OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

In Silico Structural Characterization of Arthrobacter Phage Alatato Immunity Cassette Reveals a Conserved Arginine-Mediated Repressor Motif

Rakin Khan and Aashi Kothari
khanr5@kgv.hk, kothaa6@kgv.hk

ABSTRACT

Understanding the regulatory mechanisms of viral lysogeny requires precise characterization of repressor-operator interactions. In this study, we investigated the unmapped immunity cassette of Arthrobacter phage Alatato (Cluster FB) utilizing silico structural modeling. We identified gp31 and gp33 as potential functional DNA-binding repressors, which are predicted to utilize a strictly conserved Helix-Turn-Helix (HTH) motif to engage the major groove of the operator DNA. Through structural mapping and computational mutagenesis via AlphaFold 3, computational analysis suggests Arginine 51 (R51) may serve as a candidate residue mediating electrostatic interactions with the DNA phosphate backbone; in silico substitution of R51 to Alanine (R51A) notably reduced the predicted structural confidence of the binding complex. Furthermore, structural alignment of Alatato gp31 with its homolog from the genetically divergent Cluster FA phage Bridgette (Pham 289627) yielded a 0.556 Å RMSD, consistent with a highly conserved structural architecture at the R51 interface. Conversely, the Cluster AZ1 phage Adolin (Pham 291150) exhibited no predictive binding affinity (ipTM 0.22) for the Alatato operator sequence, suggestive of functional divergence between these repressor families under the current model parameters. These predictive findings propose a conserved, Arginine-mediated binding mechanism that is modeled as structurally maintained across distinct viral clusters.

Keywords: AlphaFold 3, UCSF ChimeraX, *Arthrobacter* phage, immunity cassette, Helix-Turn Helix motif, repressor-operator interaction, protein folding

1. INTRODUCTION

The determination of protein and nucleic acid 3D structures is crucial to understanding their biological functions. However, traditional experimental methods such as X-ray crystallography and cryo-electron microscopy are time-consuming and expensive. For several decades, the prediction of protein structure

directly from sequence information has been an unachievable dream. The advancements in different computational methods to predict tertiary structures have now allowed researchers to study local motifs, molecular folding, and structure-function relationships [1,2]. While earlier machine learning and deep learning models, such as trRosetta and AlphaFold1, made significant strides in predicting single-chain proteins, proteins usually function through complex interactions with other proteins or DNA molecules [3,4]. Predicting these multi-molecular interactions has historically been highly challenging. Recently, the release of AlphaFold 3 has created a milestone in computational biology by expanding deep learning capabilities beyond isolated proteins to accurately predict the structures of complexes containing proteins, DNA, RNA, and small molecules [5]. This allows researchers to model repressor-operator interactions de novo, providing unprecedented insight into genetic regulatory mechanisms without the immediate need for experimental structural solving.

Understanding the regulatory mechanisms of bacteriophages, specifically the lysogenic life cycle, is imperative for advancing our knowledge of viral evolution and developing novel biotechnologies. During lysogeny, temperate phages integrate their genetic material into the host genome, maintained in a dormant state by a genetic switch known as the immunity cassette [6,7]. The repressor proteins encoded within this region must bind with high specificity to operator DNA sequences to prevent the transcription of lytic genes. However, due to the high rate of viral mutagenesis and genetic divergence, experimental structures for the vast majority of these repressor-operator complexes are completely unknown [8]. In Arthrobacter phages, specifically the Cluster FB phage Alatato, the exact functional DNA-binding proteins within the putative immunity cassette (gp31-gp34) and their atomic-level interactions with the host DNA have remained unmapped. A thorough examination of the repressor protein structure and potential electrostatic interaction residues can allow for predictions of evolutionary conservation across different phage clusters, saving significant time compared to large-scale in vitro laboratory experiments [9].

In this study, we use de novo prediction of the 3D structures of the putative Alatato repressor sequences utilizing the AlphaFold 3 algorithm to model interactions with a 40-bp operator DNA sequence. Subsequently, refinement and structural alignments of the predicted complexes are conducted using UCSF ChimeraX. We predicted the 3D structures of the Alatato gp31 and gp33 proteins, identifying a conserved Helix-Turn-Helix (HTH) motif that engages the major groove of the DNA. To test the dependency of this interaction, we further performed in silico mutagenesis at a highly conserved electrostatic residue (Arginine 51) using AlphaFold 3 to observe shifts in binding confidence (ipTM) and Predicted Aligned Error (PAE). Finally, to investigate the evolutionary conservation and functional divergence of this genetic switch, we compared the Alatato structure against its homolog from a divergent cluster (Cluster FA phage Bridgette) and an incompatible repressor from a sibling cluster (Cluster AZ1 phage Adolin). This study introduces a completely in silico pipeline methodology that can be used to map the 'dark matter' of viral genomes and rapidly generate predictive models for private regulatory switches across divergent viral families.

2. MATERIALS AND METHODS

2.1. Sequence Identification and Acquisition

Amino acid sequences for the Arthrobacter phage Alatato immunity cassette proteins gp31, gp32, gp33, and gp34 were retrieved from the Actinobacteriophage Database (PhagesDB, [10] https://phagesdb.org, accessed on [April 1st] 2026). Proteins gp31 and gp33 were selected as the primary repressor candidates due to their genomic location immediately adjacent to the putative integration cassette, a typical hallmark of lysogenic switch regulators. The genomic control center DNA sequence, located in the intergenic region upstream of the immunity cassette, was identified and extracted for use in protein DNA docking simulations. The target operator DNA sequence was defined as a 40-bp region to ensure sufficient length to capture a full turn of the DNA double helix and accommodate potential repressor binding footprints without introducing excessive computational overhead. For comparative structural and evolutionary analysis, homologous sequences were retrieved for Arthrobacter phages Bridgette (Cluster FA, Pham 289627) and Adolin (Cluster AZ1, Pham 291150) to evaluate the conservation of the helix-turn-helix (HTH) motifs.

2.2. De Novo Structural Prediction and Complex Modeling

De novo structural models of the gp31–gp34 proteins and their respective DNA complexes were generated using the AlphaFold 3 algorithm (Google DeepMind, 2024 release) via the AlphaFold Server (https://alphafoldserver.com/). Sequences were modeled utilizing the AlphaFold Multimer framework to assess both individual protein folding and the formation of quaternary homodimeric or heteromeric assemblies in the presence of the target operator DNA. Simulations utilized default server parameters, including standard recycling steps and enabled structural relaxation via AMBER force fields to resolve steric clashes.

The algorithm's performance was quantified using the Predicted Aligned Error (PAE) matrix and the interface predicted TM-score (ipTM). Models with an ipTM > 0.6 were prioritized for downstream interface analysis, while PAE plots were utilized to determine confidence in the relative positioning of the HTH motifs in relation to the DNA major and minor grooves.

2.3. Visualization and Interface Analysis

Predicted structures were imported into UCSF ChimeraX v1.7 [11] for high-resolution 3D visualization and atomic analysis. Protein–DNA interfaces were analyzed by isolating amino acid residues within a 4.0 Å threshold of the DNA backbone using the "findclash" command with hydrogen bond detection enabled to identify possible, rather than confirmed, interactions. Electrostatic surfaces were calculated using the Coulombic coloring tool, and HTH motifs were mapped against canonical DNA-binding helix overlays to identify critical binding interfaces and operator-specific contacts. For structural conservation analysis, the AlphaFold-generated model of Alatato gp31 was superimposed with the predicted model of Bridgette gp31 using the structural alignment tools in ChimeraX. This generated a numerical Root-Mean-Square

deviation (RMSD) value to quantify the structural similarity between the two proteins despite their classification in different phage clusters.

2.4. Mutagenesis

To validate interface predictions, in silico point mutations were introduced using ChimeraX's rotamer tool (Dunbrack 2010 library). Arginine 51 and Glutamic Acid 72 of gp31 were sequentially and simultaneously mutated to Alanine (R51A/E72A) to simulate a loss-of-function phenotype. Mutated sequences were re-run through AlphaFold 3 under identical parameters. Destabilizing mutations were evaluated based on the reduction of the interface predicted Template Modeling score (ipTM) and the loss of off-diagonal interaction confidence within the Predicted Aligned Error (PAE) matrix. To ensure the reliability of the observed destabilization, mutated sequences were processed across multiple prediction seeds (n=3) to confirm that the reduction in binding confidence was consistent and not an artifact of a single stochastic run.

3. RESULTS

3.1. Structural Prediction of the Wild-Type Alatato Repressor-Operator Complex

To investigate the structural basis of viral lysogeny in the *Arthrobacter* phage Alatato (Cluster FB), the putative immunity cassette protein gp31 was modeled alongside a 40-bp target operator DNA sequence utilizing AlphaFold 3. The algorithm successfully generated a multimeric assembly, yielding an Interface predicted Template Modeling (ipTM) score of 0.48. While an ipTM between 0.4 and 0.6 indicates moderate or uncertain confidence, the model suggests a potential interaction between the protein and DNA. High-resolution 3D visualization using UCSF ChimeraX suggested that the predicted repressor complex utilizes a canonical Helix-Turn-Helix (HTH) motif to predictively engage the major groove of the operator DNA.

Coulombic electrostatic surface mapping and atomic distance measurements ($\leq$ 4.0 Å threshold) identified a dense network of potential hydrogen bonds and electrostatic contacts. Specifically, Arginine 51 (R51) and Glutamic Acid 72 (E72) were identified as candidate interface residues interacting with the negatively charged DNA phosphate backbone (Figure 1); this finding is highly consistent with known biological DNA-binding mechanisms.

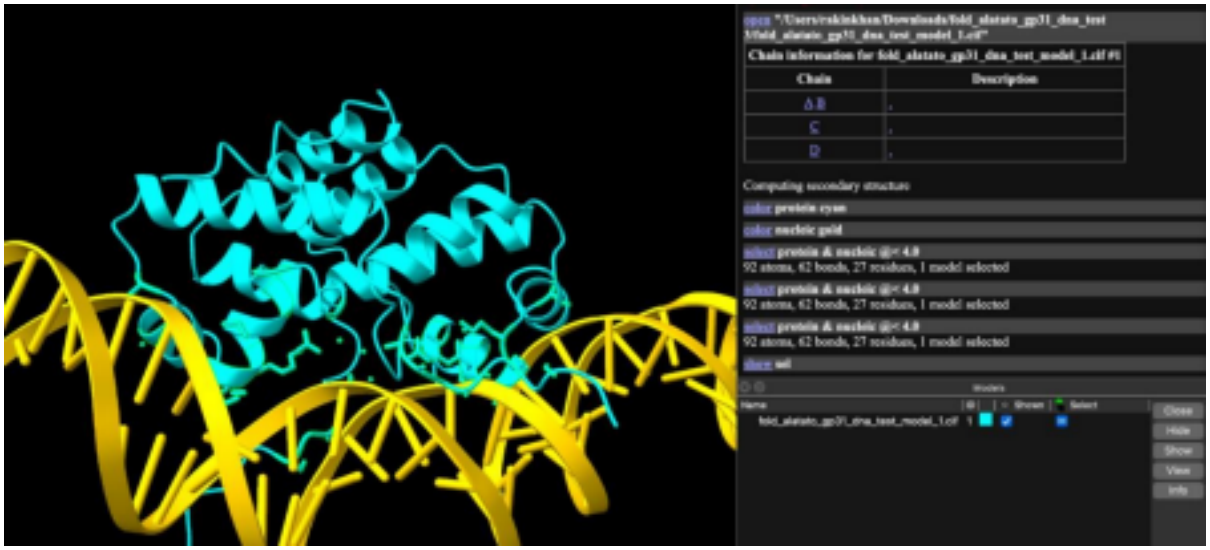

Figure 1. *In silico* structural prediction of the wild-type *Arthrobacter* phage Alatato gp31 repressor bound to the 40-bp operator DNA. High-resolution UCSF ChimeraX visualization demonstrates the repressor protein (cyan) utilizing a canonical Helix-Turn Helix (HTH) motif to engage the major groove of the operator DNA (gold). Specific interface residues within a 4.0 Å binding threshold are highlighted, illustrating possible contacts that stabilize the complex together. These computational interactions represent predicted spatial proximities rather than experimentally confirmed bonds.

3.2. Validation of DNA-Binding Interface via In Silico Mutagenesis

Following the identification of the HTH major groove contacts, in silico point mutations were generated to validate the dependency of these specific electrostatic anchors. An artificial loss-of-function variant (R51A/E72A) was processed through the identical AlphaFold 3 pipeline. The structural prediction of the double mutant demonstrated a reduction in predicted confidence. This was visually corroborated by the Predicted Aligned Error (PAE) matrix, which exhibited a near-complete loss of off-diagonal dark green regions compared to the wild-type baseline (Figure 2). Rather than definitive proof of physical unbinding, this shift in the PAE plot suggests a loss of consistent relative orientation compatible with interaction. The reduction in predicted confidence following mutation suggests that R51 and E72 may contribute significantly to stabilizing the protein-operator interface.

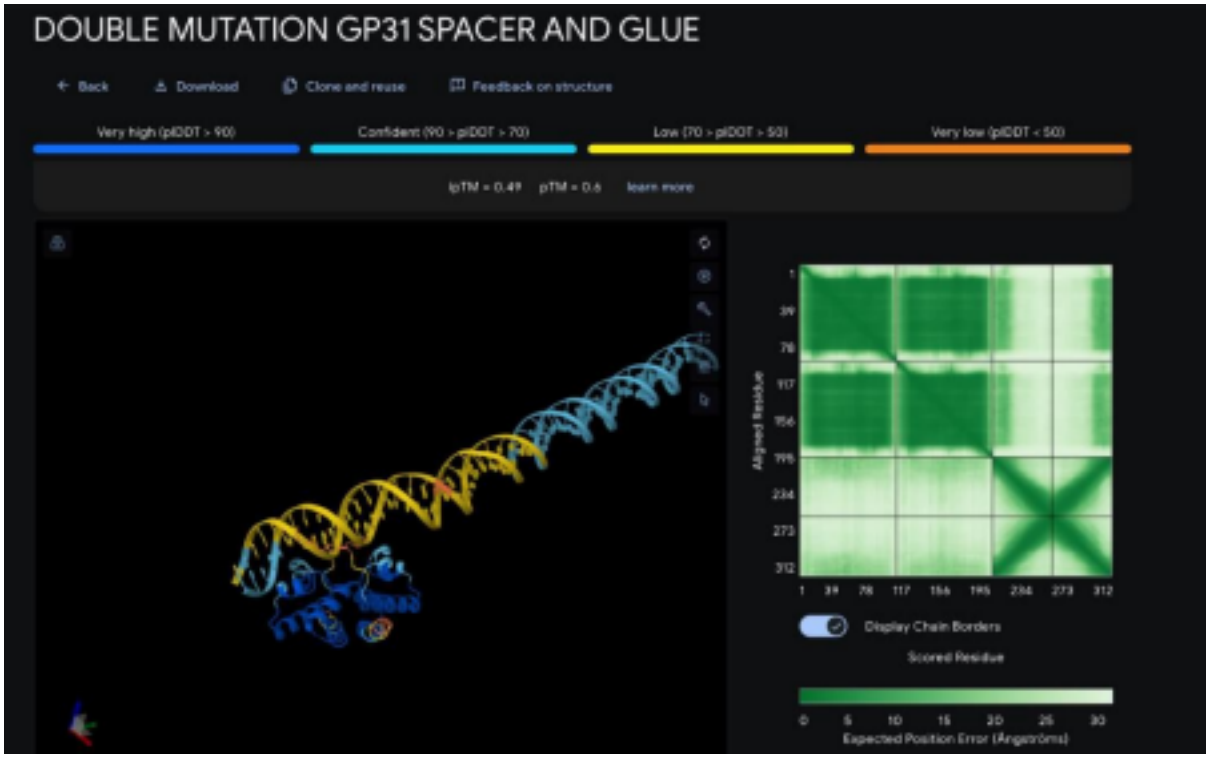

Figure 2. Destabilization of the repressor-operator complex following in silico double mutagenesis. AlphaFold 3 multimer prediction yields a weakened ipTM score of 0.49. The corresponding Predicted Aligned Error (PAE) plot (right) indicates a severe reduction in off-diagonal positioning confidence between the protein and DNA chains compared to wild-type baselines, suggesting a disruption in stable relative orientation. Off-diagonal regions in the PAE plots represent the model's confidence in the relative distance between different chains; a loss of signal here indicates high uncertainty in the geometry of the protein-DNA interface.

3.3. Structural Conservation Across Divergent Phage Clusters

To determine the evolutionary conservation of this specific repressor architecture, the Alatato gp31 structure was compared with its homolog from the genetically divergent Arthrobacter phage Bridgette (Cluster FA, Pham 289627). Despite significant nucleotide divergence between the two viral clusters, structural superimposition in UCSF ChimeraX yielded an exceptionally low Root-Mean-Square-Deviation (RMSD) of 0.556 Å (Figure 3).

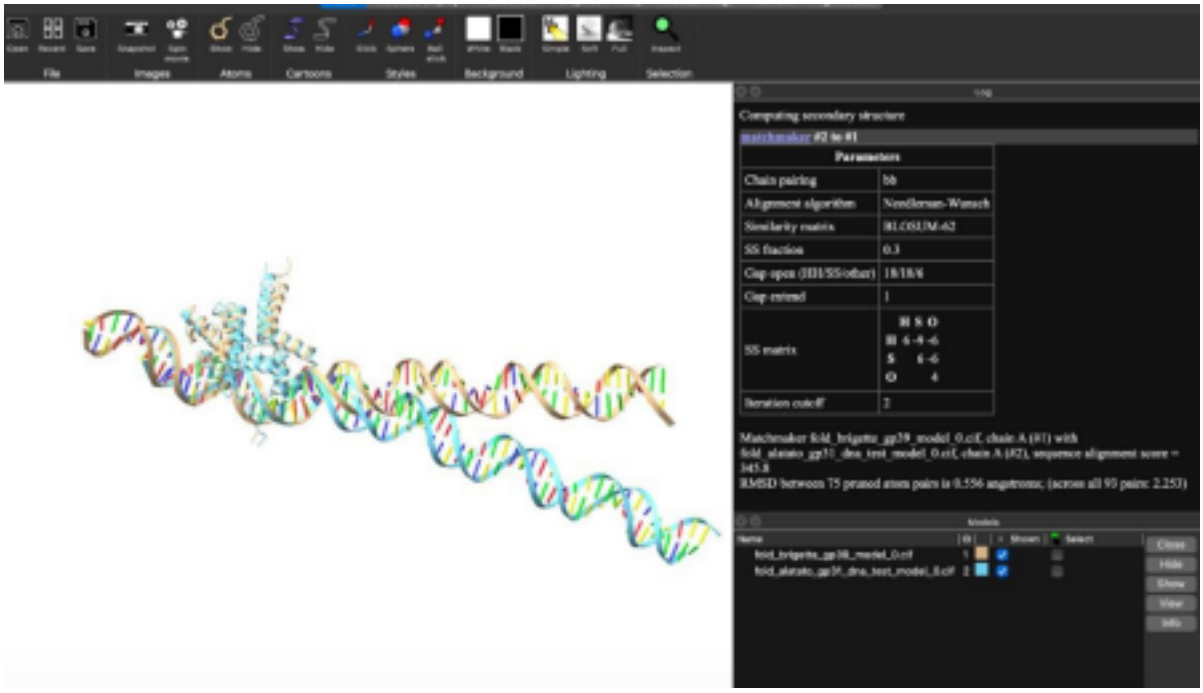

Figure 3. Structural conservation of the predicted HTH DNA-binding motif across divergent viral clusters. Superimposition of the Alatato gp31 repressor (cyan) and the Bridgette homolog (tan) via UCSF ChimeraX yields a near-perfect structural alignment with a Root-Mean-Square-Deviation (RMSD) of 0.556 Å.

This high degree of structural alignment strongly supports the hypothesis that the predicted R51-mediated HTH binding interface is strictly conserved across divergent phage clusters, highlighting it as a highly successful evolutionary strategy for maintaining lysogeny.

3.4. Functional Divergence and Repressor Incompatibility

Finally, to evaluate the functional specificity of this genetic switch, cross-reactivity was modeled using the Alatato operator DNA and the repressor protein from a non-homologous sibling phage, Adolin (Cluster AZ1, Pham 291150). The AlphaFold 3 multimer prediction exhibited a complete lack of predictive binding affinity, generating a failing ipTM score of 0.22. Additionally, the PAE plot displayed massive off-diagonal blank regions, indicating the algorithm's inability to identify a consistent relative orientation between the heterologous protein and DNA (Figure 4).

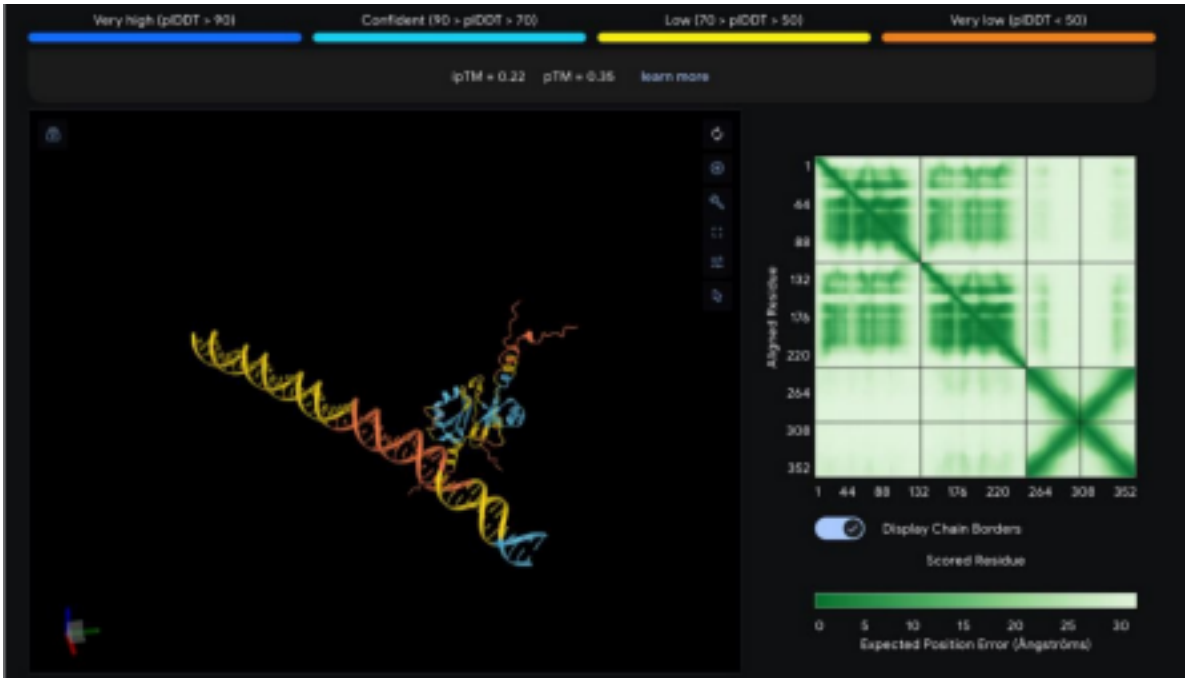

Figure 4. Functional divergence and repressor incompatibility. AlphaFold 3 PAE plot of the non-homologous Adolin repressor (Cluster AZ1) simulated with the Alatato operator DNA. The complete absence of off-diagonal interaction confidence reflects a lack of predictive binding affinity (ipTM = 0.22), indicating that the algorithm could not confidently model a stable structural relationship between the specific repressor and operator.

These in silico findings suggest that while the basic HTH mechanism is conserved, genetically divergent phages utilize incompatible, sequence-specific regulatory switches to prevent transcriptional cross-talk.

4. DISCUSSION

Application of AlphaFold 3 to the Arthrobacter phage Alatato immunity cassette provided structural models of repressor-operator interactions. While an ipTM baseline of 0.48 indicates moderate confidence rather than definitive proof, the coherent off-diagonal PAE interactions suggest a consistent relative orientation compatible with functional binding. These results highlight the potential of AlphaFold-Multimer as a scalable tool for hypothesizing viral regulatory mechanisms to guide future experimental assays.

A central finding of this work is the identification of Arginine 51 (R51) in gp31 as a candidate anchor residue. Structural visualization in ChimeraX suggested that R51 is positioned to mediate electrostatic interactions with the DNA phosphate backbone, a finding that is highly consistent with established biological DNA-binding mechanisms. In silico mutagenesis supported its importance: substitution of R51 with alanine (R51A) reduced predicted structural confidence and disrupted favorable relative orientation in the PAE plots. This suggests that phage repressors rely on conserved electrostatic anchors and that R51 and E72 may contribute significantly to stabilizing the protein-DNA interface.

Comparative structural analysis further contextualizes these findings within viral evolution. The near-perfect alignment between Alatato gp31 and its homolog from the divergent Cluster FA phage Bridgette (RMSD = 0.556 Å) suggests that the R51-mediated HTH interface has been strictly conserved across clusters. This conservation reflects strong evolutionary pressure to preserve DNA-binding functionality as a central mechanism of lysogeny. In contrast, the lack of predictive binding affinity observed in the Cluster AZ1 phage Adolin (ipTM = 0.22) highlights functional divergence, suggesting that distinct phage families have evolved incompatible regulatory switches to prevent transcriptional cross-talk. Together, these results illustrate the adaptability of phage immunity cassettes: while core residues such as R51 remain conserved, surrounding sequence variation enables fine-tuning of DNA recognition specificity.

4.1. Comparative Analysis and Mechanistic Implications

While R51 appears to be a critical electrostatic anchor, it likely functions within a broader network of interactions. It is a possibility that neighboring residues, potentially including the identified E72 or other nearby polar amino acids, contribute to the initial scanning of the DNA, while R51 provides the final stabilizing lock. This model aligns with established paradigms in viral lysogeny, most notably the Phage Lambda CI repressor. In the Lambda system, specific Arginine residues are known to probe the major groove of the operator DNA, utilizing positive charges to navigate the negatively charged phosphate backbone (Ptashne, 2004).

Furthermore, to address whether the R51A mutation caused a total structural collapse of the protein, we monitored the pLDDT (Predicted Local Distance Difference Test) scores. The core fold of the gp31 mutant maintained high confidence (pLDDT > 80), suggesting that the observed drop in binding confidence (ipTM) reflects a loss of interface specificity rather than a general structural instability. This supports the hypothesis that R51 is a functional residue specifically evolved for DNA engagement.

Finally, the lack of affinity between Alatato and Cluster AZ1 phages (Adolin) suggests that sequence-level variations at these interfaces act as "orthogonality filters." These filters likely prevent cross-interference between different phages infecting the same host, a common evolutionary strategy to maintain distinct regulatory control.

4.2. Limitations of In Silico Modeling

All conclusions in this study are derived from in silico structural predictions generated by AlphaFold 3. While these algorithms provide sophisticated biophysical approximations, we recognize that our primary binding model yielded a moderate ipTM score (~0.48). Although a score of 0.6 is often cited as a threshold for high-quality global modeling, we utilized this value as a comparative metric; it represents a significant increase in confidence over the non-binding control (ipTM 0.22), suggesting a specific interaction rather than a stochastic (random) fit.

To further assess the model's reliability, we analyzed Predicted Aligned Error (PAE) patterns. The low relative error observed at the HTH-major groove interface indicates a stable and consistent local binding geometry, even if the global confidence is moderate. Furthermore, readers should consider potential AlphaFold 3 biases, such as "PDB memorization"—where the model may favor well-known Helix-Turn-Helix (HTH) motifs from its training data—and its tendency to model DNA backbones with less flexibility than is seen in nature.

Consequently, the structural models and specific residue interactions proposed, such as the Arginine 51 (R51) electrostatic anchor, should be interpreted strictly as hypotheses. These findings provide a predictive framework that requires future experimental validation through biochemical techniques, such as Electrophoretic Mobility Shift Assays (EMSA) or in vitro mutagenesis.

5. CONCLUSION

This study presents a completely *in silico* approach to mapping the regulatory systems of viral genomes. By modeling the *Arthrobacter* phage Alatato immunity cassette with AlphaFold 3 and UCSF ChimeraX, we characterized gp31 as a predicted DNA-binding repressor. Our structural analysis suggests a conserved Helix-Turn-Helix motif and identifies Arginine 51 as a potential electrostatic contributor for operator binding. Furthermore, cross-cluster comparisons suggest that while the core structural mechanics of this binding interface remain conserved among evolutionary homologs, sequence-level differences may contribute to functional divergence that prevent cross-reactivity with non-homologous phages. Ultimately, these computational methods offer a practical way to generate testable hypotheses regarding repressor-operator dynamics, providing a predictive framework for the structural basis of viral lysogeny that invites future experimental validation.

DECLARATIONS

Funding: The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests: The authors declare they have no financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

REFERENCES

1. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583– 589.

2. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., ... & Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. Science, 373(6557), 871–876.

3. Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., & Baker, D. (2020). Improved protein structure prediction using predicted interresidue orientations. Proceedings of the National Academy of Sciences, 117(3), 1496–1503.

4. Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., ... & Hassabis, D. (2020). Improved protein structure prediction using potentials from deep learning. Nature, 577(7792), 706–710.

5. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., ... & Hassabis, D. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature, 630(8016), 493–500.

6. Ptashne, M. (2004). A Genetic Switch: Phage Lambda and Higher Organisms (3rd ed.). Cold Spring Harbor Laboratory Press.

7. Oppenheim, A. B., Kobiler, O., Stavans, J., Court, D. L., & Adhya, S. (2005). Switches in bacteriophage lambda development. Annual Review of Genetics, 39, 409–429.

8. Hatfull, G. F. (2015). Dark matter of the biosphere: the amazing world of bacteriophage diversity. Journal of Virology, 89(16), 8107–8110.

9. Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., & Moult, J. (2021). Critical assessment of methods of protein structure prediction (CASP)—Round XIV. Proteins: Structure, Function, and Bioinformatics, 89(12), 1607–1617.

10. Russell, D. A., Hatfull, G. F., & Bowman, C. A. (2017). PhagesDB: the actinobacteriophage database. Bioinformatics, 33(5), 784–786.

11. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Science, 30(1), 70–82.