Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Mycobacteriophage Cluster A Immunity Repressors: A Conserved Protein Fold Adapted for Diverse Biological Functions

Jamie Wallen

jamiewallen@wcu.edu

Western Carolina University https://orcid.org/0000-0003-3325-9177

Regan McKinney

Western Carolina University

Serim Park

Western Carolina University

Hannah Bonogafsky

Western Carolina University

Ashley Tuhro

Western Carolina University

Malcolm Turner

Western Carolina University

Thomas Hollis

Wake Forest University School of Medicine

Maria Gainey

https://orcid.org/0000-0001-8853-2092

Article

Keywords:

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10844716/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Mycobacteriophage Cluster A Immunity Repressors: A Conserved Protein Fold Adapted for Diverse Biological Functions

Regan McKinney[1], Serim Park[1], Hannah Bonogafsky[1], Ashley Tuhro[1], Malcolm Turner[1], Thomas J. Hollis[2], Maria D. Gainey[1], and Jamie R. Wallen[1*]

[1] Western Carolina University, Department of Chemistry and Physics, 111 Memorial Drive, Cullowhee, North Carolina, 28723

[2] Department of Biochemistry, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157, USA

[*]Corresponding Author Email Address:
jamiewallen@email.wcu.edu

Abstract

Mycobacteriophage cluster A immunity repressors are unusual in that they silence lytic gene expression by binding numerous DNA binding sites located throughout the viral genome. They can also provide immunity to infection by genetically similar phages that contain the same or similar DNA binding sites. Despite containing similar DNA binding sites and repressor proteins, cluster A viruses often display heteroimmune phenotypes, the molecular details of which are currently unclear. Using a combination of small-angle X-ray scattering, protein crystallography, bioinformatics, and DNA binding and *in vivo* immunity assays, we show that cluster A repressors conserve a common protein fold yet appear to provide DNA specificity through a narrow region of the novel stoperator DNA binding domain. Additionally, both viral and bacterial genomes contain cluster A-like repressor proteins fused to diverse protein domains, which suggest that these repressors have evolved to do more than just silence viral lytic gene expression.

Introduction

Temperate bacteriophages are viruses that can choose between the lytic or lysogenic replication cycle during infection of a bacterial host[1]. During a lytic infection the phage will replicate its genome, produce new virions, and lyse the host cell to promote phage propagation. In the lysogenic cycle, rather than lyse the cell, the phage's genome will either be integrated into the host chromosome or maintained extra-chromosomally, and the phage can lie dormant in the host until conditions are favorable to switch to the lytic cycle[2]. Lysogeny is maintained by a repressor protein that binds the prophage genome to silence the transcription of lytic genes[3]. The repressor can also provide superinfection immunity by binding to the genomes of other genetically similar bacteriophages to prevent a secondary phage infection[4]. A variety of immunity repressors have been characterized to date, and studies have shown that they utilize different oligomeric states to bind their respective DNA sequences. Recent studies have shown that removal of repressor genes from mycobacteriophage genomes make them more efficient as potential therapeutics for the treatment of antibiotic-resistant bacterial infections[5,6], so further work is needed to better understand how repressor systems function in actinobacteriophages.

The best characterized repressor is the cI protein from bacteriophage lambda, which as a dimer binds an operator sequence using helix-turn-helix (HTH) DNA binding domains[3,7]. Actinobacteriophage repressor systems are intriguing because they are distinct from the well-studied phage lambda repressor[8-10]. As an example, studies on repressors from several cluster A mycobacteriophages have shown that these repressors bind an ~13 bp asymmetric DNA sequence as a monomer[4,11,12]. Repressor DNA binding sites are abundant and located throughout cluster A phage genomes, and they are classified as either operator (prevent transcription initiation) or stoperator (halt transcription elongation) sites depending on if they are located in

promoter regions or intergenic spaces, respectively[11]. We recently determined the X-ray crystal structure of the repressor protein from TipsytheTRex, a cluster A2 mycobacteriophage that contains 26 operator/stoperator sites in its genome, bound to its DNA recognition sequence[13]. The structure confirmed that the protein engages DNA as a monomer, and it does so via two independent DNA binding domains that bind successive major grooves of DNA. The N-terminus of the protein is an HTH domain, while the C-terminal region contains a novel protein fold that was termed the stoperator domain. The TipsytheTRex repressor binds DNA with nanomolar affinity, and both DNA binding and superinfection immunity assays helped identify key residues that are essential for DNA recognition[13].

Prior to our crystal structure of the TipsytheTRex repressor, Mavrich *et al*. showed that while cluster A bacteriophages conserve the presence of many ~13 bp operator/stoperator sites throughout their genomes, these viruses contain different consensus repressor DNA recognition motifs and display complex immunity phenotypes[4]. Specifically, while the beginning of the DNA binding motif is fairly well conserved across cluster A phages (motif positions 1-5 in Figure 2 of Mavrich *et al.*[4]), the end of the motif (positions 9-13) shows sequence variability and likely accounts for the heteroimmunity phenotypes. Our structure of the TipsytheTRex repressor revealed that the conserved portion of the motif is engaged by the HTH, while the variable region is bound by the stoperator domain. Cluster A1 and A2 mycobacteriophages are heteroimmune[12], and an amino acid alignment of the TipsytheTRex repressor with the repressor from the cluster A1 phage Adahisdi (the focus of this report) shows that their proteins share 44% sequence identity. Given the observed differences in the DNA consensus motif in cluster A1 and A2 phages[12], we sought to determine the crystal structure of a cluster A1 mycobacteriophage repressor bound to DNA to see if the overall protein fold of the stoperator domain is conserved,

and to identify residues that contribute to DNA specificity. Here we present the X-ray crystal structure of the Adahisdi repressor bound to DNA and show that the HTH and stoperator domain folds are indeed conserved. Additionally, small-angle X-ray scattering (SAXS) studies show that in solution the A1 repressor is highly flexible in the absence of DNA, but in the presence of DNA takes on a shape that agrees well with the crystal structure. We also present AlphaFold[14,15] models of other cluster A repressors on DNA and observe that while the overall fold of the protein is conserved across all A subclusters, their amino acid makeup differs substantially in two critical areas that appear to drive DNA substrate specificity and surface charge distribution. Finally, using BLAST[16] we have identified novel proteins in both phage and bacterial genomes that contain a cluster A-like repressor fused to a variety of protein domains of both predicted and unknown functions, which suggests that the cluster A repressor fold has been adopted to perform additional functions other than transcriptional silencing.

Results and Discussion

Adahisdi Immunity Repressor Structure

The crystal structure of the Adahisdi repressor:DNA complex was determined by molecular replacement to a resolution of 2.30 Å, using an AlphaFold[15] model of the Adahisdi repressor as the search model. The Adahisdi structure reveals a monomer bound to double-stranded DNA, and the crystal contains four molecules of the repressor:DNA complex in the asymmetric unit (ASU). The four repressor:DNA complexes in the ASU show little conformational differences when superimposed (rmsd: 0.189-0.216 Å across 162 residues). A summary of data collection and refinement statistics are provided in Table 1.

The Adahisdi repressor engages DNA using the HTH and stoperator DNA binding domains (Figure 1A), and a superposition of the Adahisdi (chain A) and TipsytheTRex repressor

structures using the program Pymol (version 3.1.6.1) shows that they superimpose very well (rmsd: 0.970 Å across 139 residues, Figure 1B). An electrostatic surface potential rendering reveals important differences between the two structures. First, the Adahisdi repressor (Figure 1C) appears to have more of a positive potential in the DNA binding region of the stoperator domain as compared to TipsytheTRex (Figure 1D). In the TipsytheTRex:DNA complex there is a total buried surface area of 1702 $Å^2$, while the four Adahsdi:DNA complexes in the ASU show a greater buried surface area (chain A: 1900 $Å^2$, chain B: 1906 $Å^2$, chain C: 1957 $Å^2$, and chain D: 2039 $Å^2$). The fact that the Adahisdi repressor engages DNA more extensively agrees with our observation that when the Adahisdi repressor was purified under the same conditions as the TipsytheTRex repressor (0.5 M NaCl)[13], the resulting spectra indicated that contaminating DNA co-purified with the protein. Pure Adahisdi repressor devoid of DNA was obtained when the sodium chloride in the purification buffers was increased to 1M. Secondly, in our report on the TipsytheTRex structure we noted a substantial negative potential at the very C-terminal end of the protein that does not engage DNA, and we hypothesized that this region may play some role in protein:protein interactions in the host cell[13]. A direct comparison (Figures 1C and 1D) shows that the Adahisdi repressor lacks this strong negative potential at its C-terminus, and if this region is indeed used for protein:protein interactions, then these results indicate that despite conserving the same overall fold, these immunity repressors use this region to target different partner proteins.

Repressor Conformational Dynamics in Solution

We have been unable to crystallize a cluster A repressor in the absence of DNA, and our previous molecular dynamics work suggested that in the absence of DNA the repressor is flexible in solution[13]. The repressor shows limited solubility in a Tris/NaCl buffer, especially in

the absence of DNA, but we have found the protein is more soluble in phosphate buffer. To better understand the conformational dynamics of the A1 repressor in solution, we used small-angle X-ray scattering coupled with multi-angle light scattering in line with size-exclusion chromatography (SEC–SAXS–MALS)[17] to observe the behavior of the repressor in the presence and absence of DNA (Figure 2). In the absence of DNA, the repressor conformation in solution differs from what is observed in the crystal structure, and the SAXS data is best fit with an ensemble of models in which the HTH domain samples conformational space (Figures 2A-2C). The monomeric mass of the repressor is 22.2 kDa, and the mass calculated from MALS (30.8 kDa) is consistent with a monomer in solution. Once bound to DNA, the repressor conformation in solution matches what is observed in the crystal structure, with the only notable movement being the N-terminal His tag (Figure 2D-2F). The mass of the protein:DNA complex calculated by MALS (36.0 kDa) agrees well with the expected mass of the monomeric complex (33.9 kDa), which further supports that the cluster A1 repressor is a monomer in solution. Overall, the SEC-SAXS-MALS data shows that in the absence of DNA the repressor is flexible in solution but takes on a more rigid structure once engaged with DNA.

Structural Description for Heteroimmunity between Cluster A1 and A2 Mycobacteriophages

It is well-documented that cluster A1 and A2 mycobacteriophages are heteroimmune[4,11,12], meaning that repressors of one phage do not repress replication of the other species. An analysis of the Adahisdi genome using the MEME suite[18] identifies 37 operator/stoperator sites that are spread throughout the Adahisdi genome, with the consensus motif shown in Figure 3A. C3, T4, A13, and C15 of the motif are absolutely conserved in all 37 sequences, and this motif is similar, yet distinct, from the consensus motif identified from the 26 operator/stoperator sequences in the TipsytheTRex genome[13] (Figure 3B). Specifically, the

portion of the motif engaged by the HTH domain (CTTGAC) is strongly conserved, while the region bound by the stoperator domain (CGTAAC in Adahisdi, CCACC in TipsytheTRex) differs substantially. This supports the hypothesis that the DNA binding specificity and previously observed complex immunity phenotypes[4] across cluster A mycobacteriophages is likely driven by DNA recognition in the stoperator domain.

A DNA binding analysis of the Adahisdi and TipsytheTRex structures using DNAproDB[19] are shown in Figures 3A and 3B, respectively. In the Adahisdi structure, Q38, W42, E67, Q71, K76, and R99 all directly interact with DNA bases in the consensus and likely drive sequence recognition. Q38, W42, and R99 are all conserved in TipsytheTRex (as Q46, W50, and R108, respectively), and while mutation of Q46 in TipsytheTRex had no effect on DNA recognition, both W50 and R108 proved to be essential for function[13]. R45 in TipsytheTRex plays an essential role in DNA recognition[13], as it binds a conserved G opposite C7 of the consensus shown in Figure 3B. The equivalent residue in Adahsidi is P37, and we observe variability at the Adahsdi DNA consensus position near P37 (position 8 of the DNA consensus in Figure 3A). As illustrated in the alignment provided in Figure 3C, cluster A1 repressors conserve a proline at this position (highlighted in yellow), while an arginine is at this position for cluster A2 repressors. Additionally, Figure 3C shows that E67, Q71, and K76 of the Adahisdi stoperator domain are conserved in other A1 repressor sequences and differ from the equivalent residues in A2 sequences, which suggests that these residues likely play a role in DNA sequence specificity.

To more extensively investigate cluster A repressor protein sequence and fold conservation, we aligned representative sequences from each of the current A subclusters and generated AlphaFold models of each repressor on its DNA consensus sequence (Figure 4). To

identify each subcluster DNA consensus motif, we used MEME to analyze the genome of each phage containing the respective repressor protein, and Supplemental Figure 1 provides both the consensus sequence for each subcluster as well as the number of DNA binding sites present in each phage genome. As shown in Figure 4, the AlphaFold models of all A subclusters superimpose well with the A1 crystal structure (shown in gray) and contain excellent ipTM and pTM values (see Supplemental Figure 2). This provides strong evidence that the overall fold of the repressor is conserved across all A subclusters, but it is important to note that while the predicted AlphaFold models appear to be highly accurate, they do not account for DNA specificity. As an example, AlphaFold can generate models of the Adahisdi repressor on both an A1 and A2 DNA consensus sequence, but DNA binding assays show that the A1 repressor fails to bind an A2 DNA sequence (see reference[12] and below). For these reasons we see the A1 crystal structure as a critical component to accurately understand DNA sequence specificity in these repressors.

In the alignment provided in Figure 4, only the A1 subcluster contains a proline at position 37 (Adahisdi numbering); all other subclusters contain a positively charged residue (arginine or lysine) at this position. In looking at the DNA consensus motifs for each subcluster (Supplemental Figure 1), the CTTGAC motif that is recognized by the HTH domain is highly conserved, with the one exception being cluster A19 (Kimona). The sequence alignment in Figure 4 correlates with this observation and shows that DNA binding residues in the HTH are well conserved, as indicated with residues either in bold red (absolute conservation) or red lettering (conserved). Figure 4 reveals two key regions of the repressor that lack conservation across different A subclusters. First, the beginning of the stoperator domain (indicated by a purple line in Figure 4) shows strong sequence variability, and this is the portion of the domain

that inserts into the major groove of DNA and recognizes the portion of the consensus that varies across subclusters. This is the same region that was highlighted in Figure 3 as a key difference between cluster A1 and A2 repressors, and given the variability this appears to be a key region of determining DNA specificity across subclusters. The second variable region is at the C-terminal end of the proteins. There is strong conservation of a WX[F/W]P motif that begins at residue 163 in Adahisdi, followed by variable sequence. In comparing the Adahisdi and TipsytheTRex sequences, the TipsytheTRex sequence after the WX[F/W]P motif is both longer and more acidic, which accounts for the acidic surface patch on the A2 structure highlighted in Figure 1D. In comparing the subcluster A sequences, there is variability in both the length and charge of this C-terminal region, and electrostatic surface potential maps generated from either the crystal structure (A1 and A2) or AlphaFold models (all other A subclusters) reveal substantial surface charge distribution across these repressors that agree well with the amino acid makeup of the C-terminus (Supplemental figure 2). Based on these results, while the overall fold of these proteins is well conserved, the proteins vary greatly in their amino acid makeup and charge distribution, and these variations likely contribute to heteroimmune phenotypes.

Analysis of Protein:DNA interactions

Given the key contacts observed between E67, Q71, and K76 with DNA, and that these residues are found in the variable region identified in Figure 4, we expected to see that the mutants E67A, Q71A, and K76A would result in an A1 repressor with severely compromised DNA binding. Additionally, since R99 in Adahisdi is equivalent to the essential R108 in TipsytheTRex, we anticipated that an R99A mutant would result in a complete loss of DNA binding. As predicted based on previous reports[4,11,12], the wild-type Adahisdi repressor shows strong DNA sequence specificity, as it recognizes an A1 DNA sequence with low nM affinity

(51 ± 18 nM) but does not bind a cluster A2 DNA sequence (Figure 5A). To our surprise, point mutations of Adahisdi residues predicted to be essential for DNA recognition show binding properties that are comparable to wild-type (Figure 5A). These results are in stark contrast to the TipsytheTRex repressor where key point mutations resulted in a complete loss of DNA binding *in vitro*[13]. Both the greater positive potential in the stoperator DNA binding domain (Figure 1C) and the enhanced DNA buried surface (as compared to TipsytheTRex) could account for the fact that a single point mutation is not sufficient to generate an A1 repressor defective in DNA binding in our assay.

In contrast to the results in Figure 5A, superinfection immunity assays show that mutation of residues that contact DNA led to a substantial loss of repressor function *in vivo* (Figure 5B). In the presence of empty pMH94 vector, the wild-type Adahisdi virus can efficiently grow on *Mycobacterium smegmatis* mc2155 cells, with a titer of 1.3 x 10^{10} pfu/mL. When the wild-type repressor gene plus its endogenous promoter is placed in pMH94, Adahisi virus fails to propagate on *M. smegmatis*, with an 11-log reduction in the efficiency of plating (EOP) as compared to the empty vector control. In previous work with the TipsytheTRex repressor, virus was able to grow even in the presence of wild-type repressor[13]; the results presented here show that the Adahisdi repressor is much more efficient at superinfection immunity as compared to the cluster A2 repressor. Of all the mutants tested in Figure 5B, only Q38A showed no observable phenotype as compared to wild-type, and this result agrees with the A2 report in which the Q46A mutation in TipsytheTRex also showed minimal differences as compared to wild-type[13]. Tryptophan 42 in Adahisdi plays the equivalent role of tryptophan 50 in TipsytheTRex[13] and stacks against a DNA base, and the W42A mutant shows an ~7-log reduction in EOP as compared to wild-type. The mutants E67A, Q71A, K76A, and R99A all

show an ~6-log reduction in EOP as compared to the wild-type repressor. Overall, these results show that mutation of residues that directly contact DNA bases in the consensus cause a substantial reduction of repressor function *in vivo*.

As described above (Figure 4), cluster A1 repressors contain a proline at position 37 (Adahisdi numbering). To understand the significance of the proline at this position, we generated both P37A and P37R mutant proteins, the latter of which is designed to mimic the positively charged residue (either an arginine or lysine) present at the equivalent position in all other A subclusters. The results of Figure 5B show an ~5-6 log reduction in EOP for these mutants as compared to wild-type, and they generate smaller plaques as compared to other mutants in Figure 5B. Despite these results, the P37R mutant still binds DNA with an affinity comparable to wild-type repressor (Figure 5A). Based on the results of Figure 5B, we can conclude that the presence of a proline at position 37 of Adahisdi is important for cluster A1 repressor function.

Cluster A-like Repressors in Other Actinobacteripohage Genomes

The Adahisdi repressor belongs to a phamily[20] that contains 902 sequences (www.phagesdb.org, downloaded on 06/08/2026)[21], and along with being present in cluster A mycobacteriophages, these repressors are also observed in clusters F (44 sequences), CA (23 sequences), C (21 sequences), J (9 sequences), K (3 sequences), and two singletons. The presence of cluster A-like repressors in other phage clusters has been described previously[22,23], and of the additional clusters, a MEME[18] search can only find the repetitive operator/stoperator sites in cluster CA genomes (Supplemental Figure 1). We generated a phylogenetic tree of representative cluster A repressors, as well as repressors from clusters F, CA, C, J, K, and two singletons using the program UGENE[24] to better understand the evolutionarily relationships of

these proteins. Figure 6 shows the resulting tree, and we see several interesting observations. For cluster A repressors, we see trends that suggest subclusters sharing common ancestors, such as clusters A9 and A16 (cyan branch), A14 and A22 (dark purple branch), A7 and A19 (blue branch), A8 and A18 (blue branch), A2 and A17 (brown branch), and A11 and A12 (brown branch). For the non-A repressor sequences, the cluster K and CA repressors, as well as repressors from two singleton phages, appear to share a common ancestor with cluster A5 repressors (light purple branch). In agreement with this observation, the Lillie (cluster CA) and Aragog (A5) genomes share a similar operator/stoperator consensus motif (Supplemental Figure 1). Cluster C1 and F repressors cluster together (red branch), while cluster J repressors form their own node on the tree (black branch). Cluster C1, F, and J A-like repressors all contain a proline at the equivalent position of P37 in Adahisdi, while the K4, K7, CA, and one singleton phage (Ibantik) conserve a positively charged residue at this position (Supplemental Figure 3). For the phages that appear to lack repetitive operator/stoperator sites (clusters C1, F, J, and K), C1 phages lack an integration cassette and are lytic[8], while clusters F1, F2, J, K4, and K7 contain a cluster A-like repressor as well as a second, cluster-specific immunity repressor in their genomes. As an example, in the cluster F1 mycobacteriophage Fruitloop, gene 37 codes for a cluster-A like repressor that is 98% identical to Adahisdi's repressor, while gene 44 encodes a predicted immunity repressor protein found in cluster F1 and F2 bacteriophages[25]. One hypothesis is that these viruses have acquired a cluster-A like repressor to provide immunity to cluster A phages, and two recent reports support this hypothesis: the cluster F1 phage Mcgavigan that contains a cluster A1-like repressor that provides immunity to cluster A1 phage superinfection[22], and the A-like repressor in F1 phage Fruitloop that confers immunity to some cluster A1 phages[9]. Based on our results provided in Figure 6 and Supplemental Figure 3, we

predict that cluster K phages may have developed immunity against cluster A5 phages, while clusters C1, F2, and J, which like cluster F1 phages contain a proline at position 37, likely have immunity to cluster A1 phages.

An intriguing cluster of phages are cluster BD that infect *Streptomyces* hosts. These have been identified as temperate due to the presence of a predicted serine integrase and an immunity repressor, and although the repressor belongs to a different phamily containing 118 sequences, all present in cluster BD, their sequences are remarkably similar to cluster A repressors. Despite the similarity, MEME fails to find repetitive stoperator/operator sites in the cluster BD genomes. Figure 7A shows a sequence alignment of the Adahisdi repressor with a representative repressor from each BD subcluster, and there are several observations to note. First, the cluster BD phages do not conserve the DNA binding residues present in Adahisdi with the exception of W42 and R99; this lack of conservation strongly suggests these repressors recognize a DNA motif distinct from cluster A repressors. Second, cluster BD repressors have a truncated C-terminus and are missing both the WX[F/W]P motif and charged tail observe in cluster A repressors. Finally, cluster BD repressors are missing a proline rich region present in the stoperator domain of cluster A repressors, a sequence similar to a PXXP protein interaction motif[26,27]. The Adahisdi repressor contains the proline-rich sequence of 118-PNIPPIP-124, and in the cluster A repressor alignment in Figure 4, P118 and P122 are conserved in all cluster A repressors, while P121 and P124 are either conserved as prolines or are charged residues (either D, E, or K).

Despite lacking these two regions, an AlphaFold model of a cluster BD repressor shows it superimposes well on the Adahisdi crystal structure (Figure 7B). The missing components of the cluster BD repressor (colored in black in the Adahisdi structure and shown in Figures 7B-7D) are on the surface of the protein. W163 (the first W of the WX[F/Y]P motif) in Adahisdi is

anchored into a hydrophobic pocket formed by Y61, P62, L78, I143, L145, and I162, while the remaining portion of the C-terminal tail is solvent exposed and lacks secondary structure (Figures 7B and 7C). The proline-rich region present in Adahisdi but absent in cluster BD sequences is part of a surface exposed loop present just above the bound DNA (Figure 7D). While the function of this proline-rich region is unknown, its location is perfectly positioned to both sense a bound DNA molecule and interact with additional proteins, suggesting that this region may play some role in lysogeny regulation.

For the W163 interaction, L78 is found within the α6 helix of the stoperator domain that contacts the DNA major groove, so the observed hydrophobic pocket may be a structural element to ensure proper protein folding for DNA binding. To test this hypothesis we generated two mutants, a W163A point mutation as well as a H160stop mutant that would truncate the C-terminus to mimic cluster BD repressors. Figure 7E shows that both mutants are defective in repressor function *in vivo*. To correlate this loss of function with a defect in DNA binding, we attempted to express and purify both mutants for DNA binding studies. The H160stop mutant is not soluble, but we were able to purify the W163A Adahisdi repressor mutant. The DNA binding results in Figure 7F show a severe defect in DNA binding, as we only see a shift in DNA size at the highest protein concentrated tested. These results confirm that W163 and the resulting C-terminal tail are essential for repressor function, and W163 must play a critical role in proper protein folding. Due to these results, it is currently unclear if the cluster BD repressors are functional, and if they are, what role they play in cluster BD bacteriophages that lack the repetitive operator/stoperator sites.

Novel Functions for Cluster A-like Immunity Repressors

A BLAST[28] search using the Adahisdi repressor sequence as the query returns hundreds of sequences that are the typical repressor size of ~180 residues in length. However, we noticed that a handful of sequences identified in the BLAST search are much larger than a typical cluster A immunity repressor. Iterative rounds of PSI-BLAST identified 32 sequences from both bacteriophage and bacteria that contain a cluster A-like immunity repressor fused to other protein domains (Table 2). These proteins range in size from ~26 kDa to ~210 kDa and contain a variety of protein domains of both unknown and predicted function. The repressor domain is found at either the amino or carboxy terminus but never observed internal to the protein. Intriguingly, while in most cases it is the full-length repressor domain fused to other proteins, a handful of sequences contain just the stoperator domain.

Figure 8 shows AlphaFold models of several repressor fusion proteins and illustrates the diversity of observed protein folds. Figure 8A is a protein from *Mycobacterium sp.* ENV421 and contains a protein of unknown function at the N-terminus fused to a repressor domain. The repressor domain superimposes well with the Adahisdi repressor, and the AlphaFold model predicts that a long linker connects the two protein domains. The fusion from *Galactobacter sp*. (Figure 8B) provides an example where a domain of unknown function is fused to just the stoperator domain of the repressor. Without the HTH present, it is unclear whether this protein could bind DNA with high affinity. Predicted functions in the repressor fusions include serpins, RNA-helicase, pirins, and aminotransferases, and examples of these proteins are illustrated in Figures 8C-F. Serpins are serine protease inhibitors present in all kingdoms of life[29], and recent work showed that the activity of the prokaryotic serpin chloropin on thrombin is stimulated ~74-fold in the presence of supercoiled DNA and ~142-fold in the presence of linear DNA[30]. Serpin-repressor fusion sequences have been identified in *Mycolicibacterium mageritense*, *Gordonia*

hankookensis, and *Mycolicibacerium psychrotolerans* (Table 2, Figure 8C), and the repressor domain in these proteins may serve a similar role in that it binds DNA to stimulate the activity of the serpin domain. The AAA domain-containing protein shown in Figure 8D is a large, complex protein with a variety of predicted DNA metabolic functions. An HHPred[28,31] analysis reveals several potential functions for this protein including helicase, DNA binding protein SMUBP-2/hydrolase/helicase, mobile intron protein/homing endonuclease/hydrolase, and immunity repressor (Table 2). Given that several domains involved in DNA replication, transcription, and repair are identified in this protein, we hypothesize that the repressor domain binds DNA to aid in the complex function of this macromolecule. Pirin proteins are members of the cupin superfamily that have a conserved beta-barrel fold, bind to metals, and show a wide range of functions, including acting as transcription factors[32]. The pirin repressor fusion identified in *Nocardioides sp*. conserves the beta-barrel fold (Figure 8E), and its closest structural hit is the Quercetin 2,3-dioxygenase from *Acinetobacter baumannii* (PDBID 6D0G). The role of the repressor domain in the pirin is unclear, although it could aid in DNA binding if the protein functions in transcriptional control[33]. Finally, both *Xylanimonas cellulosilytica* and *Xylanimonas ulmi* contain fusions in which a repressor is fused to an aminotransferase domain (annotated as GntR/GabR and MocR, respectively (Table 2)). Unlike the serpin, AAA domain-containing, and pirin fusions, which all have the repressor at the C-terminus, the aminotransferase fusions contain a repressor domain at the N-terminus of the protein. The MocR/GabR family of proteins are transcriptional regulators that contain an N-terminal HTH containing DNA binding domain and a C-terminal pyridoxal 5'-phosphate (PLB) binding aminotransferase domain[34]. The GabR protein from *Bacillus subtilis* has been extensively studied, and in the presence of PLP and γ-aminobutyrate (GABA) it uses its DNA binding domain to regulate the transcription of genes

involved in GABA metabolism[35]. In the aminotransferase-repressor fusion protein (Figure 8F), the repressor domain could function as the DNA binding domain to aid in transcriptional regulation in *Xylanimonas*.

We attempted to express and purify the ENV421, serpin, AAA-domain containing, pirin, and aminotransferase fusion proteins for functional characterization; however, only the ENV421 and aminotransferase proteins were soluble. A BLASTp search on www.phagesdb.org reveals that the repressor domain of the ENV421 protein is most similar to the repressor protein in cluster A2 phage Pukovnik (67.28% identity; E-value $8e^{-72}$), and an analysis of the Pukovnik genome using MEME shows that it has 26 stoperator/operator sites with the same DNA consensus motif as TipsytheTRex and L5[11,13]. Given this similarity, we tested the ability of the ENV421 protein to bind the same A2 DNA substrate used for TipsytheTRex studies[13]. Figure 9A shows that the ENV421 protein binds the A2 DNA substrate, albeit with ~3-fold weaker affinity as compared to the A1 and A2 repressor proteins. This weaker affinity could be due to the presence of the additional protein domain fused to the repressor, or that we do not have the ideal DNA substrate for this repressor. We used SEC–SAXS–MALS to observe the solution behavior of the ENV421 protein both in the presence and absence of DNA (Figure 9A). In the presence of DNA, the SAXS data supports a model in which the repressor domain engages DNA in the same manner observed in cluster A1 and A2 repressors. In the absence of DNA, the SAXS data is best described by an ensemble of models in which the HTH domain of the repressor and the domain of unknown function sample multiple conformations in solution. Based on the MALS data, the protein is monomeric both the presence and absence of DNA.

A BLASTp search of the aminotransferase fusion reveals that the closest actinobacteriophage repressor is from the cluster BD2 phage Hank144. As noted above, we

could not find stoperator/operator sites in cluster BD phages, so we currently do not know what DNA substrate this protein binds. We analyzed the protein in the absence of DNA using SEC–SAXS–MALS, and the data support that this protein forms a dimer in solution similar to the *B. subtilis* GabR structure (Figure 9B)[35]. Collectively, these results suggest a broader function for the cluster A-like repressor protein other than silencing viral transcription, and additional work is needed to discover and characterize these novel functions in both phage and bacterial metabolism.

Methods

Oligonucleotides

Deoxynucleic acids were purchased from Integrated DNA Technologies (IDT). Double-stranded substrates were prepared by mixing complements together in a 1:1 ratio in 10 mM Tris (pH 8.0), 50 mM NaCl, and 1 mM EDTA, heated to 95 °C, and then allowed to cool to room temperature in a water bath.

Bacteriophage Isolation, Genomic Sequencing, and Virus Propagation

Bacteriophage Adahisdi was isolated from a soil sample on the campus of Western Carolina University (Cullowhee, NC) by student Aaron Stevens in 2015. Adahisdi virus was plaque purified and its full genome sequenced, with Illumina Sequencing and genome assembly performed at the Pittsburgh Bacteriophage Institute. The Adahisdi genome was annotated by students at Western Carolina University and deposited in GenBank (accession number MF919536 [https://www.ncbi.nlm.nih.gov/nuccore/1464441601]). For the current study, a plaque-picked stock of Adahisdi was grown from the original stock as described previously[13].

Protein Expression and Purification

The Adahisdi repressor gene was amplified from bacteriophage genomic DNA, then subcloned into the pET28a vector between NdeI and XhoI sites using the NEB HiFi DNA assembly kit. The resulting construct produces the repressor protein with an N-terminal His_6 tag and was confirmed by Sanger sequencing. The repressor protein was expressed in *E. coli* BL21(DE3) cells by growing in LB broth at 37 °C, with shaking, until the OD_{600} = 0.6-0.8. Cells were cooled on ice, then induced with 1 mM Isopropyl β-d-1-thiogalactopyranoside at 16 °C with shaking overnight. Initial attempts to purify the protein were performed using the same procedure described for the TipsytheTRex repressor[13]. However, purification in the presence of 0.5 M NaCl resulted in a protein contaminated with DNA. To alleviate this issue, the amount of NaCl was increased to 1.0 M during cell lysis and nickel affinity chromatography. Cells were harvested by centrifugation and resuspended in 50 mM Tris pH 8.0, 1.0 M NaCl, 0.1mM EDTA, 5% glycerol, and 0.1 mg/mL lysozyme. Cells were lysed by sonication, followed by a 1-hour centrifugation at 34,541 xg at 4 °C. The resulting supernatant was loaded on a 5 mL HiTrap nickel column (GE) equilibrated in 50 mM Tris pH 8.0, 1.0 M NaCl, 5% glycerol, and 10 mM imidazole. The column was washed with 40 mM imidazole, and the protein eluted by increasing the imidazole to 250 mM. Fractions containing the repressor protein were identified by SDS-PAGE, pooled, diluted to 0.1 M NaCl in 20 mM Tris pH 7.5, 0.5 mM EDTA, and 5% glycerol buffer, and loaded on a 5 mL HiTrap heparin column (GE). The protein was eluted from heparin using a 0.1-1.5 M NaCl gradient over 200 minutes. Fractions containing the repressor protein were identified by SDS-PAGE, pooled, then dialyzed against 2L of buffer containing 20 mM Tris pH 7.5, 0.5M NaCl, 0.5 mM EDTA, and 5% glycerol. The protein was concentrated, then flash frozen in liquid nitrogen for storage at -80 °C.

Mutant Adahisdi proteins were generated using site-directed mutagenesis and confirmed by Sanger sequencing. For DNA binding assays, both wild-type and mutant Adahisdi proteins were expressed as described above and purified using a 1.0 mL nickel gravity flow column. The column was washed with 40 mM imidazole, and the protein eluted by increasing the imidazole to 250 mM. Fractions containing the repressor protein were identified by SDS-PAGE, pooled, then dialyzed against a buffer consisting of 20 mM Tris pH 7.5, 0.5M NaCl, 0.5 mM EDTA, and 5% glycerol. The protein was concentrated, then flash frozen in liquid nitrogen for storage at -80 °C.

To generate the ENV421 and aminotransferase repressor fusion proteins, codon-optimized genes for expression in *E. coli* were synthesized and cloned into pET28a by Twist Bioscience (South San Francisco, CA). Both proteins were expressed in *E. coli* BL21(DE3) cells by growing in LB broth at 37 °C, with shaking, until the OD_{600} = 0.6-0.8. Cells were cooled on ice, then induced with 1 mM Isopropyl β-d-1-thiogalactopyranoside at 16 °C with shaking overnight. Cells were harvested by centrifugation and resuspended in 50 mM Tris pH 8.0, 0.5 M NaCl, 0.1mM EDTA, 5% glycerol, and 0.1 mg/mL lysozyme. The ENV421 protein was purified as described for the wild-type Adahisdi except that the protein was passed only over a nickel column. The MocR protein was purified exactly as described for the wild-type Adahisdi, using both Nickel and heparin columns.

X-Ray Structure Determination and Refinement

To prepare for crystallization, 250 μM repressor was mixed with an equal concentration of double-stranded DNA substrate generated from the oligos 5'-TAGGTTACGAATCAAGTCA and 5'-TGACTTGATTCGTAACCTA and allowed to incubate at room temperature for 30 minutes. Crystals of the repressor:DNA complex were obtained by hanging-drop vapor diffusion at 22 °C by mixing 1 μL protein with 1 μL reservoir solution placed over a 500 μL reservoir

solution containing 14-24% PEG 8,000, 0.1-0.4 M magnesium acetate, and 0.1M MES pH 6.0. Crystals appeared in 1-2 days and grew to full size by one week. To prepare crystals for freezing, a solution containing the reservoir plus 30% glycerol was prepared and slowly added directly to the drop containing the crystals. After equilibration, crystals were then quickly transferred to the reservoir 30% glycerol solution, then flash frozen by plunging into liquid nitrogen. X-ray data were collected at a wavelength of 1.1158 Å at beamline 8.3.1 at the Advanced Light Source on a Dectris Pilatus3 S 6M detector. The dataset were processed using XDS V20190315[36].

The structure of the repressor:DNA complex was determined by molecular replacement using Phaser[37], with an AlphaFold[15] model of the Adahisdi repressor used as the search model. The resulting solution contains four molecules of the repressor in the ASU and provided clear density for the DNA. DNA was added to the model and refined using Phenix with Amber enabled, simulated annealing, and torsion angle refinement. For the final stages of refinement, the structure was refined with subsequent rounds Phenix refinement with Amber enabled, TLS, and individual B-factor, followed by PDB-REDO[38]. The final refined model contains an R_{work}/R_{free} of 21.5/23.0% with good geometry. The His-tag and first fourteen residues of the N-terminus, as well as the C-terminal two residues were disordered in both structures and are not included in the final models. A summary of data collection and refinement statistics are provided in Table 1. Structure factors and coordinates have been deposited in the PDB as entry 36UU [https://doi.org/10.2210/pdb36UU/pdb].

Small-angle X-ray Scattering Coupled with Multi-Angle Light Scattering in Line With Size-Exclusion Chromatography (SEC—MALS—SAXS)

SEC–SAXS–MALS was performed as described previously[13] at the ALS beamline 12.3.1 LBNL Berkely, California[39]. For the Adahisdi repressor studies, either the protein alone or in the

presence of the same double stranded DNA substrate used for crystallization was prepared at a concentration of 5 mg/mL in 50 mM sodium phosphate pH 7.0, 1% w/v sucrose, 0.5 mM DTT, and 1 mM ATP and run over a Shodex 804 column. For the ENV421 repressor fusion, 5 mg/mL protein was analyzed either alone or in the presence of a double stranded DNA formed by annealing the strands 5'-TTTCGGTGGCTGTCAAGCGGG and 5'-CCCGCTTGACAGCCACCGAAA in 20 mM Tris pH 7.5, 0.25 M NaCl, and 1% glycerol, and run over a Shodex 802.5 column. For the Aminotransferase fusion protein, 5 mg/mL protein in 20 mM Tris pH 7.5, 0.4 M NaCl, and 1% glycerol was passed over a Shodex 803 column.

For SAXS solution structural modeling, the full-length apo Adahisdi repressor, the ENV421 fusion monomer with and without DNA, and the dimeric Aminotransferase fusion, were modeled using AlphaFold. For the Adahisdi repressor bound to DNA, the disordered His-tag was added to the crystal structure model. Minimal molecular dynamic simulations were performed on the models using BILBoMD[40,41]. All SAXS data has been deposited at Simple Scattering (www.simplescattering.com).

Superinfection Immunity Assays

The wild-type Adahisdi *repressor*, plus the upstream intergenic region that contains an endogenous promoter, was PCR-amplified from genomic DNA. The *repressor* plus intergenic was then cloned into the EcoRI site of the pMH94 integration shuttle-vector[42] and verified by Sanger sequencing. The pMH94 vector was a kind gift from the Hatfull laboratory. Site-directed mutagenesis of the pMH94 vector plus the *repressor* was used to generate the panel of repressor mutants. All mutant *repressor* sequences were verified via Sanger sequencing. Plasmid growth, cloning, and site-directed mutagenesis was performed using NEB5 alpha *E. coli* cells. PCR products and plasmids were purified using NEB Monarch Nucleic Acid Purification Kits.

For the superinfection immunity assays, purified pMH94 empty vector, or pMH94 plasmids containing Adahisdi *repressor* constructs, were electroporated into electrocompetent *Mycobacterium smegmatis* mc^2155 cells and plated onto Middlebrook 7H9 plates containing 5 μg/mL Kanamycin. Plates were incubated for 4 days at 37 °C. Bacterial colonies were then picked and grown to saturation in liquid culture as described above, with the addition of 5 μg/mL Kanamycin. 0.5 mL of cells were then mixed with 4.5 mL Middlebrook 7H9 top agar (5 μg/mL Kanamycin) and poured onto Middlebrook 7H9 plates (5 mg/mL Kanamycin). An Adahisdi virus stock (1.3×10^{10} PFU/mL) was then serially diluted in phage buffer, and 2 μL droplets were spotted in triplicate onto the top agar layer. Plates were incubated for 48 hours at 37 °C, then spot titers and Efficiencies of Plating (Repressor Construct Spot Titer/pMH94 Empty Vector Spot Titer) were calculated. Results provided are replicates from three independent experiments.

DNA Binding Assays

Electrophoretic mobility shift assays (EMSA) were used to monitor repressor binding to DNA substrates. For the wild-type Adahisdi repressor and all point mutations, 2 μM protein was mixed with 0.1 μM labeled DNA substrate in binding buffer that contained 20 mM Tris pH 7.0, 25 mM NaCl, 10 mM EDTA, and 5% glycerol. The double-stranded DNA substrate used was generated by annealing 5-GGTAGGTTACGAATCAAGTCAGT/36-FAM/ with 5-ACTGACTTGATTCGTAACCTACC. This complex was immediately serially diluted two-fold in a solution containing 0.1 μM labeled DNA substrate in binding buffer to monitor a protein concentration range that varied from 0.008-2 μM. Once diluted, the samples were incubated at room temperature for 30 minutes, then 15 μL of each sample was loaded on a 0.5% agarose gel prepared in 1X TAE buffer. The gels were run at 100V for 15 minutes at room temperature in 1X

TAE buffer, then imaged for fluorescein fluorescence using a Bio-Rad ChemiDoc MP System. Both free DNA and protein:DNA complex band intensities were quantified using the Image Lab V6.1 software. For the W163A mutant, the assay was performed the same as described for wild-type Adahisdi except that the starting protein concentration was 5 μM and the DNA concentration was 0.25 μM. For the ENV421 repressor fusion, the assay was performed the same as described for Adahisdi except that the starting protein concentration was 10 μM, and the DNA substrate used was designed to mimic the TipsytheTRex DNA consensus motif. The DNA substrate used for ENV421 was generated by annealing 5-GTTTCGGTGGCTGTCAAGCGGGC/36-FAM/ with 5-GCCCGCTTGACAGCCACCGAAAC. All DNA binding experiments were performed in triplicate.

Identification of Operator and Stoperator Sites

To identify operator and stoperator sites in bacteriophage genomes, bacteriophage genome fasta files were downloaded from www.phagesdb.org and loaded into the MEME sever[18] and analyzed using the same parameters previously described[4,13].

BLAST Analysis, AlphaFold Modeling of Repressor Proteins, Sequence Alignments, and Phylogenetic Tree Construction

To identify cluster A-like repressor fusion proteins, PSI-BLAST[43] was used with just the stoperator domain of the Adahisdi repressor (residues 53-170) as the query. The inclusion of the N-terminal HTH domain resulted in excessive hits that were not immunity repressors. PSI-BLAST ran for seven iterations before it found no new sequences that met the 0.005 *E-value* threshold. PSI-BLAST identified 32 sequences being greater than or equal to 250 residues in length. To generate AlphaFold models of repressor and repressor fusion proteins, the sequences identified in the BLAST searches were entereed into the AlphaFold server

(https://alphafoldserver.com/). DNA sequences and Mg^{2+} ions were included for AlphaFold repressor:DNA complex models.

Structure-based sequence alignments were performed using CLUSTAL Omega[44] and ESPript version 3.2[45]. For the phylogenetic tree construction, representative sequences from each A subcluster, as well as bacteriophages from other clusters that contain a cluster-A like repressor, were loaded into the program UGENE v53.1[24,46] and aligned using CLUSTAL W[47] using default settings. The phylogenetic tree was then constructed using a PHYLIP Neighbor Joining tree building method using default parameters.

Data Availability

Coordinates and structure factors for the crystal structure have been deposited in the PDB under accession code 36UU [https://doi.org/10.2210/pdb36UU/pdb]. The Adahisdi genome sequence has been deposited in GenBank under accession code MH697575.1 [https://www.ncbi.nlm.nih.gov/nuccore/1464441601]. All SAXS data is available at https://simplescattering.com/ under the following Simple Scattering Codes: Apo Adahisdi repressor: XSPUTHBF; Adahisdi repressor:DNA complex: XSXJBGOT; Apo ENV421 repressor fusion: XSENMPZV; ENV421:DNA complex: XSLHXGUB; Apo MocR Aminotransferase repressor fusion: XSYFC4ZF.

References

1. Hatfull, G.F. Mycobacteriophages: From Petri dish to patient. *PLoS Pathog* **18**, e1010602 (2022).
2. 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).
3. Ptashne, M. *A genetic switch : phage lambda revisited*, xiv, 154 p. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2004).

4. Mavrich, T.N. & Hatfull, G.F. Evolution of Superinfection Immunity in Cluster A Mycobacteriophages. *mBio* **10**(2019).
5. Dedrick, R.M. et al. Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant Mycobacterium abscessus. *Nat Med* **25**, 730-733 (2019).
6. Nick, J.A. et al. Host and pathogen response to bacteriophage engineered against Mycobacterium abscessus lung infection. *Cell* **185**, 1860-1874 e12 (2022).
7. Stayrook, S., Jaru-Ampornpan, P., Ni, J., Hochschild, A. & Lewis, M. Crystal structure of the lambda repressor and a model for pairwise cooperative operator binding. *Nature* **452**, 1022-5 (2008).
8. Hatfull, G.F. The secret lives of mycobacteriophages. *Adv Virus Res* **82**, 179-288 (2012).
9. Hatfull, G.F. Mycobacteriophages. *Microbiol Spectr* **6**(2018).
10. Villanueva, V.M., Oldfield, L.M. & Hatfull, G.F. An Unusual Phage Repressor Encoded by Mycobacteriophage BPs. *PLoS One* **10**, e0137187 (2015).
11. Brown, K.L., Sarkis, G.J., Wadsworth, C. & Hatfull, G.F. Transcriptional silencing by the mycobacteriophage L5 repressor. *EMBO J* **16**, 5914-21 (1997).
12. Jain, S. & Hatfull, G.F. Transcriptional regulation and immunity in mycobacteriophage Bxb1. *Mol Microbiol* **38**, 971-85 (2000).
13. McGinnis, R.J. et al. A monomeric mycobacteriophage immunity repressor utilizes two domains to recognize an asymmetric DNA sequence. *Nat Commun* **13**, 4105 (2022).
14. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
15. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493-500 (2024).
16. Sayers, E.W. et al. Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Res* **53**, D20-D29 (2025).
17. Putnam, C.D., Hammel, M., Hura, G.L. & Tainer, J.A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* **40**, 191-285 (2007).
18. Bailey, T.L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**, W202-8 (2009).
19. Sagendorf, J.M., Markarian, N., Berman, H.M. & Rohs, R. DNAproDB: an expanded database and web-based tool for structural analysis of DNA-protein complexes. *Nucleic Acids Res* **48**, D277-D287 (2020).
20. Pope, W.H. et al. Bacteriophages of Gordonia spp. Display a Spectrum of Diversity and Genetic Relationships. *mBio* **8**(2017).
21. Russell, D.A. & Hatfull, G.F. PhagesDB: the actinobacteriophage database. *Bioinformatics* **33**, 784-786 (2017).
22. Shafer, N., Dubrule, B.E. & De Buck, J. Mycobacteriophage Mcgavigan Uses Noncanonical Bxb1-Like Repressor for Heterotypic Superinfection Immunity. *J Basic Microbiol* **66**, e70133 (2026).
23. Pope, W.H. et al. Expanding the diversity of mycobacteriophages: insights into genome architecture and evolution. *PLoS One* **6**, e16329 (2011).
24. Okonechnikov, K., Golosova, O., Fursov, M. & team, U. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**, 1166-7 (2012).
25. Ko, C.C. & Hatfull, G.F. Mycobacteriophage Fruitloop gp52 inactivates Wag31 (DivIVA) to prevent heterotypic superinfection. *Molecular Microbiology* **108**, 443-460 (2018).

26. Aitio, O. et al. Recognition of tandem PxxP motifs as a unique Src homology 3-binding mode triggers pathogen-driven actin assembly. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 21743-21748 (2010).
27. Kenney, S.P., Wentworth, J.L., Heffron, C.L. & Meng, X.J. Replacement of the hepatitis E virus ORF3 protein PxxP motif with heterologous late domain motifs affects virus release via interaction with TSG101. *Virology* **486**, 198-208 (2015).
28. Gabler, F. et al. Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Curr Protoc Bioinformatics* **72**, e108 (2020).
29. Bouton, M.C. et al. The under-appreciated world of the serpin family of serine proteinase inhibitors. *EMBO Mol Med* **15**, e17144 (2023).
30. Xu, J. et al. DNA accelerates the protease inhibition of a bacterial serpin chloropin. *Front Mol Biosci* **10**, 1157186 (2023).
31. Zimmermann, L. et al. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol* **430**, 2237-2243 (2018).
32. Perez-Dominguez, F. et al. Role of Pirin, an Oxidative Stress Sensor Protein, in Epithelial Carcinogenesis. *Biology (Basel)* **10**(2021).
33. Pang, H. et al. Crystal structure of human pirin: an iron-binding nuclear protein and transcription cofactor. *J Biol Chem* **279**, 1491-8 (2004).
34. Okuda, K. et al. Role of the aminotransferase domain in Bacillus subtilis GabR, a pyridoxal 5'-phosphate-dependent transcriptional regulator. *Mol Microbiol* **95**, 245-57 (2015).
35. Edayathumangalam, R. et al. Crystal structure of Bacillus subtilis GabR, an autorepressor and transcriptional activator of gabT. *Proc Natl Acad Sci U S A* **110**, 17820-5 (2013).
36. Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-32 (2010).
37. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
38. Joosten, R.P., Long, F., Murshudov, G.N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **1**, 213-20 (2014).
39. Classen, S. et al. Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. *J Appl Crystallogr* **46**, 1-13 (2013).
40. Pelikan, M., Hura, G.L. & Hammel, M. Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen Physiol Biophys* **28**, 174-89 (2009).
41. Classen, S. et al. BilboMD: a web-accessible SAXS and AlphaFold-guided modeling pipeline. *Nucleic Acids Res* (2026).
42. Lee, M.H., Pascopella, L., Jacobs, W.R., Jr. & Hatfull, G.F. Site-specific integration of mycobacteriophage L5: integration-proficient vectors for Mycobacterium smegmatis, Mycobacterium tuberculosis, and bacille Calmette-Guerin. *Proc Natl Acad Sci U S A* **88**, 3111-5 (1991).
43. Altschul, S.F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-402 (1997).
44. Madeira, F. et al. The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res* **52**, W521-W525 (2024).
45. Robert, X., Guillon, C. & Gouet, P. FoldScript: a web server for the efficient analysis of AI-generated 3D protein models. *Nucleic Acids Res* **53**, W277-W282 (2025).

46. Rose, R., Golosova, O., Sukhomlinov, D., Tiunov, A. & Prosperi, M. Flexible design of multiple metagenomics classification pipelines with UGENE. *Bioinformatics* **35**, 1963-1965 (2019).
47. Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-80 (1994).

Acknowledgements

The authors would like to thank George Meigs and James Holton, beamline 8.3.1 at the Advanced Light Source (ALS), for their assistance in X-ray data collection and processing. Beamline 8.3.1 at the Advanced Light Source is operated by the National Institutes of Health (R01 GM124149 and P30 GM124169), and the Integrated Diffraction Analysis Technologies program of the US Department of Energy Office of Biological and Environmental Research. The Advanced Light Source (Berkeley, CA) is a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the US Department of Energy under contract number DE-AC02-05CH11231, Office of Basic Energy Sciences. SAXS data was collected at the SIBYLS beamline 12.3.1 at the Advanced Light Source (Berkeley, CA) which is supported by the National Institutes of Health (P30 GM124169), and the Integrated Diffraction Analysis Technologies program of the US Department of Energy Office of Biological and Environmental Research. The Advanced Light Source is a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the US Department of Energy under contract DE-AC02-05CH11231, Office of Basic Energy Sciences. This work was supported by the Howard Hughes Medical Institute's Science Education Alliance as well as by funds provided by Western Carolina University (MDG and JRW). We are grateful to WCU students in the 2015-2016 Phage Hunters and 2023 Biochemistry Laboratory courses for their assistance in data collection, and Megan Eckardt for assistance with the virus discovery course at WCU.

Author Contributions

RM, SP, TJH, and JRW carried out X-ray crystallization, data collection, structure determination, and refinement. HB, AT, and JRW carried out the AlphaFold modeling and bioinformatic analysis of repressor fusion proteins. AT and JRW collected and analyzed SEC-SAXS-MALS data. MT and MDG designed and carried out the *in vivo* superinfection immunity assays. RM, SP, HB, AT, and JRW performed DNA binding experiments. MDG and JRW designed and supervised the research. JRW wrote the manuscript, with input from all authors.

Competing Interests

The authors declare no competing interests.

Table 1. Data collection and refinement statistics for the Adahisdi repressor crystal structure

Data collection	
Space group	$P2_1$
Cell dimensions	
a, *b*, *c* (Å)	62.39, 170.04, 66.97
α, β, γ (°)	90, 107.16, 90
Resolution (Å)	49.66-2.30 (2.36-2.30)*
Wavelength (Å)	1.1158
Total Reflections	397826 (32213)
Unique Reflections	115413 (9631)
CC 1/2	0.999 (0.812)
CC*	1.0 (0.947)
R_{merge} (%)	3.69 (57.55)
I / σ	14.81 (1.78)
Wilson B-factor	55.21
Completeness (%)	99.06 (99.55)
Redundancy	3.4 (3.3)
Refinement	
Resolution (Å)	49.66-2.30 (2.37-2.30)*
No. reflections	58468 (4850)
R_{work} / R_{free} (%)	21.47/22.93 (33.27/35.14)
No. atoms	
Protein/DNA	8517
Water	51
Average *B*-factors	
Protein/DNA	72.12
Water	54.62
Stereochemical ideality	
RMS Bond lengths (Å)	0.003
RMS Bond angles (°)	0.71
ϕ,ψ most favored (%)	97.37
ϕ,ψ allowed (%)	2.63
ϕ,ψ outliers (%)	0.00

*Values in parentheses are for highest-resolution shell.
Coordinates have been deposited in the Protein Data Bank as entry 36UU [https://doi.org/10.2210/pdb36UU/pdb].

Table 2. Summary of repressor fusion proteins present in viral and bacterial genomes

Accession Number/Protein Description	Organism	Protein Size (kDa)	Repressor Information	Other Domain Information	Probability (%)/E-value
WP_327204034.1 hypothetical protein	*Rhodococcus pyridinivorans*	31.2	Full repressor at C-terminus (residues 106-274)	Nothing Identified	99.96/ 7.5 e^{-28} (repressor)
WP_418345807.1 hypothetical protein	*Rhodococcus pyridinivorans*	31.1	Full repressor at C-terminus (residues 107-274)	Nothing Identified	99.96/ 7.1 e^{-28} (repressor)
WP_367880880.1 hypothetical protein	*Rhodococcus pyridinivorans*	35.9	Full repressor at C-terminus (residues 151-318)	Nothing Identified	99.96/ 1.1e^{-27} (repressor)
WP_083613183.1 serpin family protein	*Mycolicibacterium mageritense*	47.0	Full repressor at C-terminus (residues 260-421)	Serpin (residues 2-246): PDBID 1JMO	99.96/ 2.4e^{-28} (serpin) 99.92/ 4.3e^{-24} (repressor)
>WP_233209289.1 hypothetical protein	*Mycobacterium sp.* ENV421	29.9	Full repressor at C-terminus (residues 100-260)	Nothing identified	99.95/ 3.3e^{-26} (repressor)
WP_427098649.1 hypothetical protein	*Mycolicibacterium vaccae*	27.3	Full repressor at C-terminus (residues 83-242)	Transcription factor/repressor proteins, but with poor E-value (PDB 5CLV)	95.04/0.19 (factor) 99.92/4.1e^{-23} (repressor)
>WP_190267212.1 serpin family protein	*Gordonia hankookensis*	65.8	Full repressor at C-terminus (residues 434-602)	Serpin (residues 44-422): PDBID 1JMO	100/ 4.1e^{-44} (serpin) 99.90/ 2.1e^{-22} (repressor)
XEN16495.1 hetero-immunity repressor	Mycobacterium phage PhesterPhotato	27.5	Full repressor at C-terminus (residues 79-238)	Transposon gamma-delta resolvase (PDBID 2GM5)	97.76/3.0e^{-4} (resolvase) 99.93/ 4.0e^{-24} (repressor)
>YP_005087008.1 transcriptional repressor	Rhodococcus phage RGL3	30.8	Full repressor at N-terminus (residues 3-175)	Nothing identified	99.96/ 6.6e^{-28} (repressor)
>WP_235677880.1 hypothetical protein	*Mycolicibacterium sarraceniae*	30.0	Full repressor at C-terminus (residues 100-261)	Nothing identified	99.93/ 2.3e^{-24} (repressor)
>WP_231748648.1 hypothetical protein	*Mycobacterium sp.* M26	31.4	Full repressor at N-terminus (residues 7-179)	Nothing identified	99.96/1.1 e^{-28} (repressor)
WP_445169897.1 hypothetical protein	*Mycolicibacterium sp. Dal123E01*	29.9	Full repressor at C-terminus (residues 101-261)	Nothing identified	99.89/4.9 e^{-22} (repressor)
>WP_427099575.1 hypothetical protein	*Mycolicibacterium vaccae*	30.2	Full repressor at the C-terminus (residues 100-260)	Anti-CRISPR protein AcrIF24 (residues 4-201: PDBID 7XI1)	99.95/ 3.3e^{-26} (repressor) 95.39/ 0.11 (CRISPR)
>WP_234792637.1 AAA domain-containing protein[%]	*Mycolicibacterium fortuitum*	169.9	Full repressor at the C-terminus (residues 1365-1516)	RNA-helicase like protein (residues 720-1001: PDBID 8FTK); DNA binding protein SMUBP-2/hydrolase/helicase(residues 13-335: PDBID 4B3F); Mobile intron protein/honing endonuclease/hydrolase (residues 1022-1128: PDBID 3R3P)	98.7/ 1.1e^{-7} (SUMBP-2) 97.84 / 2.9e^{-4} (mobile intron) 99.93/ 1.1 e^{-24} (RNA helicase) 99.34/ 1.4e^{-12} (repressor)

>OBJ98398.1 ATP-binding protein IstB[%]	*Mycolicibacterium fortuitum*	210.3	Full repressor at the C-terminus (residues 1365-1516)	RNA-helicase like protein (residues 1056-1828: PDBID 8FTK); DNA binding protein SMUBP-2/hydrolase/helicase(residues 372-507: PDBID 4B3F); Mobile intron protein/homing endonuclease/hydrolase (residues 1381-1487: PDBID 3R3P)	98.61/ 2.5e^{-7} (SUMBP-2) 97.91/ 1.5e^{-4} (mobile intron) 99.89/ 8.6 e^{-22} (RNA helicase) 99.61/ 1.4e^{-15} (repressor)
>WP_246228823.1 hypothetical protein	*Mycolicibacterium psychrotolerans*	70.3	Full repressor at the C-terminus (residues 470-646)	Serpin (residues 73-442: PDBID 1JMO)	99.9/ 1.3e^{-22} (repressor) 100/ 1.1e^{-40} (serpin)
>WP_319431842.1 hypothetical protein	*Mycobacterium sp.* RTGN5	29.4	Full repressor at the C-terminus (residues 101-260)	Nothing identified	99.94/ 3.6e^{-25} (repressor)
>AMW64332.1 hypothetical protein	*Mycobacterium phage ChipMunk*	26.5	Full repressor at the C-terminus (residues 55-229)	Nothing identified	99.97/ 4.6e^{-30} (repressor)
>AHK31009.1 Repressor-like immunity protein	*Rhodococcus opacus* PD630	37.4	Full repressor at the C-terminus (residues 170-331)	Nothing identified	99.95/ 1.1e^{-26} (repressor)
>UGL61796.1 immunity repressor	Mycobacterium phage Grub	28.2	Full repressor at the N-terminus (residues 16-182)	Nothing identified	99.88/ 2.5e^{-21} (repressor)
>GAA2655495.1 hypothetical protein	*Streptomyces spororaveus*	26.7	Full repressor at C-terminus (residues 77-226)	Nothing identified	99.79/3.2 e^{-18} (repressor)
>WP_216699470.1 hypothetical protein	*Actinotalea ferrariae*	28.1	Full repressor at N-terminus (residues 3-144)	Nothing identified	99.74/9.9 e^{-17} (repressor)
>MCW2843333.1 Pirin domain protein	*Nocardioides sp.*	49.1	Full repressor at the C-terminus (residues 300-443)	Pirin (residues 4-311: PDBID 6D0G)	100/ 4.5e^{-49} (pirin) 99.71/ 1.5e^{-16} (repressor)
>AOQ29445.1 immunity repressor	*Mycobacterium phage Bigfoot*	26.8	Full repressor at C-terminus (residues 69-205)	Nothing identified	99.78/2.2 e^{-17} (repressor)
>WP_207207318.1 hypothetical protein	*Xylanimonas protaetiae*	37.8	Full repressor at the C-terminus (residues 175-326)	Nothing identified	99.79/ 3e^{-18} (repressor)
WP_346537388.1 sigma factor-like helix-turn-helix DNA-binding protein	*Micromonospora palomenae*	26.2	Full repressor at N-terminus (residues 3-142)	Nothing identified	99.90/6.9 e^{-23} (repressor)
>WP_012877027.1 hypothetical protein	*Xylanimonas cellulosilytica*	30.7	Stoperator domain only at C-terminus (residues 169-277)	Nothing identified	99.41/ 2.62e^{-12} (repressor
WP_190092309.1 hypothetical protein	*Streptomyces melanogenes*	27.4	Full repressor at C-terminus (residues 100-242)	Structure of unknown function, but with poor E-value (PDB 2OA4)	96.17/0.032 (factor) 99.80/1.5e^{-18} (repressor)
>WP_291278509.1 hypothetical protein	*Galactobacter sp.*	32.2	Stoperator domain only at C-terminus (residues 177-275)	Nothing identified	99.58/ 1.3e^{-14} (repressor)
>WP_266766066.1 DUF6192 family protein	*Streptomyces sp. NBC_00638*	26.8	Stoperator domain at N-terminus (residues 20-109)	Nothing identified	99.04/9.2 e^{-10} (repressor)

>ACZ32369.1 putative transcriptional regulator, GntR family	*Xylanimonas cellulosilytica* DSM 15894	67.1	Full repressor (residues 5-141) at N-terminus	HTH-type transcriptional regulatory protein GabR (residues 139-585: PDBID 4TV7)	99.82/ $5.6e^{-19}$ (repressor) 100/ $1.3e^{-35}$ (GabR)
>RZS62629.1 DNA-binding transcriptional MocR family regulator	*Xylanimonas ulmi*	66.4	Full repressor (residues 5-157) at N-terminus	Aminotransferase (residues 192-576; PDBID 3AOW)	99.87/ $2.4e^{-21}$ (repressor) 100/ $6.25e^{-35}$ (aminotransferase)

The order of the sequences is based on the final output from PSI-BLAST.
Protein size information was obtained using the amino acid sequence in Expasy ProtParam. Information on the protein domains, % probability scores, and E-values were obtained using HHPred.
[%]This sequence is reported twice in the NCBI database in *Mycolicibacterium fortuitum* strain 1089381.4, once as a 169.0 kDa polypeptide and the second as a 210.3 kDa polypeptide. Both hits are in the same CDS and are the same protein.

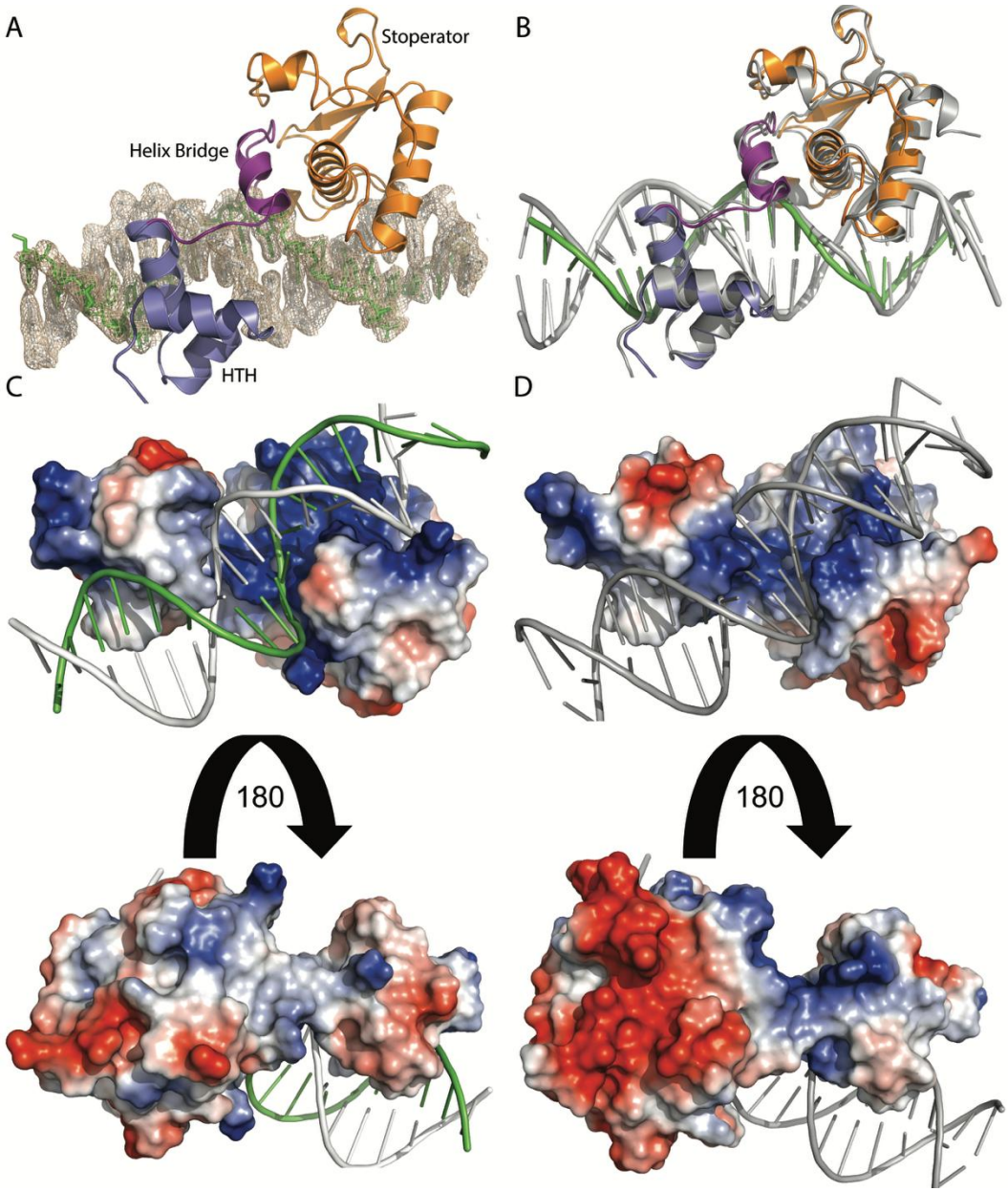


Figure 1: Structure of the cluster A1 immunity repressor
a) Cartoon representation of the repressor bound to DNA. The protein monomer is colored by domain and includes the N-terminal HTH (residues 8-47, slate blue), the helix bridge (residues 48-67, magenta), and the stoperator (residues 68-170, orange). The two DNA strands are colored in green and white, with $2F_o$-F_c density, contoured at 1 sigma, shown in sand for the DNA helix. **b**) Superposition of the Adahisdi repressor structure (color coded as in **a)**) on the cluster A2 TipsytheTRex structure (colored gray). The structures superimpose well, with a rmsd of 1.31 Å over 160 residues. **c-d**) Electrostatic surface renderings (red: negative potential, blue: positive potential, white: neutral) of the Adahisdi (panel **c**)) and TipsytheTRex (panel **d**)) repressor structures reveal a strong positive potential at the DNA binding interface. While the TipsytheTRex structure has an acidic surface at the C-terminus of the protein, the equivalent position in the Adahisdi structure lacks such strong acidic potential.

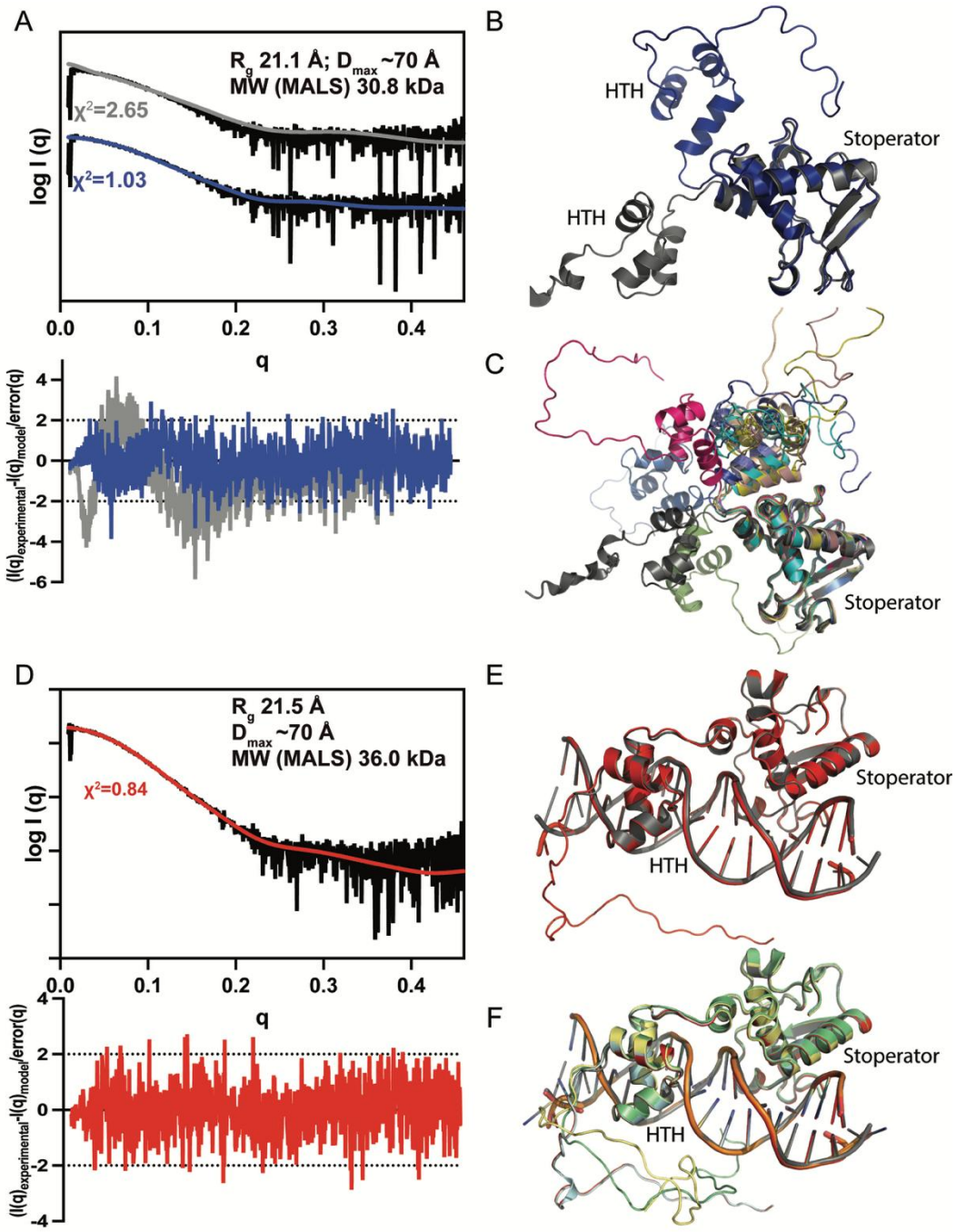


Figure 2: SAXS analysis of the Adahisdi repressor

a) Experimental SAXS curve of the Adahisdi repressor in the absence of DNA is shown in black along with theoretical scattering profiles of the Adahisdi crystal structure (with DNA removed and his tag added, gray) and a model generated from BILBOMD (blue). SAXS parameters and the molecular weight calculated from MALS are shown along with fit-residuals and χ^2 values, which show that the BILBOMD model more accurately describes the repressor structure in solution. **b**) Atomistic models used to fit the SAXS curve in panel **a**). The BILBOMD model (blue) structure takes on a more compact shape as compared the crystal structure conformation (gray), with the HTH moving closer to the stoperator domain. **c**) Along with the BILBOMD

model shown in panel **b**), four additional BILBOMD models were generated to describe the experimental SAXS curve. Panel **c**) shows a superposition of models provided by BILBOMD, with χ^2 values ranging from 0.93-1.03. While the stoperator domains all superimpose well, the his-tags and HTH domains show significant movement indicating that the protein is flexible in the absence of DNA. **d**) Experimental SAXS curve of the Adahisdi repressor in the presence of DNA (black) along with a theoretical scattering profile of a model generated from BILBOMD (red). SAXS parameters and the molecular weight calculated from MALS are shown along with fit-residuals and χ^2 values for the BILBOMD model. **e**) The BILBOMD model (red) that accurately describes the data in **d**) superimposes well with the Adahisdi repressor crystal structure (gray). **f**) BILBOMD provided three different ensembles to describe the experimental SAXS curve shown in panel **d**), with χ^2 values ranging from 0.83-0.84. All three models superimpose well on the crystal structure, with the only notable movement being the N-terminal his-tag. These results support that in the presence of DNA, the repressor in solution takes on a rigid shape that is consistent with the crystal structure.

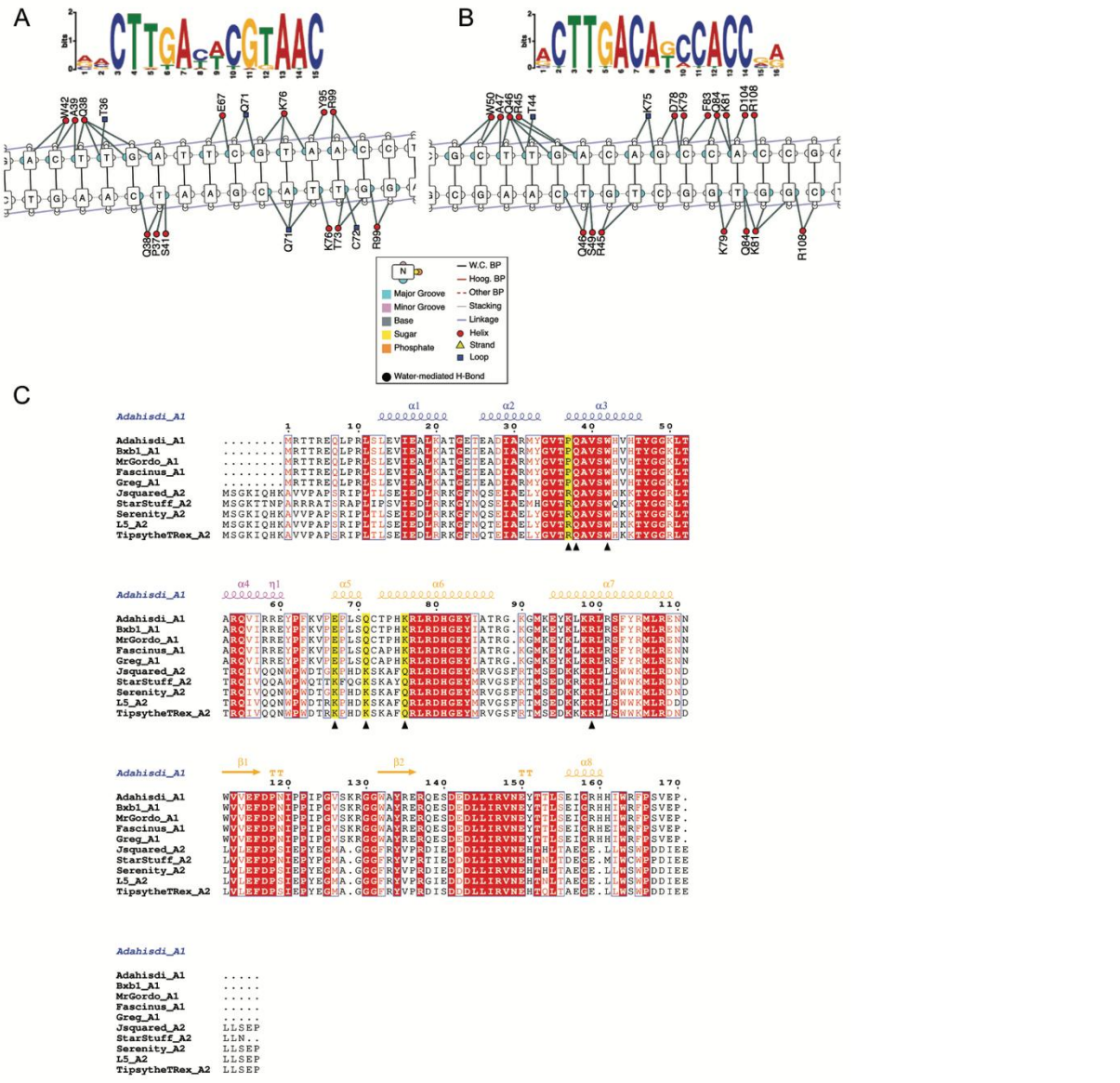


Figure 3: Protein:DNA interactions in cluster A1 and A2 mycobacteriophages
a) Adahisdi consensus operator/stoperator motif identified using MEME. The Adahisdi genome contains 37 operator/stoperator sites located throughout the genome. Also shown is a DNAProDB analysis of the interactions between the Adahisdi repressor and DNA. **b**) MEME consensus operator/stoperator sequence for the cluster A2 phage TipsytheTRex. The TipsytheTRex genome contains 26 operator/stoperator sites located throughout the genome. Also shown is a DNAProDB analysis of the interactions between the TipsytheTRex repressor and DNA. **c**) Sequence alignment of the Adahisdi repressor with a variety of cluster A1 and A2 repressor sequences. Residue numbering and secondary structural elements are based on the Adahisdi sequence and structure and are color coded as in Figure 1. Residues from the Adahisdi repressor that directly bind DNA bases are indicated with black triangles. While the majority of the sequence is highly conserved among the repressors, the yellow highlighted DNA binding residues show key differences in cluster A1 vs A2 repressors.

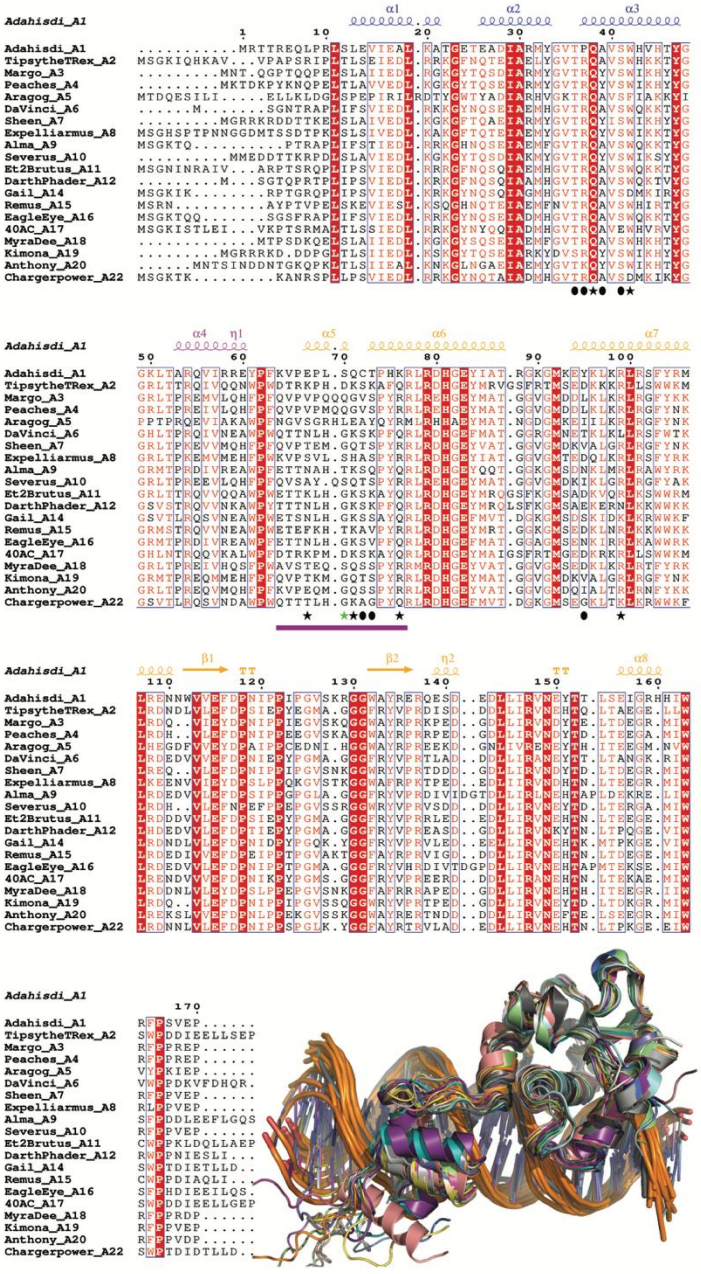
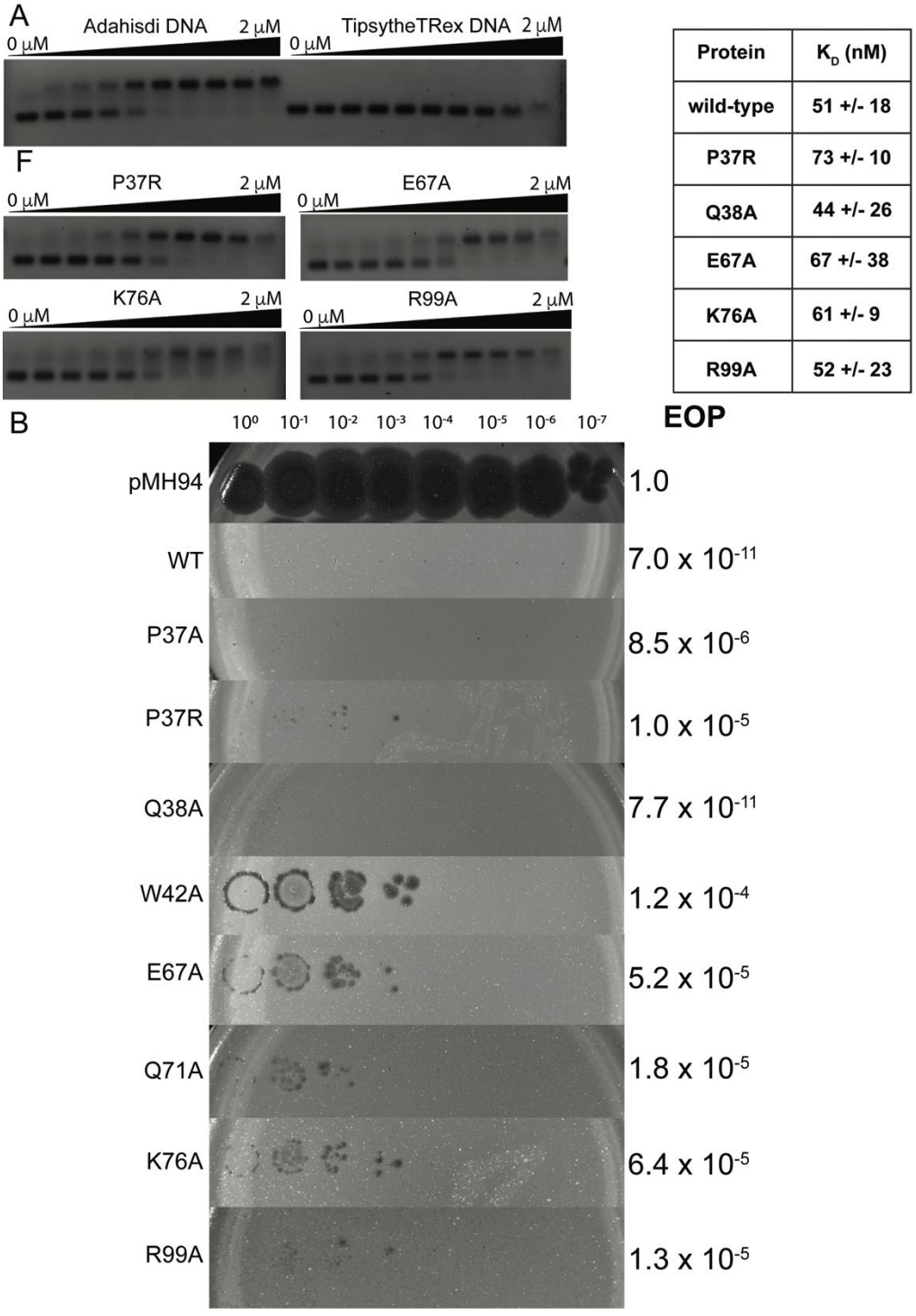


Figure 4: Sequence and structure conservation for cluster A repressors
A representative repressor from each A subcluster was selected, and all repressors were aligned using Clustal omega. Residue numbering and secondary structural elements are based on the Adahisdi sequence and structure and are color coded as in Figure 1. Black ovals represent residues in the Adahisdi structure that bind the DNA backbone, while black stars are residues that directly contact DNA bases. The green star indicates a residue in the TipsytheTRex structure that directly contacts DNA, and the equivalent residue in Adahisdi does not contact a DNA base. The purple line indicates the beginning of the stoperator domain that inserts in the major groove of DNA. This region lacks sequence conservation amongst the repressor sequences. The models shown are AlphaFold models for each sequence provided in the alignment, with the A1 repressor

crystal structure shown in dark gray. The iPTM and pTM values for all models are provided in Supplemental Figure 2. AlphaFold predicts a conservation of fold for all cluster A repressors.



Protein	K_D (nM)
wild-type	51 +/- 18
P37R	73 +/- 10
Q38A	44 +/- 26
E67A	67 +/- 38
K76A	61 +/- 9
R99A	52 +/- 23

Figure 5: Analysis of Adahisdi DNA binding residues

a) DNA binding studies of wild-type and select point mutant Adahisdi proteins. Wild-type Adahisdi was mixed with either an Adahisdi or TipsytheTRex consensus DNA motif, and DNA binding was assessed via an electrophoretic mobility shift assay. The wild-type Adahisdi protein binds the Adahisdi DNA with an affinity of 51 ±18 nM but fails to bind a TipstheyTRex DNA molecule. The point mutants were also assessed for their ability to bind Adahisdi DNA, and their affinities are provided in the table. The results show minimal differences between wild-type and

the mutants in the *in vitro* DNA binding analysis. **b**) *In vivo* superinfection immunity assays. Adahisdi virus was serially diluted and spotted onto top agar containing cells integrated with a single copy of the pMH94 empty vector, vector plus wild-type (WT) *repressor* and its endogenous promoter, or the indicated mutants. Spot titers and the efficiency of virus plating (EOP) were calculated as compared to vector-only containing cells. While Q38A has no observable phenotype as compared to WT, the other mutants tested show strong phenotypes and support that these residues are critical for repressor function *in vivo*.

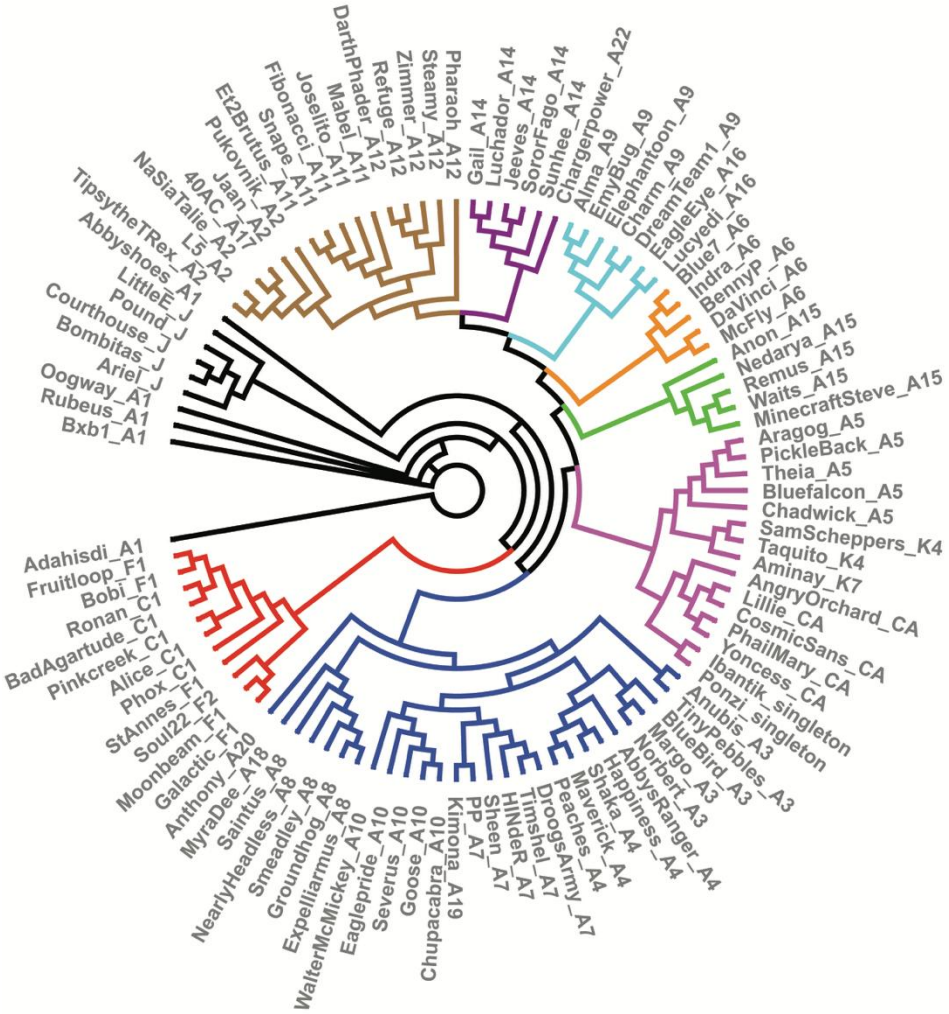


Figure 6: Evolutionary relationships of cluster A and cluster A-like immunity repressors. Representative sequences from each A subcluster, as well as cluster A-like repressors from other subclusters, were aligned in UGENE to build a phylogenetic tree. The tree is color-coded to represent sequences that share a common ancestor.

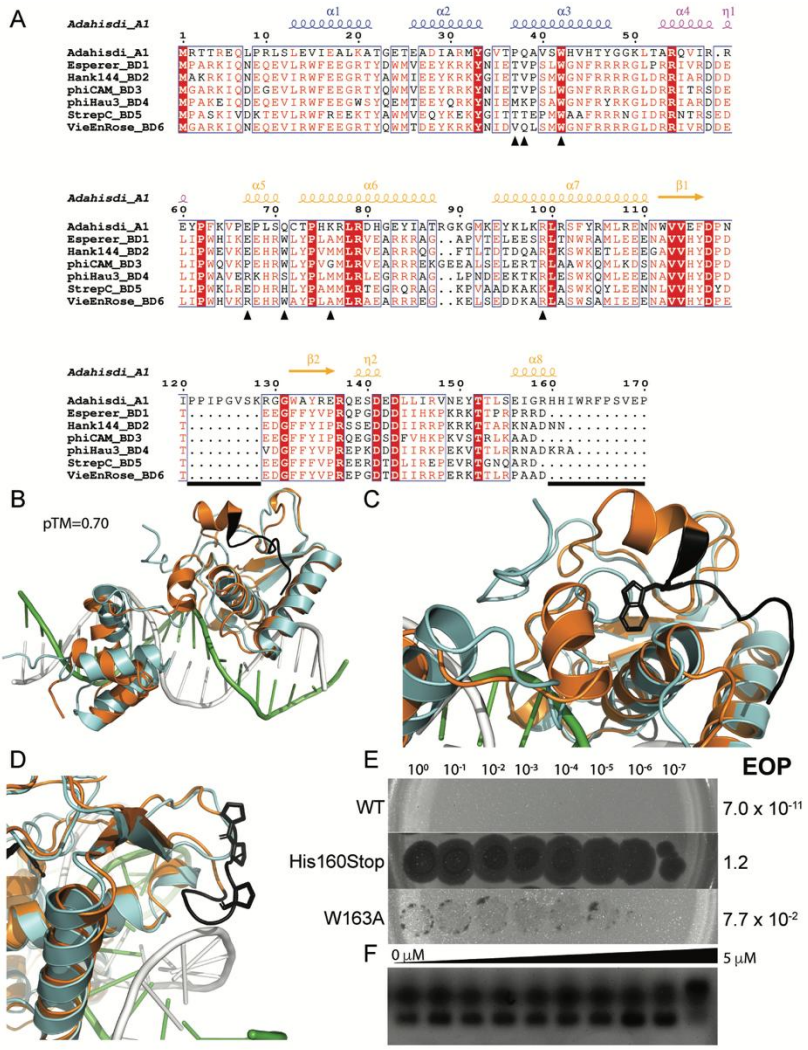


Figure 7: Cluster BD repressors conserve the repressor fold but lack elements essential for cluster A repressor function.

a) Alignment of the Adahisdi repressor with a representative of each BD subcluster. Residue numbering and secondary structural elements are based on the Adahisdi sequence and structure and are color coded as in Figure 1. Black triangles indicate residues in the Adahisdi structure that bind DNA. The two black lines highlight regions of the BD repressors that are missing as compared to the cluster A1 repressor. **b**) Structural superposition of the Adahisdi crystal structure (Orange protein: Green/White DNA) with an AlphaFold model of the Hank144 repressor (Cyan). The reported pTM value is for the Hank144 model. The models superimpose with a rmsd of 2.53 Å over 95 residues. Colored in black are the two regions of the Adahisdi structure that are missing in Hank144. **c**) A zoom in of the C-terminal tail of Adahisdi that is missing in the Hank144 model. The missing loop is colored black, and the Adahisdi W163 side chain is shown as sticks. **d**) A close-up view of the proline loop of Adahisdi missing in the Hank144 structure. The prolines are shown as sticks. **e**) Superinfection immunity assay results wt, His160stop, and W163A Adahisdi proteins. The His160stop mutant has completely abolished repressor function *in vivo*. The W163A mutant also has a significant phenotype, with an 11-log reduction in the efficiency of plating as compared to wt. **f**) DNA binding results for the W163A mutant, which reveals it is defective in DNA binding.

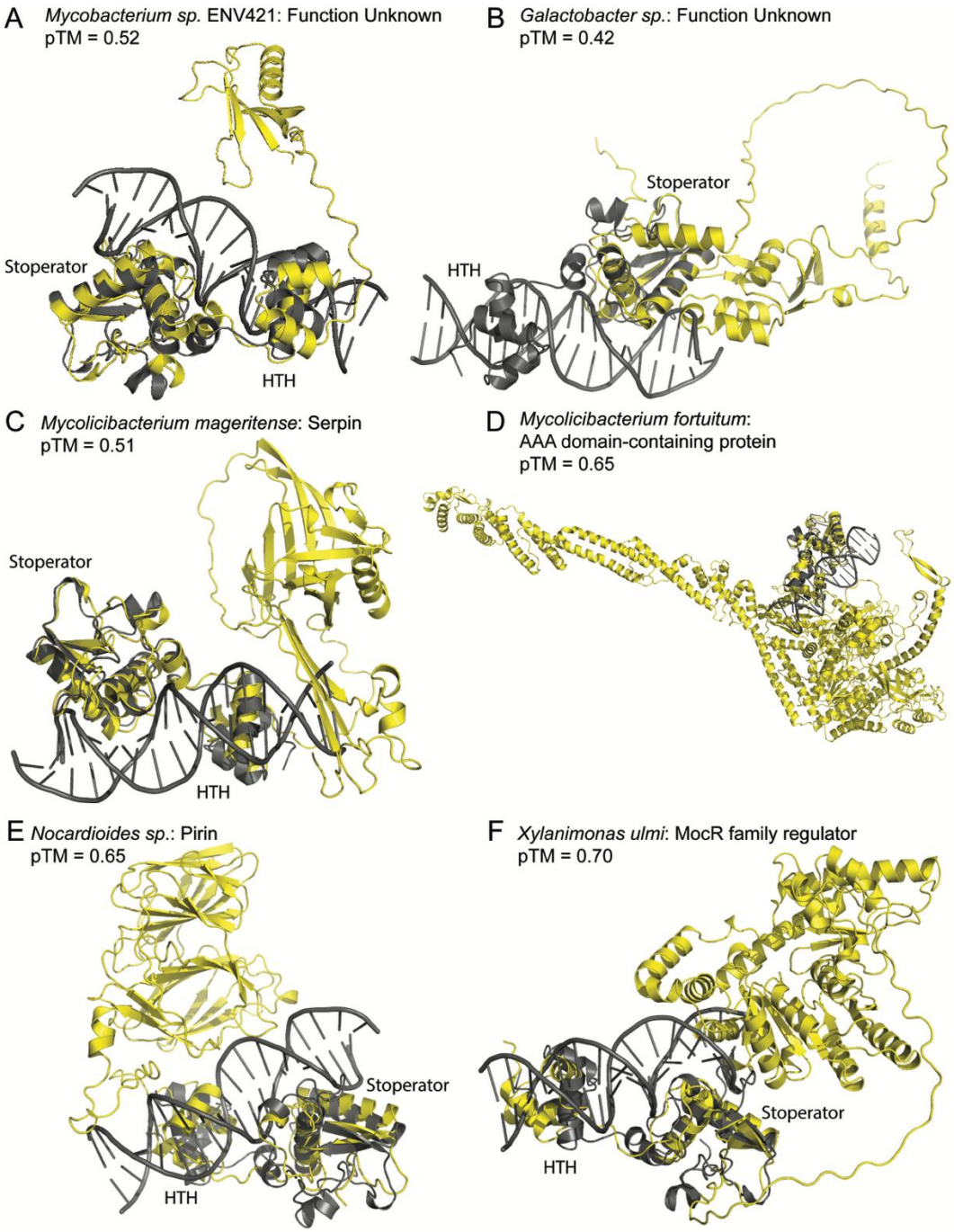


Figure 8: AlphaFold models of repressor fusion proteins show a conserved repressor fold. Repressor fusion AlphaFold models are presented for **a**) *Mycobacterium sp*. ENV421 repressor fusion of unknown function, **b**) *Galactobacter sp.* repressor fusion of unknown function, which lacks the HTH of the repressor protein, **c**) *Mycolicibacterium mageritense* serpin fusion, **d**) *Mycolicibacterium fortuitum* AAA domain-containing protein fusion, **e**) *Nocardioides sp.* pirin fusion, and **f**) *Xylanimonas ulmi* MocR family regulator (aminotransferase). All models are presented as monomers, and the repressor domain of each fusion is superimposed on the Adahisdi crystal structure, which is shown in dark gray. It is important to note that in all models the lower pTM values are mainly due to predicted flexible regions.

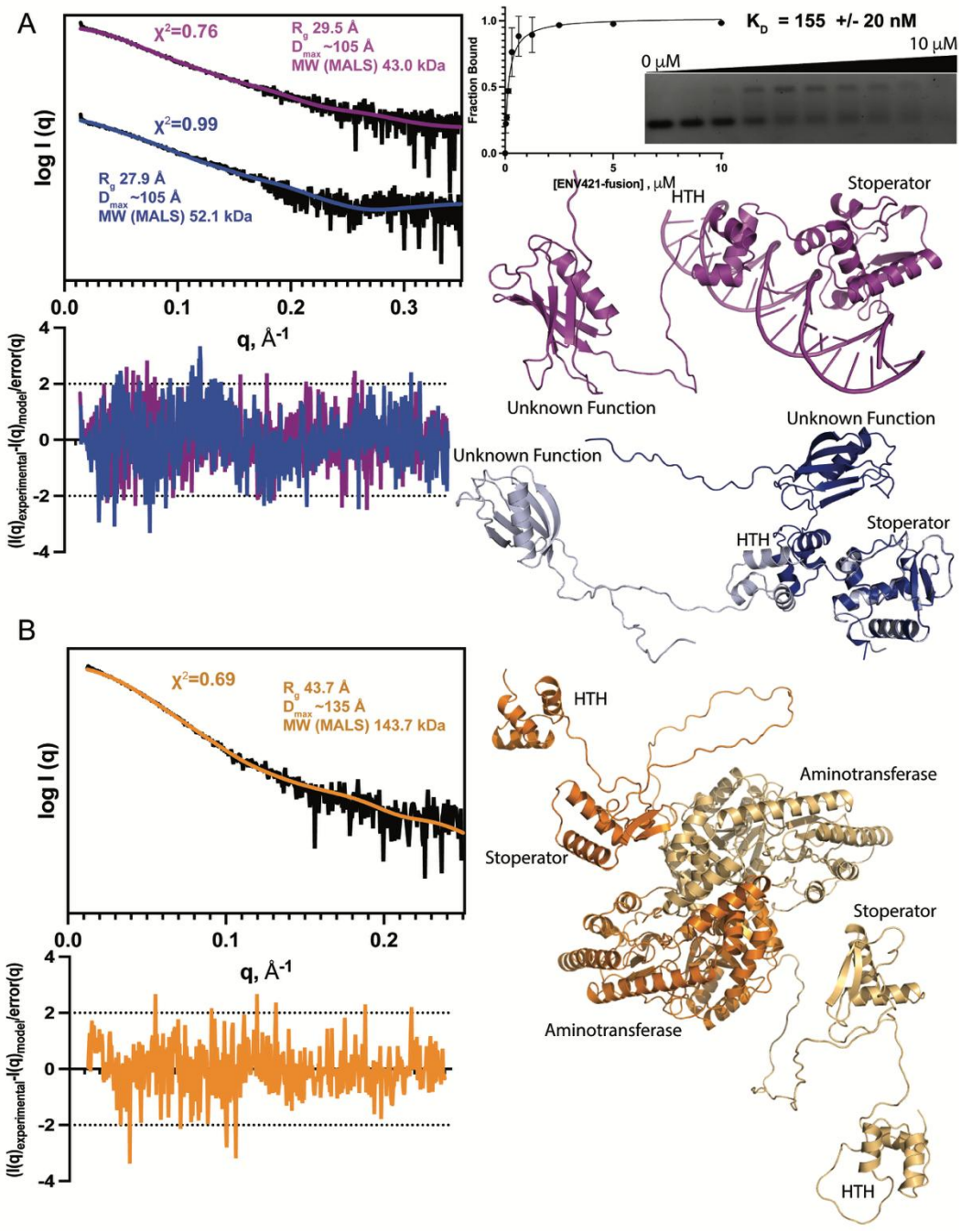


Figure 9: Solution structure and DNA binding properties of repressor fusion proteins
a) SEC-MALS-SAXS analysis of the ENV421 repressor fusion in the absence (blue) and presence (purple) of DNA. The plots on the left show the experimental SAXS curve (black) as well as the theoretical scattering profiles (solid blue and purples lines) of the BILBOMD models shown on the right. χ^2 values for the model fits as well as the residuals of the fits are provided for each sample, along with the calculated SAXS parameters and molecular weights as determined by MALS. The ENV421 fusion protein binds an A2 consensus DNA motif, and shown is the binding curve for the DNA binding assay. The ENV421 protein binds DNA ~3-fold weaker than the Adahisdi and TipsytheTRex repressor proteins. **b**) SEC-MALS-SAXS analysis of the MocR

aminotransferase repressor fusion in the absence of DNA. The plot on the left shows the experimental SAXS curve (black) as well as the theoretical scattering profile (orange) of the BILBOMD dimeric model shown on the right. The monomers of the dimer are colored in light orange and dark orange. The χ^2 value for the model fit as well as the residuals of the fit are provided, along with the calculated SAXS parameters and molecular weight as determined by MALS.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FiguresSupplemental.docx