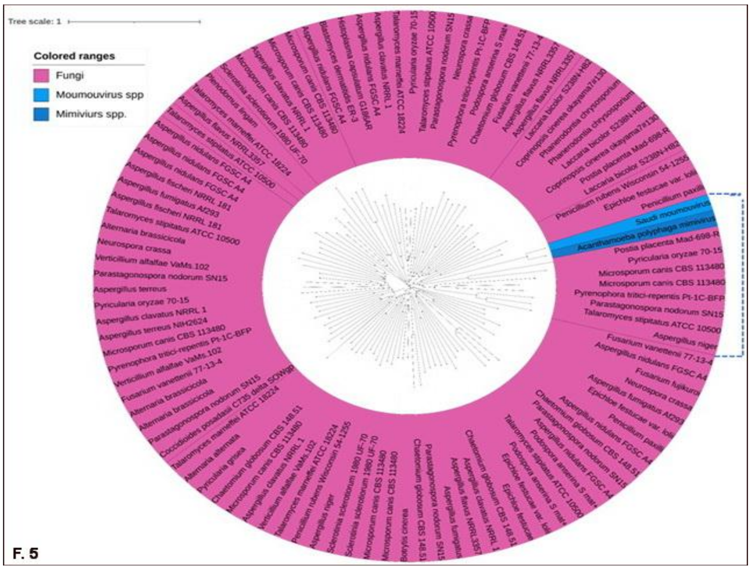

Fig. 5: Neighbor joining phylogenetic subtree of SMV cytochrome P450 fused to 2OG-fe(II) clustering away from all sequences with highest similarity to domain A of Posita placenta Mad-698-R CYP450.

Three-dimensional structure prediction: AlphaFold2 produced a high confidence model of the SSMV CYP450. The predicted structure revealed a monomer of two well-predicted domains, domain A (residues: 1-476) and domain B (residues: 477-703) with high confidence levels indicating a highly reliable prediction varied between 80 and >90% pLLDT as detected on the expected position error plot (Fig. 6). The high confidence score for each domain was illustrated in superimposition of the two CYP450 structures and the corresponding Ramachandran plot (Fig. 7). Both showed a very low homology level in the transmembrane region and beyond boundaries of each domain indicated that this positioning was rare in the existing proteins database whole predicted structure compared to *P. placenta* CYP450 monooxygenase 110 (Fig. 8).

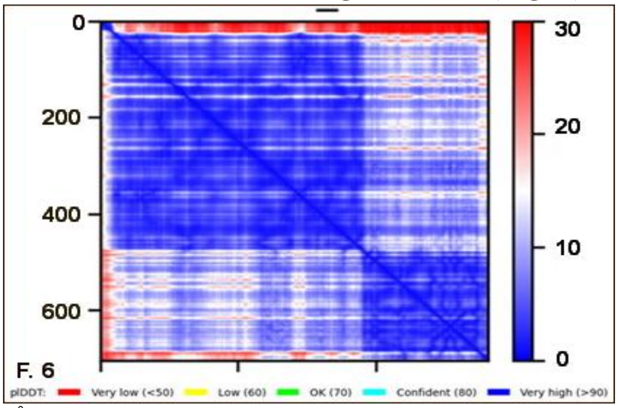

Fig. 6: Expected position error (Ångströms) of Saudi moumouvirus cytochrome P450 fused to 2OG-fe(II) oxygenase domain.

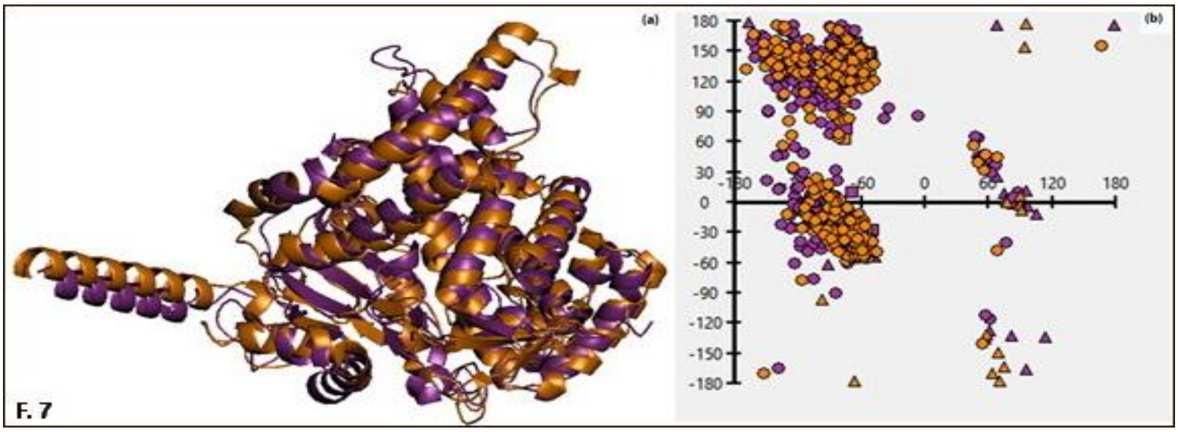

Fig. 7: Saudi moumouvirus fused to 2OG-fe(II) oxygenase domain A (residues: 1-476) and CYP450 and Postia placenta CYP450 monooxygenase 110 domain structural homology. (a) Superimposition image of the Saudi moumouvirus CYO450 in cyan and Postia placenta CYP450 in orange. (b) Ramachandran plot: Corresponding site of each amino acid represented by x-axis & Ψ angle on y-axis. ▲Glycine, ■ Proline and ● Other residues

Domain A (residues: 1-476): Low confidence levels were detected in the transmembrane helix toward the linker and N-terminus predicted (Fig. 6) and verified using the hydropathy analysis of the sequence (Fig. 7). Inferred from homology with domain A of the Postia placenta Mad-698-R and overall topology of the predicted structure, the domain consists of three regions. First region is a transmembrane region of a single helix (helix A). Transmembrane region prediction was verified using the hydropathy analysis of the sequence (Fig. 8) and compared with homologous protein (Fig. 9). Analysis displayed the position of the expected region (residues: 1-19) and indicated the rest posi tion of the domain to the internal side of the membrane. Transmembrane region thickness was much smaller than *Postia placenta* Mad-698-R protein indicated the host smaller membrane thickness. Blot showed three internal hydropathic stretches implying the interaction of this domain B (Fig. 8). Second one is an auxiliary redox partner showed a distinct ferredoxin-like fold and the third one is P450 haem reg- ion towards the C-terminus. Both regions interact with each other, but the distinct triangular shape of

CYP450 was clear on the front view, and the ferredoxin was more distinguishable on the domain back view (Fig. 9), and three regions were in predicted protein structure (Fig. 10).

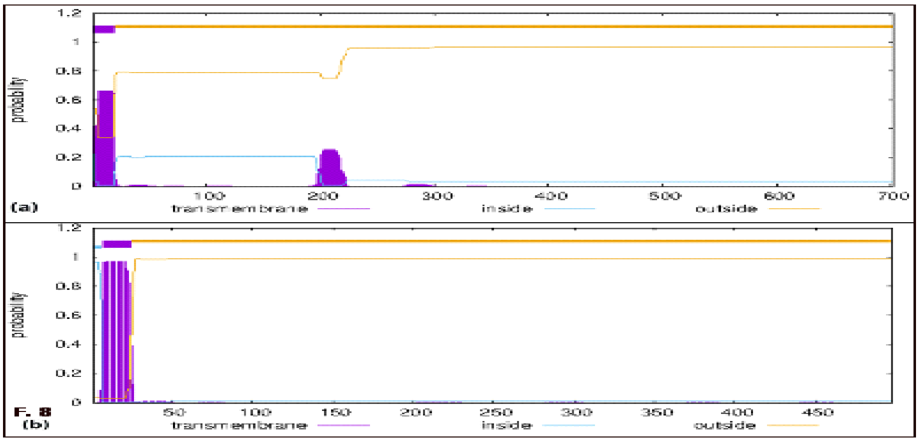

Fig. 8: Hydropathy plot of Saudi moumouvirus cytochrome P450 fused to 2OG-fe(II) (a)& Postia placenta Mad-698-R cytochrome P450 (b).

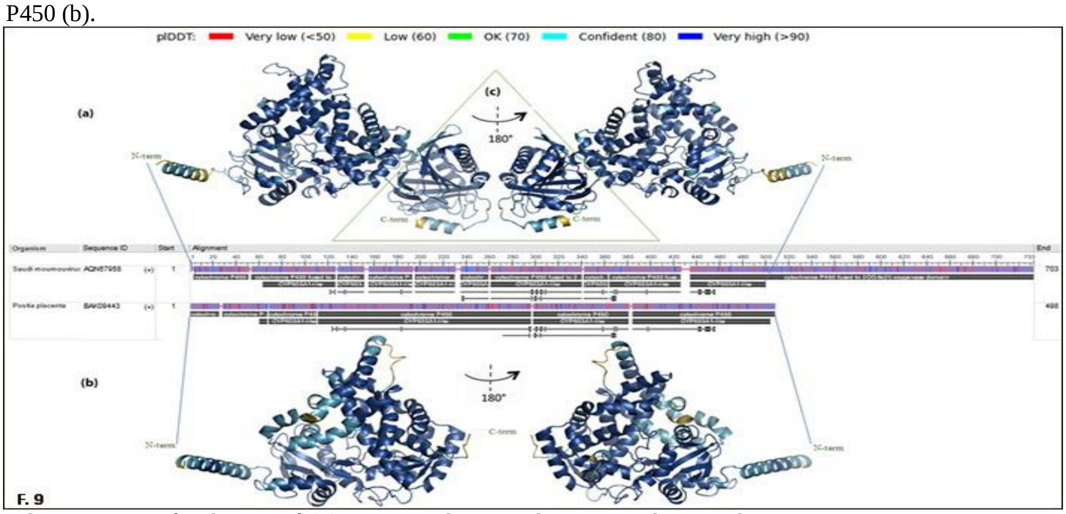

Fig. 9: Saudi moumouvirus fused to 2OG-fe(II) oxygenase domain and CYP450 and *Postia placenta* CYP450 monooxygenase 110 pairwise alignment and protein structures front and back views coloured in pLLDT confidence level colours . (a) Predicted structure of Saudi moumouvirus. (b) *Postia placenta* CYP450 monooxygenase 110. (c) 2OG-fe(II) oxygenase domain shown within green triangle.

Domain B (residues: 477-703): This domain showed a core of four antiparallel β-strands flanked by short α-helices oxidoreductase 2OG-Fe(II) (Fig.10). This topology is emblematic of FMN-binding redox partner domains in self-sufficient P450s (Eser *et al*, 2021) rather than an Fe(II)/2OG-oxygenase fold. The domain orientation relative to the other redox partners is not expected but, functional enzymatic studies considerable domain repositioning might take place as a result of substrate binding refolding (Kitazu me *et al*, 2007; Lamb and Waterman 2013; Zhang *et al*, 2018; Blum *et al*, 2020).

Machine learning functional prediction and metabolic pathway database search: Both UniProt and InterPro analysis confirmed the enzyme identity as a cytochrome P450 enzyme, but did not assign any specific pathway for it. UniProt identified significant similarity to entry L7RBR2 of the cytochrome P450 fused to 2OG-fe(II) oxygenase from *A. polyphaga* moumouvirus with 99.6% similarity and indicated oxidoreductase activity. Interpro identified the enzyme sequence as a CYP450 based on haem presence bin-ding site (GO:0020037) and iron ion binding site: (GO:00055506). Based on predicted structure, it predicted oxidoreductase activity on a paired donor (GO: 0016705) & monooxygenase activity (GO: 0004497). Pfam database didn't detect any similarity to both virus domains, but detected homology to domain A of CYP503 A1-like & 20G-fe (II) oxygenase domain. ProtNLM's AI-based annotation indicated a cytochrome P450 of highest probability for enzyme sequence with lanosterol 14-alpha demethylase. A lower probability of CYPP450 monooxygenase also indicated and suggested potential oxidoreductase activity and steroid metabolism though a precise substrate could not be determined.

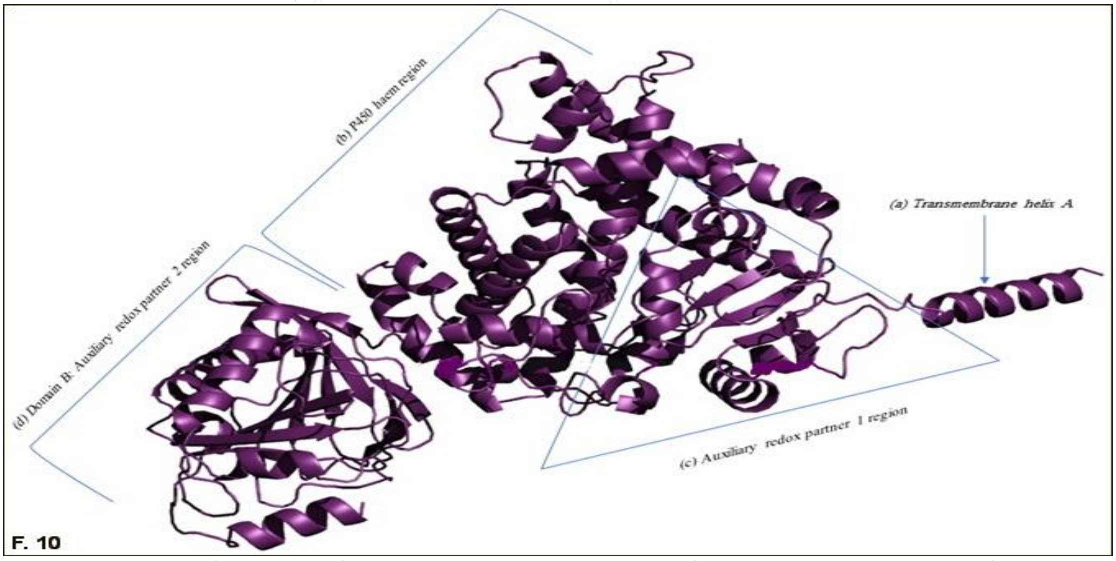

Fig. 10: Saudi moumouvirus fused to 2OG-fe(II) oxygenase domain main regions of predicted structure. First region from N terminus on right (b) P450 haem region with a transmembrane helix (a), (c) auxiliary redox partner 1 () & on (d) cytochrome P450 reductase domain region.

Discussion

The characteristic of the SDMV CYP450 and those of other members of the Mimiviruse genus are unique among all CYP450 enzymes identified by both sequence and functional analysis. The molecular weight, atomic composition and amino acids profile of SDMV CYP450 are of a higher value compared to other members of the annotated enzyme family members. The half-life and the instability index indicate a relative stability of the enzyme ruling out any errors in the sequencing and consequently the structure prediction of the enzyme. These also align the predicted two-domain structure results obtained by Alphafold2.

Using the BLASTp search to classify the SDMV CYP450 based on sequence homology has not been successful as none of the highly similar sequences of Acanthamoeba polyphaga moumouvirus and Moumouvirus Monve have been experimentally analysed. However, this analysis has pointed to an evident relationship with the CYP450 of *Folsomia candida* and *Dinothrombium tinctorium* based on the high similarity on the whole amino acid sequence of the CYP450 which was not seen with any CYP450 out of the Viruses Super kingdom. The similarity also included variable sequences from other *Folsomia candida* CYP450 with variable calculated values of similarity but the CYP450 was clearly the closest to the Query CYP450 from Saudi moumouvirus.

Conserved domain analysis of the enzyme structure clarifies the reason for the annotation of the enzyme as CYP51 (family 503, subfamily A, polypeptide 1 and similar cytochrome P450). The results show that this annotation was based on the homology with *A. polyphaga* mimivirus CYP450 (Bajrai *et al*, 2016). But, the present phylogenetic analysis indicated inaccuracy in this annotation based on the high distance seen in the phylogenetically analysed SDMV enzyme from all other members of the family. This result was supported by the high-confidence three-dimensional enzyme structure prediction by AlphaFold2 as the SDMV CYP450 enzyme has a two-domain structure compared to a single domain structure seen in all other members of this family. The three-dimensional artificial intelligence predicted the enzyme's structure to be a monomer of two domains with a high confidence level. It also displays the three components of the enzyme which are the haem, ferredoxin and ferredoxin reductase. This structure is characteristic to self-sufficiency CYP450 enzymes (Eser *et al,* 2021). Also, the machine learning functional prediction and metabolic pathway database search support the annotation of the SDMV as a CYP450 and reinforce its self-sufficiency structure.

Conclusion

This study showed that CYP450 of Saudi Moumouvirus amoebophage (CYP503 fused to 20G-fe (II) oxygenase domain) encoded by the R96 gene showed a unique structure of the enzyme among currently recognized CYP450. Uniqueness of CYP450 structure elucidated the current discrepancies in the enzyme's classification.

Also, the study characterized the enzyme among the self-sufficient CYP450 enzymes, which hold considerable promise in the field of biotechnology especially in therapeutic biotechnology because of their effectiveness in biological conditions and their potential for biomolecular redesigning. Besides, the study focused on the enzyme structure, further computational and functional studies clarified the range of substrates of the enzyme, elucidate their associated metabolic pathways and characterise the electron flow in the system.

Further comparative genomic analysis of the viral enzyme gene in relationship to other organisms including *Folsomia candida* may also deepens the knowledge of CYP450 enzyme evolution and fill knowledge gap in giant virus's evolution history in general.

Author declaration: Author neither has any conflict of interest nor received any funds.

References

Al Qurashi, Y, Alansari, M, Almutiri, R*, et al,* 2025: Molecular epidemiology of Omicron's CH.1.1 lineage: evidence of recombination event. Rev. Res. Med. Microbiol. 36(1).

Al Qurashi, YMA, Abdulhakim, JA, Alkhalil, SS*, et al,* 2024: Molecular epidemiology of omicron CH.1.1 Lineage: Genomic and phenotypic data perspective. Cureus 16, 2:e53496.

Alissa Alkhalaf, M, Al Qurashi, YMA, Guiver, M*, et al,* 2014: Genome sequences of three species d adenoviruses isolated from AIDS patients. Genome Announcements 2, 1:e01267-13.

Alqurashi, YMA 2024: Systematic Review of the effects of *Acanthmoeba polyphaga* mimivirus (APMV) infection on virulence of *Acanthamoeba polyphaga* (AP). JESP 54, 3:369-80.

Altschul, SF, Madden, TL, Schäffer, AA*, et al,* 1997: Gapped BLAST and PSI-BLAST: A new gene-ration of protein database search programs. Nucl. Acids Res. 25,17:3389-402.

Altschul, SF, Wootton, JC, Gertz, EM*, et al,* 2005: Protein database searches using compositionally adjusted substitution matrices. Febs J. 272, 20:5101-9.

Bajrai, LH, de Assis, FL, Azhar, EI*, et al,* 2016: Saudi Moumouvirus, the first group B mimivirus isolated from Asia. Front. Microbiol. 7: 2029.

Blum, M, Chang, H-Y, Chuguransky, S*, et al,* 2020: The InterPro protein families and domains data-base: 20 years on. Nucl. Acids Res. 49, D1: D344-54.

UniProt Consortium, 2024 UniProt: The universal protein knowledgebase in 2023. Nucl. Acids Res. 51, D1:D523-31.

Eser, BE, Zhang, Y, Zong, L*, et al,* 2021: Self-sufficient Cytochrome P450s and their potential applications in biotechnology. Chinese J. Chem. Eng. 30:121-35.

Gane, A, Bileschi, M, Dohan, D*, et al,* 2022: ProtNLM: model-based natural language protein annotation. Available from: https://storage.googleapis.com/brain-genomics-public/research/proteins/protnlm/uniprot_2022_04/protnlm_preprint_draft.pdf.

Gasteiger, E, Hoogland, C, Gattiker, A*, et al,* 2007: Protein identification and analysis tool on the ExPASy Server. 112:571-607.

Hallgren, J, Tsirigos, KD, Pedersen, MD*, et al,* 2022: DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv:04.08.487609.

Høie, MH, Kiehl, EN, Petersen, B*, et al,* 2022: NetSurfP-3.0: accurate and fast prediction of protein structural features by protein language models and deep learning. Nucl. Acids Res. 50. W1:W510-5.

Jumper, J, Evans, R, Pritzel, A*, et al,* 2021: Highly accurate protein structure prediction with AlphaFold. Nature 596, 7873:583-9.

Kitazume, T, Haines, DC, Estabrook, RW*, et al,* 2007: Obligatory intermolecular electrontransfer from FAD to FMN in dimeric P450BM-3. Biochemistry 46, 42:11892-901.

Kuznetsov, A, Bollin, CJ 2021: NCBI Genome Workbench: Desktop Software for Comparative Genomics, Visualization, and GenBank Data Submission. Meth. Mol. Biol. 2231:261-95.

Lamb, D, Waterman, M 2013: Unusual properties of the cytochrome P450 superfamily. Philosophical transactions of the Royal Society of London. Series B, Biol. Sci. 368: 20120434.

Lamb, DC, Follmer, AH, Goldstone, JV*, et al,* 2019: On the occurrence of cytochrome P450 in viruses. Proc. Natl. Acad. Sci. USA 116, 25: 12343-52.

Letunic, I, Bork, P 2019: Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. Nucl. Acids Res. 47, W1:W256-9.

Mirdita, M, Schütze, K, Moriwaki, Y*, et al,* 2022: ColabFold-making protein folding accessible to all. BioRxiv:2021.08.15.456425.

NCBI 2023: National Center for Biotechnology InformationI, Conserved Protein Domain Family cytochrome_P450. Available from: https:// www.ncbi.nlm.nih.gov/Structure/ cdd/cdd. shtml [Accessed 28 Apr 2025].

Papadopoulos, JS, Agarwala, R, 2007: COBALT: Constraint-based alignment tool for multiple protein sequences. Bioinformatics 23, 9: 1073-9.

Paysan-Lafosse, T, Blum, M, Chuguransky, S, *et al,* 2022: InterPro in 2022. Nucl. Acids Res. 51, D1:D418-27.

Sievers, F, Higgins, DG 2018: Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. 27, 1:135-45.

Zhang, H, Yokom, AL, Cheng, S*, et al,* 2018: The full-length cytochrome P450 enzyme CYP-102A1 dimerizes at its reductase domains and has flexible heme domains for efficient catalysis. J. Biol. Chem. 293, 20:7727-36.