



OPEN

Substantiation of propitious “Enzybiotic” from two novel bacteriophages isolated from a wastewater treatment plant in Qatar

Ramya Ramadoss¹, Moza Al-Shukri¹, Basem Shomar², Valentin Alekseevich Ilyin³ & Annette Shoba Vincent¹✉

Lysin of bacteriophages isolated from a particular ecosystem could be inducted as a bio-controlling tool against the inhabiting pathogenic bacterial strains. Our study aims at both experimental and computational characterization of the identical lysin gene product inherent in the genomes of two novel *Myoviridae* bacteriophages, *Escherichia* Phage C600M2 (GenBank accession number OK040807, Protein ID: UCJ01465) and *Escherichia* Phage CL1 (GenBank Genome accession number OK040806.1, Protein ID: UCJ01321) isolated from wastewater collected from the main water treatment plant in Qatar. The lysin protein, evinced to be a globular *N*-acetyl-muramidase with intrinsic “cd00737: endolysin_autolysin” domain, was further expressed and purified to be experimentally validated by turbidimetric assay for its utility as an anti-bacterial agent. Comprehensive computational analysis revealed that the scrutinized lysin protein shared 85–98% sequence identity with 61 bacteriophages, all native to wastewater allied environments. Despite varied Host Recognition Components encoded in their genomes, the similitude of lysins, suggests its apparent significance in host–pathogen interactions endemic to wastewater environment. The present study substantiates the identical lysin from *Escherichia* Phage C600M2 and *Escherichia* Phage CL1 as propitious “enzybiotic”, a hybrid term to describe enzymes analogous to anti-biotics to combat antibiotic-resistant bacteria by *in silico* analysis and subsequent experimental validation.

The issue of water scarcity around the world is prevalent today and its sustainability and security are especially critical in middle eastern countries such as Qatar. With a population of more than 2.68 million as of December 2021, Qatar’s water consumption is one of the highest in the world per capita at 500 L per day¹. Water available for use originates from sources such as abstraction of fresh and saline groundwater, seawater desalination and re-use of treated sewage effluent, with seawater desalination accounting for 99% of Qatar’s drinking water supply. As a result, Qatar has several water sustainability programs carried out by water authority: Kahramaa, among its top priorities. These programs are majorly focused on finding out major and minor chemical and microbial contaminants in drinking water. Wastewater treatment plants have been set up by the authorities to explore alternatives to desalination of seawater and abstraction of Qatar’s limited fresh groundwater resources. As of 2015, 34% of treated wastewater was used for agriculture irrigation and 16% for green space irrigation.

Despite of the continuous efforts, assuring adequate water quality still remains challenging². Several recent research publications suggest that wastewater treatment plants worldwide confirm prevalence of bacterial communities with antibiotic resistance, irrespective of operational efficiency^{3–6}. On the bright side, the same environment facilitates rapid discovery of bacteriophages predacious towards the native bacteria⁷ and could be enlisted as potent as bio-control⁸ for sustainable wastewater treatment. For instance, two novel *E. coli* bacteriophages isolated from the Zayandehrood River in Iran were used in wastewater treatment processes⁹.

¹Biological Sciences, Carnegie Mellon University Qatar, PO box 24866, Doha, Qatar. ²Environmental Science Center (ESC), Qatar University, PO box 2713, Doha, Qatar. ³Computational Biology, Carnegie Mellon University Qatar, PO box 24866, Doha, Qatar. ✉email: annettev@andrew.cmu.edu

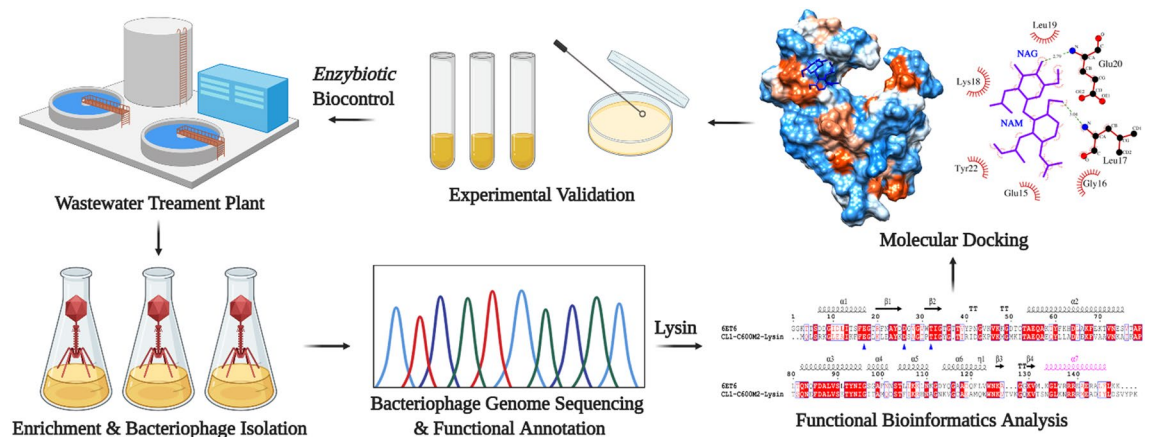


Figure 1. Summary of the research study. Wastewater samples were drawn from main wastewater treatment plant in Qatar, followed by enrichment and isolation of inherent bacteriophages, whole genome sequencing using Ion Torrent S5 next-generation sequencing (NGS) platform and functional annotation of the assembled bacteriophage genome. Gene products annotated as lysins were further analyzed for their potentiality as “enzybiotics” by in silico and experimental studies. This schematic representation was created with BioRender.com.

These bacteriophages isolated from wastewater could also serve as a reservoir of lytic enzymes with the capability of anti-biotics, coined as “enzybiotics”¹⁰ to control drug-resistant pathogenic bacteria¹¹. The lytic enzymes target pathogenic bacteria while leaving commensal microflora unaffected¹² and have proven to be effective in various clinical applications as alternatives to replace antibiotics¹³. Moreover, lysins are highly stable and could be produced in large scale¹⁴ and do not spread Anti-biotic resistance genes through horizontal transfer, thus making them more reliable than bacteriophages for biocontrol and pharmaceutical applications¹³. For instance, endolysins isolated from phages predacious to contaminating drug-resistant pathogens have been proved as efficient “enzybiotic” agents for water treatment in aquacultures^{15,16}.

Lysins induce rapid lysis of the bacterial peptidoglycan cell wall layer and initiate or culminate the infection cycle by either being part of the phage-tail, enacting localized degradation termed as the “lysis from without” phenomenon¹⁷, for phage-DNA entry or as endolysins, facilitating bacterial cell wall lysis mediated by holin for progeny release¹⁸ termed as the “lysis from within” phenomenon. Structurally, lysins are categorized into either globular or modular. Most lysins earmarked for Gram-negative bacteria are globular, comprising a single Enzymatically Active Domain (EAD)^{14,19} while some are modular with two domains—the N-terminal EAD and the C-terminal Cell-wall Binding Domain (CBD)²⁰. Though the EAD of lysins are highly conserved and categorized into muramidase, glucosaminidase, endopeptidase and L-alanine amidase, the CBD is variable and facilitates specific binding to bacteria cell wall²¹.

Enzymes like lysins are catalytic proteins whose efficiency and biological activity are determined by their interaction with induced binding partners. Ideally, the molecular interactions of a Protein–Ligand duo are studied by arduous and expensive X-ray Crystallography or Nuclear Magnetic Resonance (NMR) methodology and the binding sites are pinpointed by techniques like mass-spectrometry. Alternatively, the interactions could be studied by computational modelling of Protein–Ligand complex through molecular docking studies to identify the binding-sites and estimate binding affinity by accurate scoring functions²². A notable study by Kemege et al.²³ proved I-TASSER²⁴ to be one of the most reliable structure prediction tools to accurately model three-Dimensional structure akin to high-resolution X-ray Crystallography and molecular docking studies well postulates the correlation of enzyme-ligand docking interactions with experimental bioactivity²⁵.

As summarized in Fig. 1, in this study, we propose the identical lysin integral to two novel bacteriophages isolated from wastewater²⁶ as propitious “enzybiotic” with anti-bacterial potential by in silico analysis including molecular docking followed by experimental validation using turbidimetric reduction assay which may mark a new approach in the area of modern environmental biotechnology.

Materials and methods

Analysis of CL1-C600M2-lysin sequence. Lysin protein sequences derived from the genomes of *Escherichia* Phage C600M2 (Protein ID: UCJ01465) and *Escherichia* Phage CL1 (Protein ID: UCJ01321) were identical (Supplementary Fig. S1). Both the lysin protein sequences would hence be collectively termed as CL1-C600M2-lysin hereafter. Domain Analysis of CL1-C600M2-lysin was done using Conserved Domain Database (CDD)²⁷. Further, the protein sequence of CL1-C600M2-lysin was subjected to protein-blast (accessed on 13 October 2021) against Refseq_Protein database to identify homologous proteins. Protein sequences sharing sequence identity >70% and low E-value are 90% probable to share functional similarity²⁸. Hence, such phage lysin hits sharing 100% query coverage were chosen, resulting in 61 protein entries with identity ranging from 85% to 98% (Supplementary Table 1). The accommodating environment of these lysin embodying phages were obtained from their respective GenBank records and publications.

Phylogenetic tree of closely related phage lysin sequences. The retrieved lysin protein sequences from “[Analysis of CL1-C600M2-lysin sequence](#)” were aligned with CL1-C600M2-lysin using MUSCLE Alignment tool²⁹. The best-fit protein substitution model JTT + G4 was estimated based on the Bayesian Information Criterion (BIC) using ModelTest-NG³⁰. Then Maximum Likelihood (ML) Phylogenetic tree was generated using IQ-TREE³¹ by first creating 1000 ultrafast bootstraps (UFBoot)³² to minimize overestimation of bootstrap support (-bnni) and minimum correlation coefficient (-bcor) for UFBoot convergence criterion. Secondly, Shimodaira-Hasegawa like approximate likelihood ratio test (SH-aLRT) with 1000 replicates was performed on the consensus tree derived from the previous run. The Standard bootstrap support (SBS) values for the ML analysis was estimated by concatenation of the generated bootstrap trees after 100 iterations with same alignment and substitution model mentioned above. A consensus tree using the original MUSCLE alignment input file was created. The support values UF/SH-aLRT/SBS were mapped to the ML tree and further annotated using Interactive Tree Of Life (iTOL) online tool³³.

Comparative analysis of phage proteome. NCBI Batch Entrez tool³⁴ was used to download the complete proteomes of the 61 bacteriophages with their lysin gene product 85–98% identical to CL1-C600M2-lysin encoded in their genomes. The protein sequence dataset thus derived, along-with proteomes of *Escherichia* Phage CL1 and *Escherichia* Phage C600M2 was subjected to clustering based on sensitive search of sequences. There were 8105 protein sequences in total and those sharing at-least 80% identity and 100% query coverage were clustered using MMseqs2 tool³⁵ to identify orthologous proteins in the protein dataset.

Docking study of CL1-C600M2-lysin. Structure prediction of the CL1-C600M2-lysin protein sequence was done using I-TASSER protein modelling server²⁴ with default settings. The model with the best C-Score was selected and validated with PROCHECK³⁶ and assessed in ProSA³⁷ prior molecular docking with prominent bacterial cell wall sugar receptors from literature³⁸. The chemical entities analogous to bacterial cell wall receptors were identified from published studies and extracted from Protein Data Bank (PDB) as detailed in Table S1. The docking studies were performed using MTiAutoDock webserver²² in blind docking mode. The interacting residues were identified using LigPlot V.2.2³⁹ and PDBsum⁴⁰ and visualized using Chimera⁴¹ and PyMOL (Schrödinger, LLC) visualization software. The docked protein–ligand complexes were also visualized by embedding the ConSurf^{42,43} output derived by using the protein model of CL1-C600M2-lysin and the multi-sequence MUSCLE alignment (“[Phylogenetic tree of closely related phage lysin sequences](#)”) as input.

Experimental validation for lytic activity of the CL1-C600M2-lysin. *Plasmid construction and cloning.* CL1-C600M2-lysin coding sequence was chemically synthesized (Integrated DNA Technologies) and cloned into the pRSET-emGFP expression vector (ThermoFisher Scientific) using BamH1 and EcoRI restriction sites. This incorporated a N-terminal polyhistidine tag (His₆). A stop codon was incorporated prior to the emGFP tag at the C-terminal resulting in CL1-C600M2-lysin with His Tag of molecular weight 21.269 KDa. Successful cloning was verified using PCR (Supplemental Figure S4).

Expression of lysin and protein purification. Transformed *Escherichia coli* BL21 (DE3) pLysS were grown in Luria broth supplemented with ampicillin and chloramphenicol media in specialized Erlenmeyer flasks to an OD of 70 Klett units using a Klett colorimeter. These cultures were induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and shaken at 37 °C for 2.5 h at 225 rpm. Next, the bacteria were pelleted at 5000×g for 5 min and resuspended in 4 ml B-PER reagent (ThermoFisher Scientific) containing 1× protease inhibitor cocktail per gram cell pellet. Following incubation at room temperature for 10 min, the suspension was homogenized by sonicating 20 s, followed a 1-min rest on ice which is repeated for a total 4 bursts. The lysate was centrifuged at 15,000×g for 5 min at 4 °C to separate the soluble proteins from the insoluble proteins. Ni–NTA chromatography was used to purify the his-tagged CL1-C600M2-lysin from the soluble proteins fraction. Proteins were eluted from the His-Bind Resin, Nickel charged (Novagen, EMB Biosciences, Darmstadt, Germany) using 6 volumes of elution buffer composed of 1 M imidazole, 0.5 M NaCl, 20 mM Tris–HCl, pH 7.9). The imidazole-eluted fractions that contains the recombinant CL1-C600M2-lysin (as verified by SDS-PAGE and Western blot, Supplemental Figure S5), were combined and dialyzed for overnight at 4 °C in deionized water containing 0.1% Trifluoroacetic acid (TFA) and 5 mM Dithiothreitol (DTT). Purified CL1-C600M2-lysin was stored at –20 °C until further analysis.

Turbidity reduction assay. Turbidity reduction assay was carried out according to Vander Elst et al.⁴⁴ with some modifications. Briefly, *Escherichia coli* C, C600 (K-12), HB101 and B/R strains were grown in separate tubes overnight in Luria Broth at 37 °C in the air shaker. The cells were pelleted and washed with PBS and resuspended with a 1:1 mixture with purified CL1-C600M2-lysin to OD₆₀₀ = 2. The OD₆₀₀ was measured every 15 s at 37 °C for 1 h, shaking the 96-well plate between each measurement in the Varioskan™ LUX multimode microplate reader (ThermoScientific).

Results

Domain analysis of CL1-C600M2-lysin. The CL1-C600M2-lysin was interrogated in silico to set the precedents prior experimental confirmation for the possibility to be an “enzymobiotic”. The lysin had intrinsic domains–“cd00737: endolysin_autolysin” with no CBD, suggesting it to be a globular N-acetyl-muramidase^{14,19,20}. A search for globular endolysins in Protein Data Bank (PDB) yielded entry 6ET6⁴⁵ as the only X-ray crystal

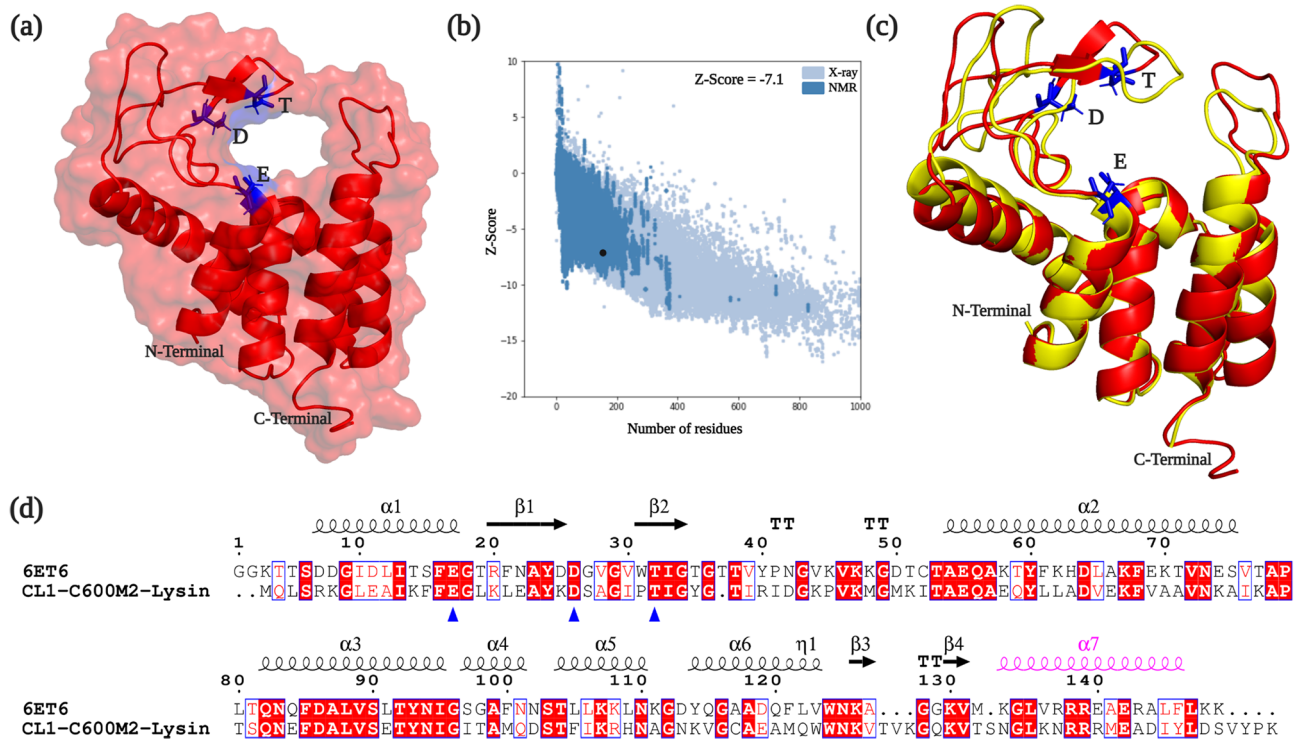


Figure 2. Sequential and Structural analysis: (a) three-dimensional CL1-C600M2-lysine modelled structure predicted by I-TASSER with labelled N and C-terminal region along with the catalytic triad Glu(E)-Asp(D)-Thr(T) represented as blue sticks. (b) Z-score plot generated by ProSA for predicted model quality assessment. (c) Superposition of CL1-C600M2-lysine (Red) with AcLys (PDB Code: 6ET6, Yellow). The catalytic triad is represented as blue sticks and labelled. (d) Multiple sequence alignment of AcLys (PDB Code: 6ET6) and CL1-C600M2-lysine. The catalytic triad is marked by blue triangles and the C-terminal α -helix enriched with positively charged residues is highlighted in Pink.

structure of endolysin AcLys classified as *N*-acetyl-muramidase antimicrobial protein encoded in genome of *Acinetobacter baumannii* AB 5075.

Structure of CL1-C600M2-Lysin was modelled using I-TASSER webserver (Fig. 2a), validated with PROCHECK³⁶ (Supplementary Fig. S2) and Quality-assessed using ProSA³⁷ (Fig. 2b) followed by structural comparison with 6ET6 (Fig. 2c). Z-Score of CL1-C600M2-lysine was -7.1 (Fig. 2b), well within the range of experimentally published structures. CL1-C600M2-lysine and 6ET6 have conserved Glu-Asp-Thr catalytic triad as represented in Fig. 2d. Superposition of the CL1-C600M2-lysine structure with PDB:6ET6 (Fig. 2c) revealed close similarity with RMSD between corresponding $\text{C}\alpha$ -atoms of 0.46 Å. C-terminal α -helix of AcLys, enriched with positively-charged residues significantly facilitates cell-wall lysis⁴⁵. Similar domain architecture could be observed in the CL1-C600M2-lysine as well (Fig. 2d).

Phylogenetic tree analysis. The CL1-C600M2-lysine sequence was subjected to protein-blast against Ref-seq_Protein database to identify homologous proteins. 61 Phage protein sequences annotated as either lysin, endolysin or putative lysin with 100% query coverage and sharing 85–98% sequence identity was determined as listed in Supplementary Table 1 and summarized in Table S2. A ML phylogenetic tree was estimated from the derived lysin sequences to discern their relatedness (Fig. 3). Apart from wastewater, these lysins were inherent in phages thriving in various similar environments like sewage, fecal samples, soil from poultry and animal farms etc. Among the 49 bacteriophages with eminent information about isolation source, 27 (55%) were native to sewage or wastewater. This clearly indicates the involvement of CL1-C600M2-lysine in host–pathogen interactions endemic to wastewater environment.

Moreover, it can be observed from the ML tree (Fig. 3) that lysins inherent in bacteriophages—*Shigella* phage Z31, *Salmonella* phage FelixO1, *Escherichia* phage ekra, *Escherichia* phage warpig, *Shigella* phage JK55, *Escherichia* phage mio, *Salmonella* phage GEC_vB_B1, *Salmonella* phage Mushroom, *Enterobacteriophage* UAB_Phi87, *Salmonella* phage DaR-2019b, *Escherichia* phage skuden, *Escherichia* phage finno cluster well with CL1-C600M2-lysine. These lysin sequences share identity ranging from 98% to 96%. Since, lysins sharing high similarity with CL1-C600M2-lysine were involved in host–pathogen interactions of multiple species in the *Enterobacteriaceae* family, this could indicate the potency of CL1-C600M2-lysine against multiple hosts.

Interestingly, lysin from *Staphylococcus* phage SA1 shares 96% identity with CL1-C600M2-lysine. It has been published that *Staphylococcus* phage SA1 is effective against infection by *Staphylococcus*, a Gram-Positive bacteria in animal models⁴⁶. This could further suggest the effectiveness of CL1-C600M2-lysine against both Gram-positive

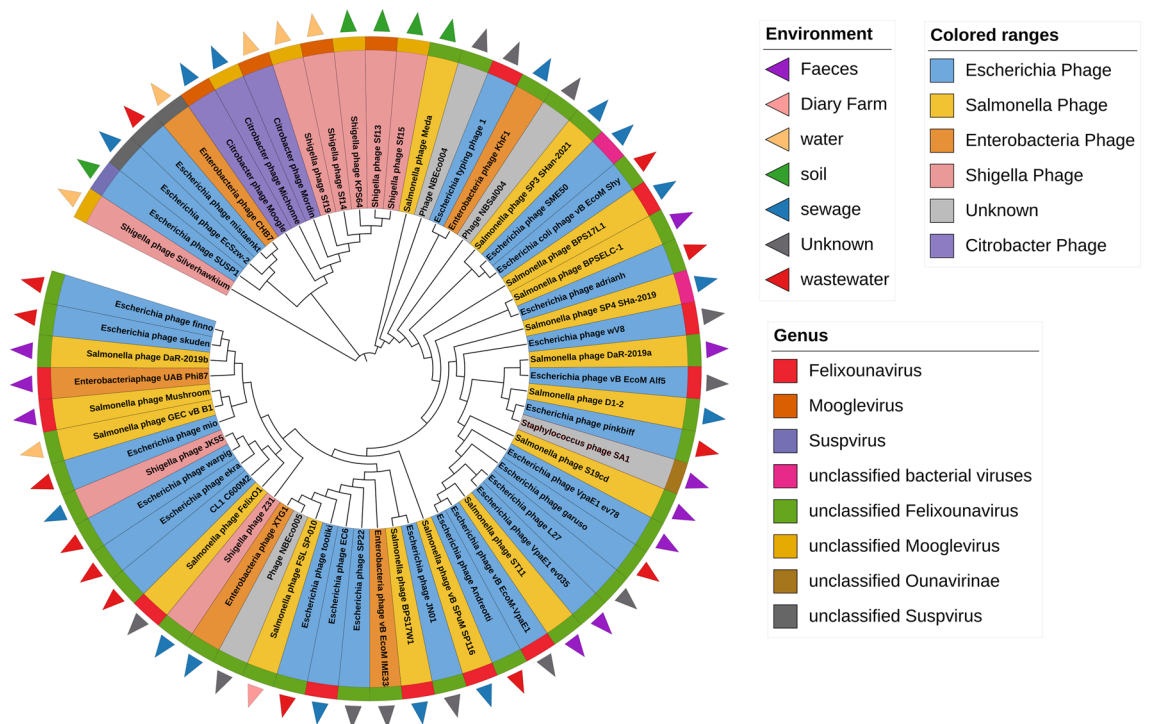


Figure 3. ML phylogenetic tree: 61 highly identical phage lysin protein sequences sharing 100% query coverage were derived by querying CL1-C600M2-lysin in Protein Blast tool against Refseq_Protein database. Colored triangles represent the generalized isolation source of the Phage as detailed in the legend titled Environment. Colored ranges indicate the host organism of the Phage. The outer circle of the ML Tree is colored respective to the Genus of the Phage.

and Gram-negative bacteria. Further, the affinity of CL1-C600M2-lysin to bacterial cell wall receptors was predicted using molecular docking studies.

Clustering analysis of phage proteome. A total of 8105 protein sequences were retrieved as described in “Comparative analysis of phage proteome” for sensitive-search clustering analysis using MMseqs2 tool³⁵. The protein sequences clustered into 797 independent clusters, such that the sequences in each cluster share at least 80% identity with 100% coverage (listed in Supplementary Table 2). There were about 368 unique clusters (46%) with only a single sequence and around 369 clusters with number of sequences ranging between 2 and 49, leaving only 60 clusters constituted by at least 50 proteins. In other words, only 60 clusters had orthologous proteins from at least 50 candidate phage species (Table S3). The conserved protein clusters majorly consisted proteins involved in Structural architecture (baseplate assembly protein, head maturation protease, hypothetical protein/baseplate assembly protein, hypothetical protein/baseplate protein, hypothetical protein/putative membrane protein, hypothetical protein/putative portal protein, hypothetical protein/structural protein, hypothetical protein/tape measure chaperone, major capsid protein and tape measure chaperone) followed by nucleotide metabolism (ribonucleoside triphosphate reductase alpha/beta subunit, glutaredoxin, thymidylate synthase, dNMP kinase, anaerobic NTP reductase, ribose-phosphate pyrophosphokinase and dihydrofolate reductase), replication module (DNA ligase, DNA primase/helicase, exonuclease, homing endonuclease/NAD synthetase, Hypothetical protein/dsDNA binding protein and terminase large subunit), Lytic module (endolysin/lysin, holin-hypothetical or putative holin, Hypothetical protein/i-spanin and rIIB lysis inhibitor) and Hypothetical protein of unknown function. Among the 60 conserved protein clusters, 31 of them comprised of Hypothetical proteins of unknown function (50–63 proteins of the kind in each cluster) followed by 2 clusters of Baseplate assembly proteins (63 proteins of the kind in each cluster) and the rest were disjoint clusters (50–63 proteins of each kind in each cluster).

Further to it, search for phage host recognition proteins in the protein dataset, revealed the prevalence of 19 host recognition proteins (annotated as Tail Fiber Protein, Putative Tail Fiber Protein, Putative Tail Protein, Putative Tail Fiber Protein Gp37, Tail Sheath Protein, Putative Tail Tape Measure Chaperone, Hk97 Major Tail Subunit, Head–Tail Preconnector Protein, Side Tail Fiber Protein, Long Tail Fiber Protein, Conserved Tail Assembly Protein, Tail Tube Protein, Tail Protein, Putative Tail Tape Measure Protein, Tail Tape Measure Protein, Tail Assembly Protein, Tail Length Tape Measure Protein and Minor Tail Protein) clustered within 69 clusters (Table S3).

Molecular docking study of CL1-C600M2-lysin. It can be noted from Supplementary Fig. S3, and Fig. 4a,b that the residues of CL1-C600M2-lysin interacting with WTA and TS were highly conserved amongst

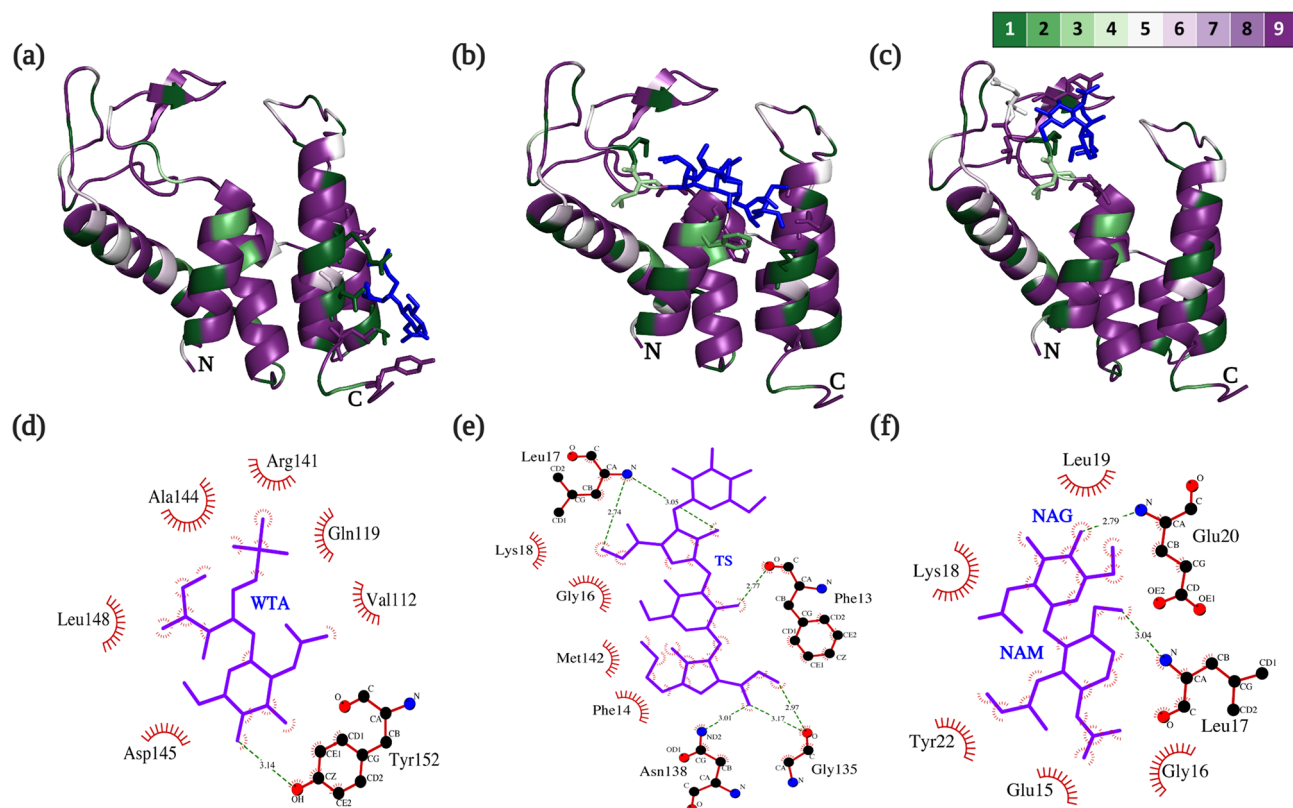


Figure 4. Protein–ligand docked complexes: the 3D structure of CL1–C600M2–lysin is represented using ConSurf^{42,43} output complexed with ligands (a) WTA; (b) TS and (c) PG. Interacting residues are denoted as sticks and colored based on the represented conservation scale with dark purple being highly conserved and dark green being highly variable. N and C terminal of each structure is indicated by N and C, respectively. All ligands are represented in dark blue irrespective of their kind. Ligplot maps for interacting residues for protein ligand complexes, (d) CL1–C600M2–lysin–WTA, (e) CL1–C600M2–lysin–TS and (f) CL1–C600M2–lysin–PG. The sugar components of PG, β -(1,4) linked *N*-acetylglucosamine (NAG) and *N*-acetylmuramic acid (NAM) are labelled.

Cell wall receptor	Binding affinity (kcal/mol)	Interacting residues
WTA	– 10.67	Val112, Gln119, Arg141 , Ala144 , Asp145 , Leu148 , Tyr152*
TS	– 11.15	Phe13* , Phe14 , Gly16 , Leu17* , Lys18 , Gly135* , Asn138* , Met142
PG	– 6.53	Glu15 , Gly16 , Leu17* , Lys18 , Leu19 , Glu20* , Tyr22

Table 1. Summary of molecular docking analysis of CL1–C600M2–lysin with selected bacterial cell surface receptors as represented in Ligplot interaction map in Fig. 4d–f. The binding affinity of all the protein–ligand partners were for the 1st ligand conformation as predicted by MTiAutoDock. Sequential amino acid residues are in bold. *Residues forming hydrogen bond.

similar phage lysin sequences derived from protein-blast analysis. It can be deduced from Fig. 4c that PG substrate binding residues are within the binding cavity surrounding the catalytic triad and mostly conserved.

Table 1 summarizes the interacting residues with each ligand and the corresponding binding affinity values. Sequential interacting amino acid residues could probably suggest an interacting motif in CL1–C600M2–lysin that facilitates binding to the corresponding ligand. In CL1–C600M2–lysin–WTA protein–ligand complex, the motif Arg141–Ala144–Asp145–Leu148–Tyr152 could be the motif involved in recognition of Gram-Positive bacterial cell wall with Tyr152 forming a hydrogen bond of length 3.14 Å (Fig. 4d). While in CL1–C600M2–lysin–TS complex, the motif Phe13–Phe14–Gly16–Leu17–Lys18 might be involved in recognition of Gram-negative bacterial cell wall with Phe13 (2.77 Å) and Leu17 (2.74 Å and 3.25 Å) forming hydrogen bonds (Fig. 4e). The dual purpose of motif Phe13–Phe14–Glu15–Gly16–Leu17–Lys18 and efficiency of CL1–C600M2–lysin against Gram-negative bacterial cell wall could be speculated from its interaction and hydrogen bond formation with both TS and PG (Fig. 4e,f).

Assessment of lytic activity of CL1–C600M2–lysin. To assess the lytic capacity of CL1–C600M2–lysin turbidity reduction assay was performed on *Escherichia coli* C, C600 (K-12), HB101 and B/R strains (Fig. 5) by

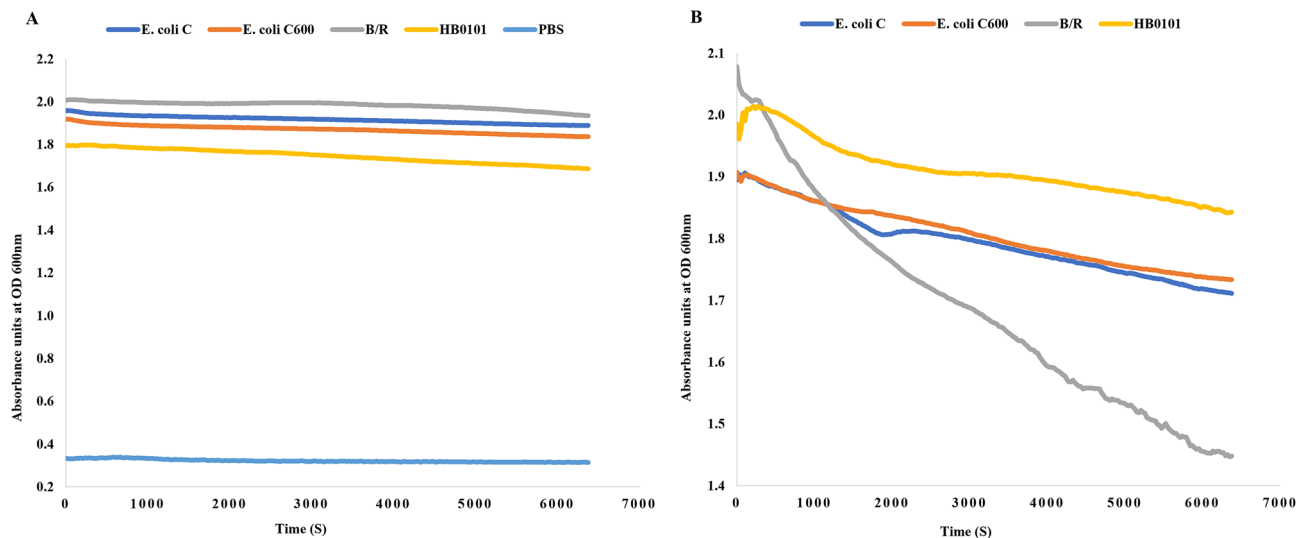


Figure 5. Turbidity reduction assays: (A) *Escherichia coli* strains C, C600, B/R and HB101 in PBS serve as negative controls. (B) The same *Escherichia coli* strains with 80 ng of purified CL1-C600M2-lysins.

mixing 80 ng of purified CL1-C600M2-lysins. The bacterial strains in PBS served as negative control for comparison. It can be well observed from Fig. 5B that the CL1-C600M2-lysins was most effective against *Escherichia coli* B/R followed by *Escherichia coli* C, C600 and HB101 strains. While the negative control does not show reduction in the OD values. Thus, the turbidity reduction assay using CL1-C600M2-lysins clearly predicates its utility as a possible anti-bacterial agent.

Discussion

Phage infection machinery majorly consist of tail fibers, base plates and lysins. Host recognition occurs through a reversible interaction of the tip of the long tail fibers with bacterial outer-membrane (OM) components⁴⁷. This activates the baseplate and binding of short tail fibers on cell-surface receptors, followed by contraction of the outer tail sheath and penetration of inner tail tube into cell-membrane and injection of the viral-DNA into the bacteria^{48,49}. Each Phage-Host interaction is attuned with great specificity^{50,51}.

In-silico analysis of endolysins. Endolysins are bifunctional enzymes marshalling precise binding to bacterial-receptor and enzymatic hydrolysis of its cell-wall which is exhorted by crucial domains. The endolysins of most phages that infect gram-negative bacteria, comprise a single catalytic domain, whereas gram-positive bacterial phages embody N-terminal EAD and C-terminal CBD. The location and sizes of the domains are known to vary. The CBD and its appurtenant binding site eventually dictate the endolysin's substrate-affinity and therefore robust bacterial infection⁵².

In consensus, endolysins such as AcLys and PlyE146 categorized as muramidases and tested for their anti-microbial ability against pathogens *A. baumannii*, *E. coli*, *P. aeruginosa*, *K. pneumoniae* and *S. enterica* have N-terminal “cd00737: endolysin_autolysin” domain and a highly positive-charged stretch of amino acids at their C-Terminal^{45,53}. This domain architecture highly correlates with CL1-C600M2-lysins accentuating it as promising enzymatic candidate. Protein-Blast of CL1-C600M2-lysins (Fig. 3) corroborates the likelihood of broad-range of target pathogens like *Enterobacteria*, *Escherichia*, *Shigella* and *Salmonella*.

A study by Grose et al.⁵⁴ classifies bacteriophages into 56 clusters inclusive of 32 lytic and 24 temperate clusters based on available genomic sequence data. Inter-cluster members share genomic similarity but significant dissimilarity intra-cluster. It could be inferred from Fig. 3 that the CL1-C600M2-lysins corresponds to the lytic cluster “Felix-O1-like” it is 99% identical with *Salmonella phage felix-O1*. This particular lysins is widely proven to target majority of *Salmonella* family and employed in effective biocontrol of pathogenic organisms^{55,56}. Lysins of *Escherichia phage vB_EcoM_VpaE1* with host range of VpaE1 *E. coli* B strains like BE, BL21, BL21(DE3), B40, BE-BS⁵⁷ also shares 98% identity with CL1-C600M2-lysins. *Salmonella phage Mushroom* is a constituent of IntestiPhage, a combination of 23 phages capable to infect a range of enterobacteria strains⁵⁸ while *Samonella Phage vB_SpuM_SPI16* is capable of infecting 9 serotypes of *Salmonella*, namely, *Pullorum*, *Enteritidis*, *Indiana*, *Typhimurium*, *Infantis*, *Montevideo Heidelberg*, *Paratyphi A*, and *Derby*. Both the afore-mentioned phages share 96% and 97% identity with the CL1-C600M2-lysins, respectively. Interestingly, the research study by Low et al.⁵⁹ correlates net-positive charge of the catalytic domain of lysins with bactericidal efficiency. CL1-C600M2-lysins has a net charge of +6.1 at Ph-7.0 (using <http://protcalc.sourceforge.net/>), upholding it to be a potential “enzymatic”. Further in silico analysis through clustering of Proteome from wastewater-endemic phages, molecular docking studies, and experimental validation has suggested the same.

Analysis of proteome from wastewater-endemic phages. Systematic clustering analysis of the protein dataset derived from proteome of 61 phages, *Escherichia phage CL1* and *Escherichia phage C600M2* was

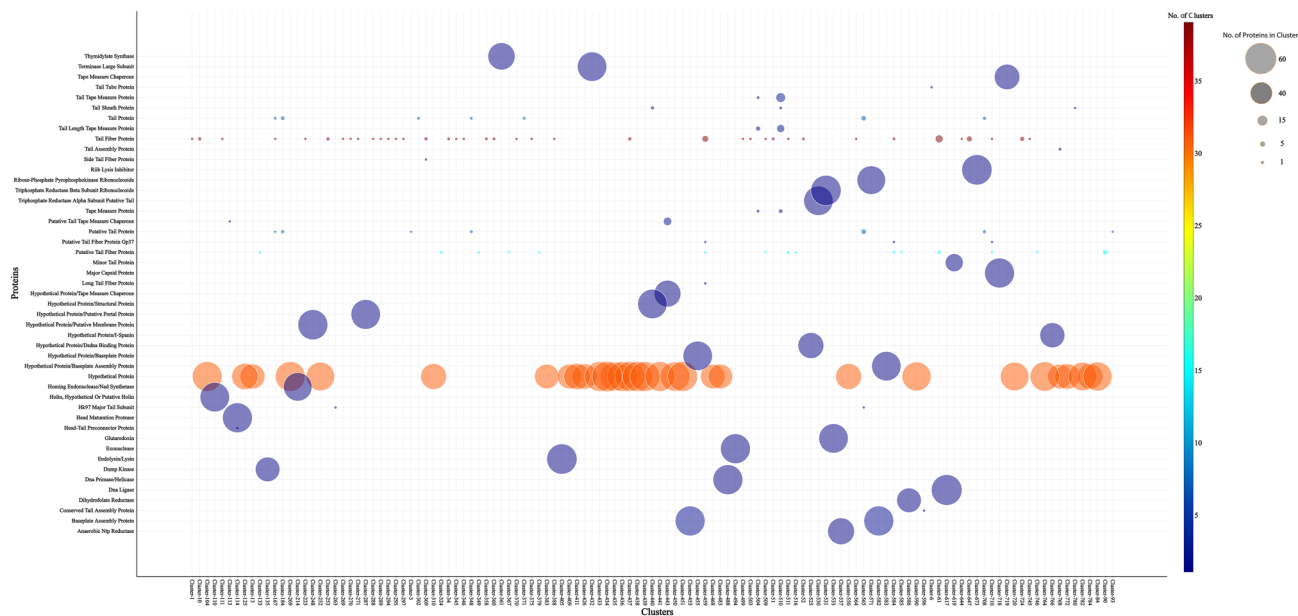


Figure 6. Bubble plot of crucial protein clusters. Crucial proteins clusters distilled by clustering of Phage proteome are represented as bubbles. Color of bubble represents the number of clusters that contain the protein kind, ranging from dark brown to dark blue. And the size of the bubble represents the quantity of the protein kind within the cluster.

done to unravel their inherent biodiversity in spite of likeness of their lysin proteins. The two major cluster sets of interest were the conserved clusters and the host-recognition protein clusters. It can be deciphered from Fig. 6 that majority of the host recognition proteins of the candidate phages were grouped into multiple clusters, especially the Tail fiber proteins (39 clusters, quantity: 1–12), Putative Tail fiber proteins (15 clusters; quantity: 1–6), Putative Tail protein (7 clusters; quantity: 1–6) and Tail proteins (7 clusters; quantity: 2–6). Interestingly, all the 35 minor tail proteins in the dataset clustered into a disjoint cluster.

Typically, phage–host adsorption is a two-step process involving preliminary reversible attachment to the host cell receptors followed by irreversible attachment either by strengthening the initial bond or by binding to secondary receptors. Distal phage tail elements like Tail fibers form reversible interactions with exposed and highly accessible host cell wall components while irreversible interactions with secondary host-receptors are facilitated by short/minor tail proteins⁶⁰. The tail structures of bacteriophages are the key determinants of host specificity and the observation of diverse distal tail protein elements despite a highly conserved lysin among these bacteriophages endemic to wastewater could indicate the potency of CL1–C600M2-lysin against multiple bacterial species.

Molecular docking studies. Phage lysins typically recognize varied bacterial cell-wall receptors depending on their host range. However, due to insufficient research on designated ligands, it can be presumed that lysins mostly target cell wall carbohydrates and the specificity of its cell wall binding region determines its range of target organisms⁵². WTA and LPS are one of unique cell surface ligands that distinguish between gram-positive and gram-negative bacteria. Minimal repeating glyco-polymers of the above-mentioned ligands used in similar experimental initiatives were used in molecular docking studies with CL1–C600M2-lysin to identify propitious cell wall binding motifs. The synthetic PG analog used in the study of T4 phage lysozyme’s enzyme–substrate interactions was also implemented in docking studies to computationally model and interrogate substrate affinity of CL1–C600M2-lysin.

With reference to Fig. S3, both the N-terminal catalytic and C-Terminal cell-wall binding regions of CL1–C600M2-lysin are well conserved among the lysins from bacteriophages isolated from wastewater and similar environments (Table S2). Probably suggesting multi-host versatility of its cell-wall binding region and proving to be the right candidate to designate as an “enzymic” bio-control agent for wastewater treatment purposes.

Adding to it, most interacting residues of CL1–C600M2-lysin irrespective of the ligand were hydrophobic (Fig. 7). Corroborating the study by Yan et al.⁶¹, residues in proximity of predicted cell-wall binding motifs of CL1–C600M2-lysin could be modified to be more hydrophobic for improved efficiency for exogenous application.

Antimicrobial activity of CL1–C600M2-lysin. The turbidity reduction assay was used to demonstrate the antimicrobial activity of CL1–

C600M2-lysin in several *Escherichia coli* strains. The decrease in turbidity in a bacterial solution is an indirect measure of cell death. The effect of factors such as buffer components and osmotic pressure changes are accounted for in the bacteria and PBS controls (Fig. 5A). Therefore, the turbidity reduction observed in the four *Escherichia*

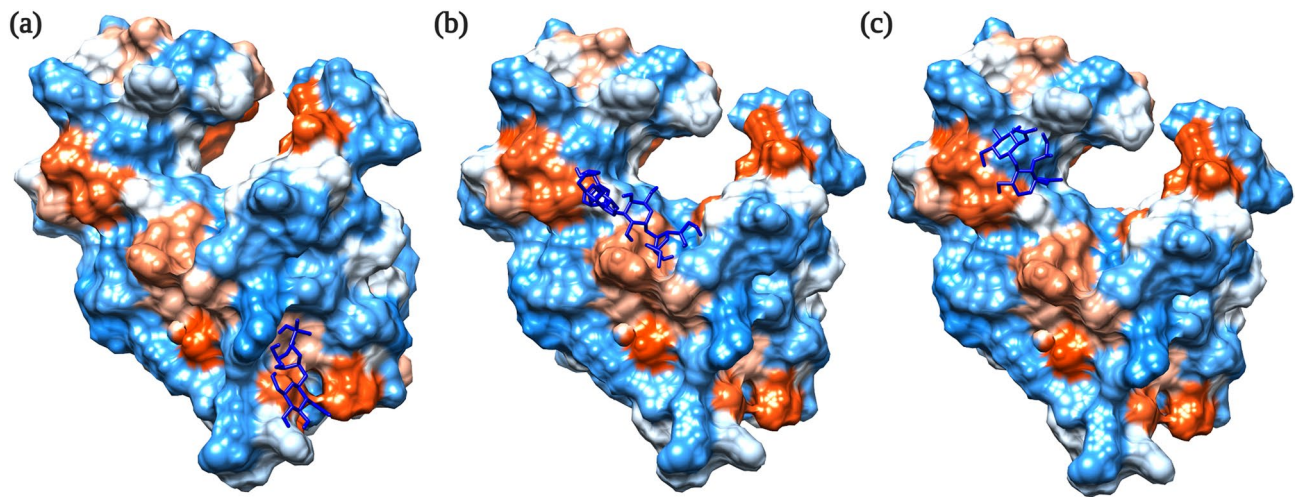


Figure 7. Hydrophobicity of protein–ligand complexes. The 3-dimensional structure of CL1-C600M2-lysin is represented using the hydrophobicity attribute in Chimera, complexed with ligands (a) WTA; (b) TS and (c) PG. The protein surface is colored based on hydrophobicity of the underlying residues ranging from blue, being highly hydrophilic to red, being highly hydrophobic. All ligands are represented in dark blue irrespective of their kind.

coli strains in the presence of CL1-C600M2-lysin is indicative of the possible “enzybiotic” nature of the lysin and corroborates with the bioinformatics analysis.

Overall, we present comprehensive functional bioinformatics analysis and experimental validation of identical lysin gene products (collectively termed as CL1-C600M2-lysin) identified from two novel *Myoviridae* bacteriophages, *Escherichia* Phage C600M2 and *Escherichia* Phage CL1, which were isolated from wastewater treatment plant in the State of Qatar. CL1-C600M2-lysin was analyzed in-silico to gain insights for their practicability as prospective “enzybiotics” for water-treatment and set the necessary precedents prior experimental venture. Followed by experimental assessment of the lytic activity of CL1-C600M2-lysin using turbidimetric reduction assay.

Conclusions

In summary, by using efficient computational strategies of comparative sequence analysis, proteome clustering, protein structure modelling, protein structural analysis and molecular docking studies we present complete in silico characterization of identical lysin, CL1-C600M2-lysin, from two bacteriophages isolated from the wastewater samples collected from a treatment plant in State of Qatar. Further experimental investigation of detrimentality of CL1-C600M2-lysin towards laboratory strains of *Escherichia coli* revealed the lysin to be most lethal towards *Escherichia coli* B/R followed by *Escherichia coli* C, C600 and HB101 strains.

Encompassing comprehensive computational characterization of CL1-C600M2-lysin to set the essential premises for substantial experimentation withal experimental confirmation of the lysin’s utility as a potential “enzybiotic”, this study presents a novel amalgamation of research strategies to corroborate prospectiveness of CL1-C600M2-lysin to be purposed as a biocontrol agent distinctive to wastewater environment.

Given the necessity to preserve scarce water resources, reclamation of treated wastewater is an essential step towards water security. CL1-C600M2-lysin can potentially be used as an efficient “enzybiotic” to biocontrol wastewater endemic bacteria. CL1-C600M2-lysin could eradicate the residual pathogenic bacteria still persistent in treated wastewater to assure safety and quality-control prior recycling. Moreover, research inquiry for the therapeutic potential of CL1-C600M2-lysin could further widen the scope of this study.

Data availability

Escherichia Phage C600M2 (GenBank accession: OK040807.1) and Lysin (GenBank accession: UCJ01465.1) Phage Phage CL1 (GenBank accession: OK040806.1) and Lysin (GenBank accession: UCJ01321.1).

Received: 4 March 2022; Accepted: 20 May 2022

Published online: 31 May 2022

References

- Shomar, B. *et al.* Optimization of wastewater treatment processes using molecular bacteriology. *J. Water Process Eng.* **33**, 101030 (2020).
- Howard-Varona, C., Hargreaves, K. R., Abedon, S. T. & Sullivan, M. B. Lysogeny in nature: Mechanisms, impact and ecology of temperate phages. *ISME J.* **11**, 1511–1520 (2017).
- Da Silva, M. F. *et al.* Antibiotic resistance of enterococci and related bacteria in an urban wastewater treatment plant. *FEMS Microbiol. Ecol.* **55**, 322–329 (2006).
- Novo, A. & Manaia, C. M. Factors influencing antibiotic resistance burden in municipal wastewater treatment plants. *Appl. Microbiol. Biotechnol.* **87**, 1157–1166 (2010).

5. Watkinson, A. J., Micalizzi, G. B., Graham, G. M., Bates, J. B. & Costanzo, S. D. Antibiotic-resistant *Escherichia coli* in wastewaters, surface waters, and oysters from an urban riverine system. *Appl. Environ. Microbiol.* **73**, 5667–5670 (2007).
6. Łuczkiwicz, A., Jankowska, K., Fudala-Książek, S. & Olańczuk-Neyman, K. Antimicrobial resistance of fecal indicators in municipal wastewater treatment plant. *Water Res.* **44**, 5089–5097 (2010).
7. Loc-Carrillo, C. & Abedon, S. T. Pros and cons of phage therapy. *Bacteriophage* <https://doi.org/10.4161/bact.1.2.14590> (2011).
8. Adibi, M., Mobasher, N., Ghasemi, Y., Mohkam, M. & Mobasher, M. A. Isolation, purification and identification of *E. coli* O157 phage for medical purposes. *Trends Pharm. Sci.* **3**, 43–48 (2017).
9. Maal, K. B., Delfan, A. S. & Salmanzadeh, S. Isolation and identification of two novel *Escherichia coli* bacteriophages and their application in wastewater treatment and coliform's phage therapy. *Jundishapur J. Microbiol.* **8**, 25 (2015).
10. Veiga-Crespo, P., Ageitos, J. M., Poza, M. & Villa, T. G. Enzybiotics: A look to the future, recalling the past. *J. Pharm. Sci.* **96**, 1917–1924 (2007).
11. Czaplewski, L. *et al.* Alternatives to antibiotics—a pipeline portfolio review. *Lancet Infect. Dis.* **16**, 239–251 (2016).
12. Schuch, R., Nelson, D. & Fischetti, V. A. A bacteriolytic agent that detects and kills *Bacillus anthracis*. *Nature* **418**, 884–889 (2002).
13. Dams, D. & Briens, Y. Enzybiotics: Enzyme-based antibacterials as therapeutics. In *Therapeutic Enzymes: Function and Clinical Implications* 233–253 (Springer, 2019).
14. Fischetti, V. A. Bacteriophage endolysins: A novel anti-infective to control Gram-positive pathogens. *Int. J. Med. Microbiol.* **300**, 357–362 (2010).
15. Srinivasan, R. *et al.* Recombinant engineered phage-derived enzybiotic in *Pichia pastoris* X-33 as whole cell biocatalyst for effective biocontrol of *Vibrio parahaemolyticus* in aquaculture. *Int. J. Biol. Macromol.* **154**, 1576–1585 (2020).
16. Matamp, N. & Bhat, S. G. Phage endolysins as potential antimicrobials against multidrug resistant *Vibrio alginolyticus* and *Vibrio parahaemolyticus*: Current status of research and challenges ahead. *Microorganisms* **7**, 84 (2019).
17. Mayer, M. J., Narbad, A. & Gasson, M. J. Molecular characterization of a *Clostridium difficile* bacteriophage and its cloned biologically active endolysin. *J. Bacteriol.* **190**, 6734–6740 (2008).
18. Fischetti, V. A. Development of phage lysins as novel therapeutics: A historical perspective. *Viruses* **10**, 310 (2018).
19. Callewaert, L., Walmagh, M., Michiels, C. W. & Lavigne, R. Food applications of bacterial cell wall hydrolases. *Curr. Opin. Biotechnol.* **22**, 164–171 (2011).
20. Vidová, B., Šramková, Z., Tišáková, L., Oravkinová, M. & Godány, A. Bioinformatics analysis of bacteriophage and prophage endolysin domains. *Biologia (Bratisl.)* **69**, 541–556 (2014).
21. Fischetti, V. A. Bacteriophage lytic enzymes: Novel anti-infectives. *Trends Microbiol.* **13**, 491–496 (2005).
22. Labbé, C. M. *et al.* MTiOpenScreen: A web server for structure-based virtual screening. *Nucleic Acids Res.* **43**, W448–W454 (2015).
23. Kemege, K. E. *et al.* Ab initio structural modeling of and experimental validation for *Chlamydia trachomatis* protein CT296 reveal structural similarity to Fe (II) 2-oxoglutarate-dependent enzymes. *J. Bacteriol.* **193**, 6517–6528 (2011).
24. Yang, J. *et al.* The I-TASSER Suite: Protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
25. Azam, S. S. & Abbasi, S. W. Molecular docking studies for the identification of novel melatoninergic inhibitors for acetylserotonin-O-methyltransferase using different docking routines. *Theor. Biol. Med. Model.* **10**, 63 (2013).
26. Ramadoss, R., Al-Marzooqi, F., Shomar, B., Ilyin, V. A. & Vincent, A. S. Genomic characterization and annotation of two novel bacteriophages isolated from a wastewater treatment plant in Qatar. *Microbiol. Resour. Announc.* <https://doi.org/10.1128/MRA.01090-21> (2022).
27. Lu, S. *et al.* CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Res.* **48**, D265–D268 (2020).
28. Joshi, T. & Xu, D. Quantitative assessment of relationship between sequence similarity and function similarity. *BMC Genom.* **8**, 222 (2007).
29. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
30. Darrriba, D. *et al.* ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* **37**, 291–294 (2019).
31. Minh, B. Q. *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
32. Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
33. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
34. Batch Entrez BT - *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine*. In 131 (Springer, 2006). https://doi.org/10.1007/3-540-29623-9_6283.
35. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
36. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* <https://doi.org/10.1107/s0021889892009944> (1993).
37. Wiederstein, M. & Sippl, M. J. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **35**, W407–W410 (2007).
38. Zhang, R. *et al.* Lysozyme's lectin-like characteristics facilitates its immune defense function. *Q. Rev. Biophys.* **50**, 25 (2017).
39. Laskowski, R. A. & Swindells, M. B. LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* <https://doi.org/10.1021/ci200227u> (2011).
40. Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S. & Thornton, J. M. PDBsum: Structural summaries of PDB entries. *Protein Sci.* **27**, 129–134 (2018).
41. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
42. Glaser, F. *et al.* ConSurf: Identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* **19**, 163–164 (2003).
43. Landau, M. *et al.* ConSurf 2005: The projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299–W302 (2005).
44. Vander Elst, N. *et al.* Characterization of the bacteriophage-derived endolysins PlySs2 and PlySs9 with in vitro lytic activity against bovine mastitis *Streptococcus uberis*. *Antibiotics* **9**, 621 (2020).
45. Sykilinda, N. N. *et al.* Structure of an Acinetobacter broad-range prophage endolysin reveals a C-terminal α -helix with the proposed role in activity against live bacterial cells. *Viruses* **10**, 309 (2018).
46. Cui, Z. *et al.* Safety assessment of *Staphylococcus phages* of the family Myoviridae based on complete genome sequences. *Sci. Rep.* **7**, 1–8 (2017).
47. Yu, F. & Mizushima, S. Roles of lipopolysaccharide and outer membrane protein OmpC of *Escherichia coli* K-12 in the receptor function for bacteriophage T4. *J. Bacteriol.* **151**, 718–722 (1982).
48. Aksyuk, A. A. *et al.* The tail sheath structure of bacteriophage T4: A molecular machine for infecting bacteria. *EMBO J.* **28**, 821–829 (2009).

49. Kostyuchenko, V. A. *et al.* The tail structure of bacteriophage T4 and its mechanism of contraction. *Nat. Struct. Mol. Biol.* **12**, 810–813 (2005).
50. Dimitrov, D. S. Virus entry: Molecular mechanisms and biomedical applications. *Nat. Rev. Microbiol.* **2**, 109–122 (2004).
51. Skehel, J. J. & Wiley, D. C. Receptor binding and membrane fusion in virus entry: The influenza hemagglutinin. *Annu. Rev. Biochem.* **69**, 531–569 (2000).
52. Jarábková, V., Tišáková, L. & Godány, A. Phage endolysin: A way to understand a binding function of C-terminal domains a mini review. *Nov. Biotechnol. Chim.* **14**, 117–134 (2015).
53. Larpin, Y. *et al.* In vitro characterization of PlyE146, a novel phage lysin that targets Gram-negative bacteria. *PLoS One* **13**, e0192507 (2018).
54. Grose, J. H. & Casjens, S. R. Understanding the enormous diversity of bacteriophages: The tailed phages that infect the bacterial family Enterobacteriaceae. *Virology* **468**, 421–443 (2014).
55. Radford, D. *et al.* Characterization of antimicrobial properties of Salmonella phage Felix O1 and Listeria phage A511 embedded in xanthan coatings on Poly (lactic acid) films. *Food Microbiol.* **66**, 117–128 (2017).
56. Whichard, J. M., Sriranganathan, N. & Pierson, F. W. Suppression of Salmonella growth by wild-type and large-plaque variants of bacteriophage Felix O1 in liquid culture and on chicken frankfurters. *J. Food Prot.* **66**, 220–225 (2003).
57. Šimoliūnas, E., Vilkaitytė, M. & Kaliniene, L. Incomplete LPS core-specific Felix01-like virus vB_EcoM_VpaE1. *Viruses* **7**, 6163–6181 (2015).
58. Tolen, T. N., Xie, Y., Hernandez, A. C. & Everett, G. F. K. Complete genome sequence of *Salmonella enterica* serovar *Typhimurium myophage* Mushroom. *Genome Announc.* **3**, 25 (2015).
59. Low, L. Y., Yang, C., Perego, M., Osterman, A. & Liddington, R. Role of net charge on catalytic domain and influence of cell wall binding domain on bactericidal activity, specificity, and host range of phage lysins. *J. Biol. Chem.* **286**, 34391–34403 (2011).
60. Bertozzi Silva, J., Storms, Z. & Sauvageau, D. Host receptors for bacteriophage adsorption. *FEMS Microbiol. Lett.* **363**, 25 (2016).
61. Yan, G. *et al.* External lysis of *Escherichia coli* by a bacteriophage endolysin modified with hydrophobic amino acids. *AMB Express* **9**, 1–7 (2019).

Author contributions

R.R. conducted investigation, visualized the study, prepared Figs. 1, 2, 3, 4, 6 and 7 and edited the manuscript text. M.A. conducted investigation and prepared Fig. 5. B.S. Obtained samples and wrote original draft manuscript. V.A.I. supervised the study and assisted in writing original manuscript draft. A.S.V. conceptualized the study, supervised, administered the project and acquired funding. All authors reviewed the manuscript.

Funding

This publication was made possible by NPRP 10-0119-170197 from the Qatar National Research Fund (a member of Qatar Foundation). The findings herein reflect the work and are solely the responsibility of the authors. Article Processing Charge (APC) fund was provided by Carnegie Mellon University Libraries and Biological Sciences, Carnegie Mellon University Qatar.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-13171-8>.

Correspondence and requests for materials should be addressed to A.S.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022