Computational and Structural Biotechnology Journal 27 (2025) 2051–2062

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research article

Tevenvirinae phages encode a family of LSm-like fold proteins

Jinshil Kim [a], Chin-Hsien Tai [b], Natalie K. Livingston [a,1], Jennifer Patterson-West [a,2], Oliver Stearns [a,3], Bokyung Son [c,*], Deborah M. Hinton [a,*]

[a] Gene Expression and Regulation Section, Laboratory of Biochemistry and Genetics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA
[b] Laboratory of Molecular Biology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA
[c] Department of Food Biotechnology, Dong-A University, Busan 49315, South Korea

ARTICLE INFO

Keywords:
T4
GoF
LSm-fold
Hfq
AlphaFold

ABSTRACT

Uncharacterized bacteriophage proteins typically have little homology outside the phage world. An example is the T4 early protein GoF. Although the function of wild type *goF* is not known, the GoF mutant (D25Y) affects the level of T4 gene *41* mRNA under certain conditions. To investigate possible GoF functions, we leveraged the power of AlphaFold3. We found that despite having very dissimilar sequences, GoF and 2 other uncharacterized T4 early proteins, MotB.1 and Frd.2, are structurally similar with predicted N-terminal LSm-like fold motifs. Since this motif, which is found throughout biology, is frequently associated with an RNA function and the GoF (D25Y) mutation is found within the predicted LSm-like fold, we hypothesized that these proteins may affect gene expression. Consequently, we used a fluorescent translational *mCherry* reporter system and RT-qPCR to investigate if and how the presence of the proteins affect the expression of an *mCherry* gene placed downstream of the T4 gene *41* 5' untranslated region. We find that the heterologous expression of *goF*(D25Y) increases the level of mCherry post-transcriptionally by increasing the stability of the RNA. However, neither WT GoF nor MotB.1 have this effect. We speculate that GoF(D25Y) may represent a gain-of-function mutant that can increase RNA stability. Using AlphaFold3 models we speculate how the D25Y mutation in GoF might facilitate or enhance RNA binding. Our work reveals the power of AlphaFold to find unexpected structure/function relationships among uncharacterized proteins.

1. Introduction

Like complex organisms, simple viruses require the orderly, temporal expression of their genes. In the developmental cycle of bacteriophage T4, a virus that infects *Escherichia coli*, genes are expressed in 3 classes: early, middle, and late. Essential replication proteins, such as DNA polymerase (gene *43*) and DNA helicase (gene *41*), are middle gene products, while morphological proteins are encoded by late genes. However, the functions of many of the early genes, most of which are non-essential under laboratory growth conditions, are unknown [1]. Most of these early genes are thought to encode proteins that establish an optimal infection or facilitate the progression to middle gene expression. Understanding the functions of these proteins may reveal novel mechanisms of host takeover and anti-bacterial strategies. However, common protein BLAST searches [2] typically fail to identify homologs for uncharacterized phage proteins outside of the phage world, making it difficult to assign possible functions [3].

How T4 initiates transcription from its early, middle, and late promoters has been well-studied [1, 4, 5]. However, how T4 RNA is regulated post-transcriptionally is less understood. Previous work has speculated that the T4 early gene *goF* might be involved. This is because T4 does not grow in *E. coli* containing particular mutations in Rho, the major host transcription termination factor [6–8]. Mutations in a T4 gene, called *goF* (or *comCα*), were isolated that restore growth [6–10], and sequencing indicated that one of these mutants has a D25Y substitution within the 141 amino acid, *goF* gene product [11]. An analysis of T4 protein levels in the mutant *rho* infections revealed low amounts of several T4 replication proteins, consistent with poor T4 growth [12].

* Corresponding authors.
E-mail addresses: bkson@dau.ac.kr (B. Son), dhinton@helix.nih.gov (D.M. Hinton).
[1] Center for Engineering in Medicine and Surgery, Massachusetts General Hospital, Boston, MA, USA, 02114
[2] NIAID Office of Research Training and Development, National Institutes of Health, Bethesda, MD, USA, 20892
[3] Northeastern University, Boston, MA, USA, 02115

https://doi.org/10.1016/j.csbj.2025.05.028
Received 30 January 2025; Received in revised form 19 May 2025; Accepted 20 May 2025
Available online 21 May 2025

These proteins are encoded by T4 middle genes, which are expressed before replication begins. For the essential DNA helicase (gene *41*, required for DNA replication), a low level of *41* mRNA was found to be responsible for the poor amount of 41 protein [13]. In particular, more of the RNA ends at a specific site upstream of gene *41* in the *rho* mutant infection than in a WT infection [13,14]. However, in the presence of GoF (D25Y), the correct level of RNA extends into gene *41* [13].

One explanation for the effect of GoF(D25Y) postulated that the *rho* alleles are better transcription terminators, and GoF(D25Y) provides an anti-termination function [8,9]. However, the presence of the unrelated, plasmid-encoded RNA binding protein, Rop, which is not an anti-terminator, also allows T4 to plate on the *rho* alleles. This suggested that GoF(D25Y) and perhaps also WT GoF work post-transcriptionally in a way to protect the RNA [15]. However, determining the function of GoF(D25Y) and WT GoF has been hampered because there have been no protein homologs outside the phage world for comparison.

In this paper, we have investigated GoF and GoF(D25Y) guided by AlphaFold3 predicted structures. We show that *goF* belongs to a family of T4 genes, which also includes the unknown early T4 genes *motB.1* and *frd.2*, and that this family is well-conserved throughout T4-like phages. AlphaFold3 predicts that all three proteins share a similar, N-terminal LSm-like fold domain, a motif that is found throughout biology and is frequently involved in functions associated with RNA [16–18]. For example, the bacterial RNA chaperone Hfq, which facilitates the binding of sRNAs to their mRNA targets as well as binds mRNA substrates that are targeted by nucleases, is a well-characterized LSm-fold protein [18–21]. Using a translational reporter system and RT-qPCR, we demonstrate that the presence of GoF(D25Y) post-transcriptionally increases expression of a reporter gene by increasing the stability of the RNA. Interestingly, WT GoF and MotB.1 only slightly affect expression of the reporter gene in the fluorescence assay. Using AlphaFold3 modeling, we provide a speculative model of how GoF(D25Y) might interact with RNA. To our knowledge this is the first identification of a phage-encoded LSm-like fold family. Since uncharacterized phage proteins typically do not share sequence homologies outside of the phage world, this work provides an example of how AlphaFold3 can be exploited to discover potential functions.

2. Results

2.1. Biological and biochemical characterizations of GoF and GoF(D25Y)

In our initial attempts to understand the function of the WT GoF and GoF(D25Y), we investigated their biological and biochemical properties. First, we estimated the amount of WT GoF during T4 infection by observing its levels in previously published 2-D protein gels after infection of *E. coli B* at 37° C [22]. From a comparison of GoF to another T4 early protein MotB, which we had previously determined to have a concentration of ~ 40,000 molecules per cell at 5 min post-infection [23], we estimate a GoF level of ~20,000 GoF proteins per cell. Previous work has also indicated that GoF is synthesized early, and its synthesis continues even past 5–7 min, when replication begins [11,22]. Thus, GoF is an abundant phage protein during the prereplicative phase, with some protein still remaining later in infection.

The abundance of GoF suggests that the phage expends resources to produce and maintain this protein throughout much of infection. However, previous work has shown that *goF* is located within an 11 kbp region of the T4 genome that can be deleted without an effect on phage growth under normal laboratory conditions [24]. To ask if the loss of only the *goF* gene affects T4 growth, we generated both a T4 deletion in *goF* (Δ*goF*) and a T4 with an amber mutation in *goF* (T4 goF^{am}), in which the first serine codon was mutated to an amber translation stop (site shown in bold italic in Fig. 1A). We compared the efficiency of plating (EOP) for the amber mutant on the nonsuppressing strain *E. coli* NapIV *vs.* the amber suppressing strain *E. coli* NapIV *supD*, which inserts serine at the amber codon. An EOP close to 1 was obtained, indicating that a knock-down of *goF* is not deleterious under our growth conditions (Table S1). When using T4 Δ*goF*, we found that its burst size and latent period were similar to WT T4 in an infection of the *E. coli* strain TOP10 (Table S2).

To investigate whether GoF or GoF(D25Y) can interact with itself or with other factors, we expressed plasmid-borne C-terminal His-tagged *goF* (pBAD33-*goF*-his_6) or *goF*(D25Y) [pBAD33-*goF*(D25Y)-his_6] in the *E. coli* K12 strain TOP10 and then incubated the lysate with Dynabeads His-Tag affinity beads to 'pull-down' the tagged GoF as well as any accompanying species. As seen in Fig. 2, SDS-PAGE in the absence of reducing agent showed the presence of either the GoF-His_6 (panel A) or GoF(D25Y)-His_6 (panel B) as well as multiple protein species present in the clarified supernatant lysate (lanes 2,8). After elution from the beads,

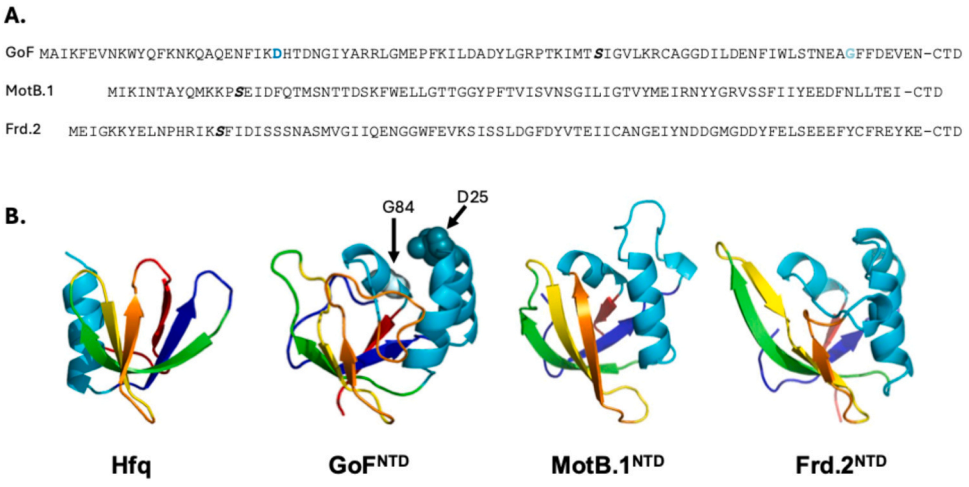

Fig. 1. Sequences and AlphaFold3 predicted structures of GoF^{NTD}, $MotB.1^{NTD}$, and $Frd.2^{NTD}$. (A) Protein sequences of the LSm-like fold NTDs of GoF, MotB.1, and Frd.2 are shown. The serine residues that were mutated to an amber stop in T4goF^{am}, T4$motB.1^{am}$,and T4$frd.2^{am}$ phage are indicated in bold italic, and GoF D25 and G84 are indicated in deep teal bold and light cyan bold, respectively. (B) The top-ranked AlphaFold3 models for GoF^{NTD}, $MotB.1^{NTD}$, $Frd.2^{NTD}$ compared to the structure of a monomer of Hfq [PDB: 1HK9 [70]]. In each case the 5 antiparallel β-sheets, which comprise the LSm-like fold are colored as blue β1, green β2, yellow β3, orange β4, and red β5, and α-helices are shown in cyan. (Two β-sheets going in the same direction comprise β3 in Frd.2.) The mutations in GoF that support T4 growth in Rho026 are shown as the deep teal sphere (D25Y) and the light cyan sphere (G84R) [11].

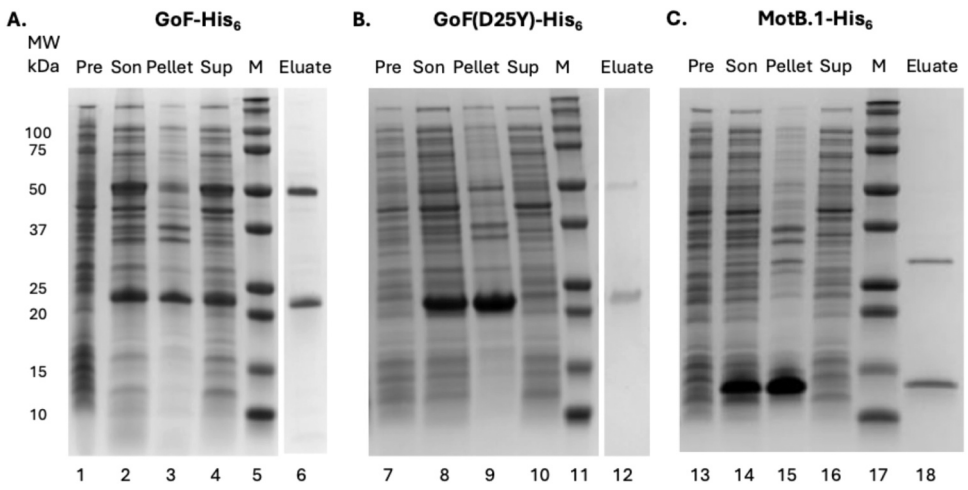

Fig. 2. Representative SDS-PAGE gels (1 of 2) of fractions obtained from *E. coli* TOP10 containing pBAD33-*goF-his*$_6$ (A), pBAD33-*goF*(D25Y)-*his*$_6$ (B), and pBAD33-*motB.1-his*$_6$ (C) before and after elution from Dynabeads His-Tag affinity beads. Lanes 1, 7, and 13, immediately before induction with arabinose (Pre); lanes 2, 8, and 14, sonicated fraction after 2 h of induction (Son); lanes 3, 9, and 15, pellet after centrifugation of sonicated fraction (Pellet); lanes 4, 10, and 16, clarified supernatant after centrifugation of sonicated fraction (Sup) that was then incubated with Dynabeads His-Tag beads; lanes 5, 11, and 17, molecular weight markers [M; Precision Plus Protein Dual Color Standard (Bio-Rad, Hercules, CA, USA)] with sizes in kDa indicated on the far left; lanes 6, 12, and 18 eluate from beads. Lanes 1–6, lanes 7–12, and lanes 13–18 represent lanes from the same gel; the white line indicates where lanes were removed. For GoF(D25Y) (B) and MotB.1 (C) only a trace soluble amount remained in the supernatant after centrifugation of the sonicated fraction.

only 2 dominant bands were observed (lanes 6, 12). One band migrated as expected for GoF-His$_6$ with a molecular weight (MW) of ~23 kDa. [Although the actual MW for GoF-His$_6$ is 17.9 kDa (16.7 kDa for GoF + 1.2 kDa for the tag), previous work has shown that WT GoF migrates abnormally with a size of ~22 kDa [11].] The other band migrated as a species of around 50 kDa, or about 2-fold larger.

After excision from the gel and trypsin digestion, both bands present in the GoF gel were identified by mass spectrometry. In this analysis, the Score SEQUEST HT (Table S3) reflects the level of confidence in the identification, with a score > 10 being considered significant. Table S3 lists the proteins that were identified in either the T4 or the *E. coli* protein databases with scores > 10. GoF is by far the protein with the highest score (in the hundreds) for both the upper and lower bands, confirming that GoF-His$_6$ is the species in the lower band and suggesting that the upper band of ~50 kDa species, observed for both WT GoF and GoF(D25Y), is a dimer.

2.2. *Phylogeny and AlphaFold3 predicted structures suggest that GoF, and two other T4 early proteins, MotB.1 and Frd.2, compose a family of phage LSm-like fold proteins*

Although our biological and biochemical characterizations of GoF indicated that GoF is an abundant T4 protein that may exist as both a monomer and dimer, these analyses failed to provide insight into function. Consequently, we turned to phylogenetic and secondary structure analyses. We were initially guided by deep BLAST searches performed by Dr. Aravind Iyer (NCBI), who suggested that the N-terminal domains (NTDs) of GoF as well as two other uncharacterized T4 early proteins, MotB.1 and Frd.2, shared a homologous β-barrel domain even though conventional BLASTP searches did not indicate any sequence similarity among the proteins (Fig. 1B).

To ask whether *motB.1* and *frd.2* are essential under normal laboratory conditions, we generated T4 *motB.1*^{am}, *frd.2*^{am}, and Δ*motB.1* phage. Like we had seen for *goF*^{am}, we found similar EOPs for the amber mutants on the suppressing *vs.* nonsuppressing NapIV strains (Table S1). Like Δ*goF*, the burst size and the latent period of the Δ*motB.1* phage were similar to those of WT T4 (Table S2). In addition, using our pull-down assay, we found that MotB.1, like GoF and Gof(D25Y), migrates as both a monomer (13.8 kDa) and dimer (~ 28 kDa) after SDS-PAGE in the absence of reducing agent (Fig. 2C, lane 18).

To investigate the conservation of the T4 proteins, we performed 5 runs of PSI-BLAST sequence profile searches against the NCBI ClusteredNR database. This analysis demonstrated that GoF, MotB.1, and Frd.2 are conserved throughout *Tevenvirinae* (T4-like) viruses. Within the generated phylogenetic tree (Fig. 3), the GoF and Frd.2 subfamilies are more related, stemming from a common branch that is separate from that of the MotB.1 subfamily.

To further investigate possible structural similarity among the three T4 proteins, we predicted the structures using AlphaFold2 and then AlphaFold3. In both cases, the N-terminal domains (NTDs), which comprise the first half of each protein, were composed of similar bent, five-stranded, anti-parallel β-sheets, which were connected to C-terminal domains (CTDs) of various structures by flexible linkers (Fig. 1B, Fig. S1). We then performed a Dali server exhaustive search in PDB25, a representative subset of the PDB containing structures that are less than 25 % identical in sequence, to identify proteins similar to the models of GoF^{NTD} (residues 1–91), MotB.1^{NTD} (residues 1–82), and Frd.2^{NTD} (residues 1–88) (Fig. 1). A detailed description of the hits, the Dali Z-score, root mean square deviation (RMSD), and number of aligned residues are listed in Table S4. GoF, Frd.2, and MotB.1 were the most closely related, and similar to the result from sequence phylogeny, GoF and Frd.2 were again derived from a single branch. Several protein structural hits that appeared in all 3 phage protein searches are shown in bold in Fig. 4.

A group of proteins containing SH3-like topologies [25] and related to metal regulation/transport was predicted to be the closest relatives to the T4 family (Fig. 4). This group includes ferrous iron transport protein A (FeoA in *E. coli*), whose function is unknown, but may interact with the Fe^{2+} transmembrane transporter component FeoB [26]. Other relatives in this set are DNA-binding regulators that are also associated with metal usage, such as *Streptococcus mutans* SloR [27] and mycobacterial IdeR [28,29]. However, these DNA-binding proteins contain multiple histidine residues needed to coordinate metal ion, a motif which GoF lacks.

Another close relative is Hfq, the major bacterial RNA chaperone, which contains an LSm-fold. This finding was particularly intriguing given that GoF(D25Y) affects the level of the T4 *41* mRNA [13]. In addition, besides Hfq, numerous other LSm-fold proteins found throughout biology serve as RNA chaperones [16–18]. The predicted NTD structures (Figs. 1B) and 2D topology diagrams of GoF, Frd.2, and MotB.1 exhibited the 5 anti-parallel β-strand motif seen in these

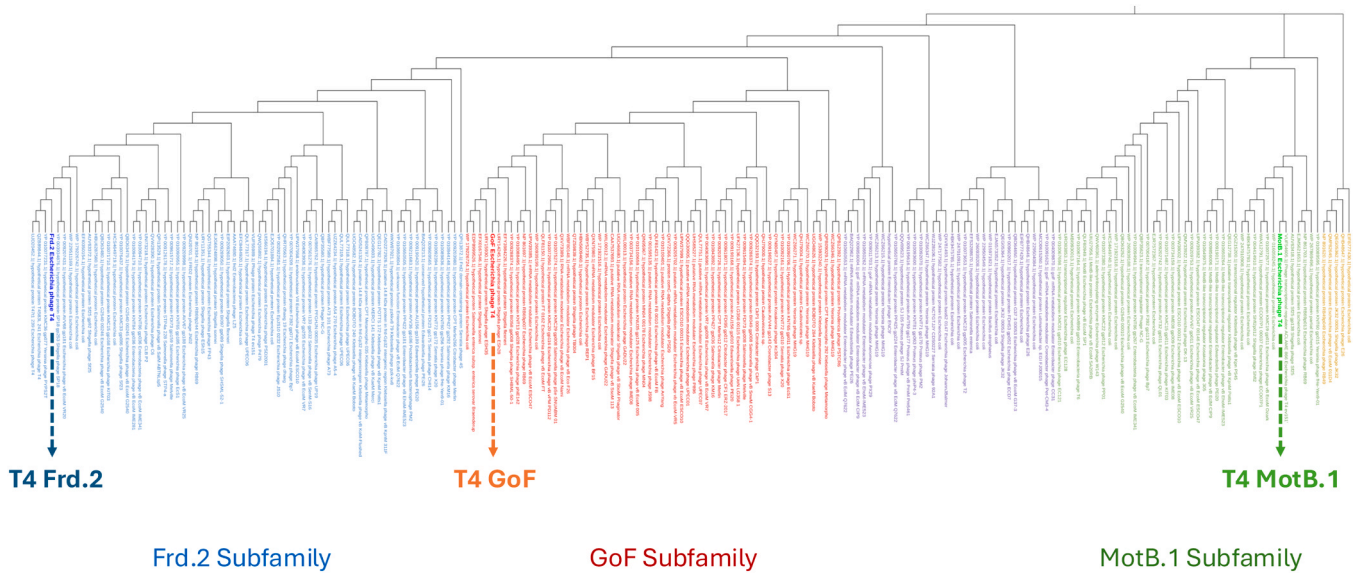

Fig. 3. Phylogenetic tree for GoF, MotB.1, and Frd.2. PSI-BLAST sequence profile searches against the NCBI ClusteredNR database were performed for T4 GoF^{NTD}, T4 MotB.1^{NTD}, and T4 Frd.2^{NTD}, and multiple sequence alignment was used to construct the phylogenetic tree as described in Materials and Methods. Subfamilies containing GoF^{NTD}, MotB.1^{NTD}, and Frd.2^{NTD} are colored in red, green, and blue, respectively. The subfamily that is shared by GoF^{NTD} and Frd.2^{NTD}, but is a different branch from MotB.1^{NTD}, is colored in purple, while another separate subfamily is colored in orange.

proteins, which we will refer to as an LSm-like fold (shown as pink arrows in Fig. S2). While most LSm-fold proteins, including Hfq, have an N-terminal α-helix, it should be noted that this is not a requirement. The structure of an uncharacterized LSm-fold protein, possibly from cyanophage, exhibits the 5 anti-parallel β-strand structure, but with a C-terminal α-helix [30].

Two mutations within GoF have been shown to allow T4 growth on *rho* alleles that restrict WT T4 growth [11]: D25Y (dark teal sphere in GoF in Fig. 1B) and G84R (light cyan sphere in GoF in Fig. 1B). Both of these are located within the predicted LSm-like fold region and on the same face of the predicted GoF structure. However, while the D25Y substitution is known to increase the level of T4 gene *41* RNA [13], there is no direct evidence linking the G84R substitution to RNA levels.

LSm-fold proteins frequently interact as dimers/multimers [16]. To test the possibility of the T4 LSm-fold proteins forming homodimers or homohexamers, we submitted both the untagged and His_6-tagged GoF, GoF(D25Y) and MotB.1 sequences to the AlphaFold server. However, none of the top 5 complex dimer or hexamer models had ipTM scores > 0.3 or pTM scores > 0.4, indicating that AlphaFold3 does not have confidence in their homomultimer formation. Consequently, we were unable to generate reliable homomultimer models for the proteins.

2.3. *Overexpression of goF(D25Y) increases the level of a reporter mCherry in a translational fusion assay*

Previous work has suggested that GoF(D25Y) increases the level of both *41* mRNA and 41 protein under certain conditions [13]. Given this finding and the fact that AlphaFold3 predicted that the T4 proteins have an LSm-like fold, which can be associated with RNA, we employed a tandem fluorescence reporter assay [31] to investigate whether GoF (D25Y), WT GoF, or MotB.1 affects gene *41* expression post-transcriptionally. In this system, WT *goF*, *goF*(D25Y), or *motB.1* were present on a plasmid and placed downstream of the arabinose-inducible promoter P_{BAD} (Fig. 5A). In the chromosome, a strong constitutive promoter (P_{con}) was located upstream of 2 reporter genes, *mCherry* followed by *gfp* (Fig. 5A). *gfp* had its own RBS (RBS2, Fig. 5B) while various RBS-containing sequences were placed upstream of mCherry for testing (Fig. 5B).

Because previous work had shown a significant effect of GoF(D25Y) on T4 gene *41* expression [13], we used the 358 bp region upstream of T4 *41* (*41* 5'UR) as well as just the 20 bp immediately upstream of 41 [41 5'UR 20 (nucleotides) nt], since these regions contain the natural sequence present in the T4 genome. However, because we knew that GoF(D25Y) can affect the levels of several proteins [12], we also tested other general RBSs: RBS1 and RBS2 (20 nt). In addition, for the 41 5'UR 20 nt and the RBS2 (20 nt), we tested both a GUG translation start, which is present in gene *41*, and the more common AUG.

To quantify our results, we measured the growth of the cells as monitored by OD_{600} and the levels of mCherry and GFP fluorescence (Fig. S3-S5). We then normalized the fluorescence relative to the OD_{600} and determined the ratio of the fluorescence observed with the induced plasmid relative to the uninduced plasmid. Thus, if production of the T4 protein increased the level of mCherry fluorescence, without a corresponding change in the level of GFP fluorescence, we could conclude that *mCherry* expression is altered post-transcriptionally. This could arise from a change in the stability of the RNA and/or by a change in translation.

As expected, addition of arabinose to cells containing the empty vector did not appreciably change the normalized level of mCherry fluorescence in any construct (Fig. 6A and B and S6). In contrast, overexpression of *goF*(D25Y) increased the mCherry fluorescence dramatically with the *41* 5'UR construct (Fig. 6A). After a 2 h induction, there was a ~2.6-fold increase in mCherry fluorescence in the presence of GoF(D25Y). This increase was observed without a corresponding increase in GFP fluorescence (Fig. 6A). In fact, the level of GFP fluorescence was somewhat lower after induction. These results are consistent with the idea that the combination of GoF(D25Y) and the *41* 5'UR markedly increases mCherry fluorescence post-transcriptionally.

While there was a dramatic increase in mCherry fluorescence from the *41* 5' UR-*mCherry* construct with GoF(D25Y), we saw a much smaller effect with WT GoF or MotB.1 (Fig. 6A) even though WT GoF, MotB.1, and GoF(D25Y) were produced at similarly high levels (Fig. S6A). Thus, we conclude that the D25Y mutation imparts a change in the activity of GoF in this assay.

Interestingly, although GoF(D25Y) showed a strong effect on mCherry fluorescence when using the 41 5' UR, our assays indicated that GoF(D25Y) also increased the mCherry fluorescence somewhat when using other RBSs: RBS1 (Fig. 6B), 41 5'UR (20 nt, AUG or GUG,

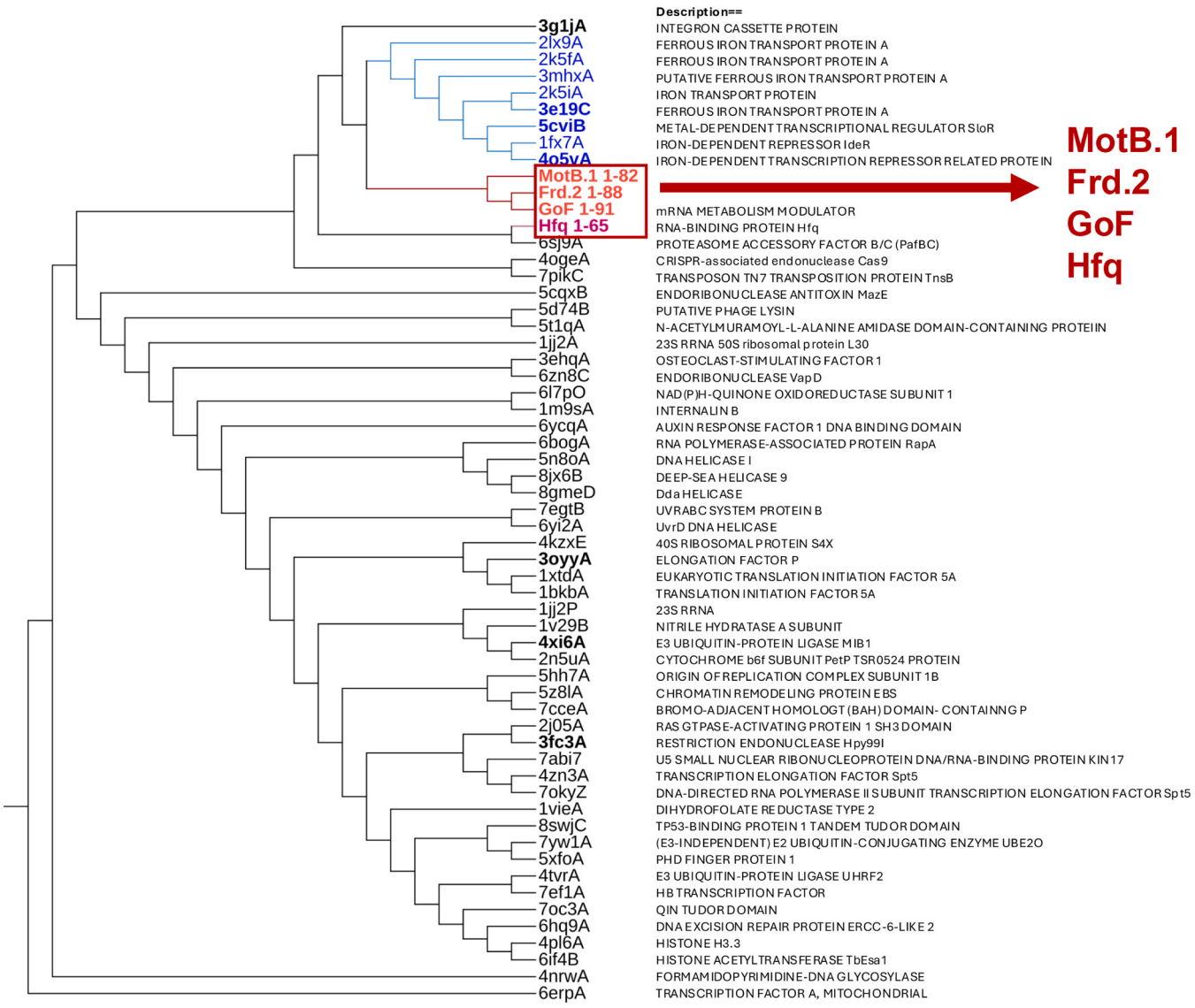

Fig. 4. Structural similarity dendrogram of GoF^{NTD}, MotB.1^{NTD}, Frd.2^{NTD}, Hfq, and their structural neighbors in PDB25. An average linkage clustering of 55 similar structures in PDB25 found by the Dali Server, along with GoF^{NTD} (residues 1–91, red), MotB.1^{NTD} (residues 1–82, red), Frd.2^{NTD} (residues 1–88, red), and Hfq (residues 1–65, dark red) is shown here. The branch closest to GoF^{NTD}, MotB.1^{NTD}, and Frd.2^{NTD} consists of 8 proteins, shown in blue, which have the SH3-topology domain present in FeoA (ferrous iron transport protein A) (CATH classification 2.30.30.90). Hits that appeared in the searches with all 3 of the T4 proteins are in bold. A detailed description of the hits and the Dali Z score, RMSD, and number of aligned residues are listed in Table S5 (GoF-MotB.1-Frd.2-DaliPDB25hits.xisx).

Fig. S7A), or RBS2 (20 nt, AUG or GUG, Fig. S7B). Furthermore, GoF (D25Y) increased the mCherry fluorescence with RBS2 (20 nt)-*mCherry,* but no increase in GFP fluorescence was seen by GoF(D25Y) with RBS2-*gfp* (Fig. 6 and Fig. S7). It is known that RNA structure contributes to mRNA stability and translation [32], so this difference may reflect the overall structures and stabilities of the *mCherry vs. gfp* mRNAs.

Taken together, our translational reporter assays indicated that GoF (D25Y), but not WT GoF or MotB.1, considerably increases the post-transcriptional expression of mCherry when *41* 5'UR is present upstream. In addition, a modest increase is observed with GoF(D25Y) and the other *mCherry* constructs, suggesting that GoF(D25Y) may provide a general increase in mCherry fluorescence that is independent of the specific UR sequence.

2.4. GoF(D25Y) increases the stability of the 41 5'UR-mCherry RNA

One way that a protein can increase the level of an RNA post-transcriptionally is by increasing the stability of the RNA. To investigate this possibility for GoF(D25Y), we observed the half-lives of the *mCherry* and *gfp* RNAs in the *41* 5'UR-mCherry strains containing either the vector pBAD33, pBAD33-*gof*, or pBAD33-*goF*(D25Y). At 2 h post-induction, we added rifampicin to stop transcription, and then we followed the decay of the *mCherry* and *gfp* transcripts by RT-qPCR using primers for the *mCherry* and *gfp* RNA (Fig. 7A). At the earliest time point (2 min), the presence of either GoF or GoF(D25Y) was associated with an increase in mCherry RNA stability over that seen with the vector. However, at the later time points (4, 6, and 8 min), the *mCherry* RNA was significantly more stable in the presence of GoF(D25Y) *vs.* the vector or WT GoF. In contrast, the stability of *gfp* RNA was similar when comparing the vector *vs.* GoF(D25Y) and lower with GoF. While we do not understand the decreased stability of the *gfp* RNA with GoF, these results are consistent with the idea that GoF(D25Y) specifically increases the stability of the *mCherry* RNA in this system.

We would expect that an increase in stability should result in an overall higher steady-state level of the *mCherry* RNA with GoF(D25Y). To investigate we used RT-qPCR to compare the FC (induced

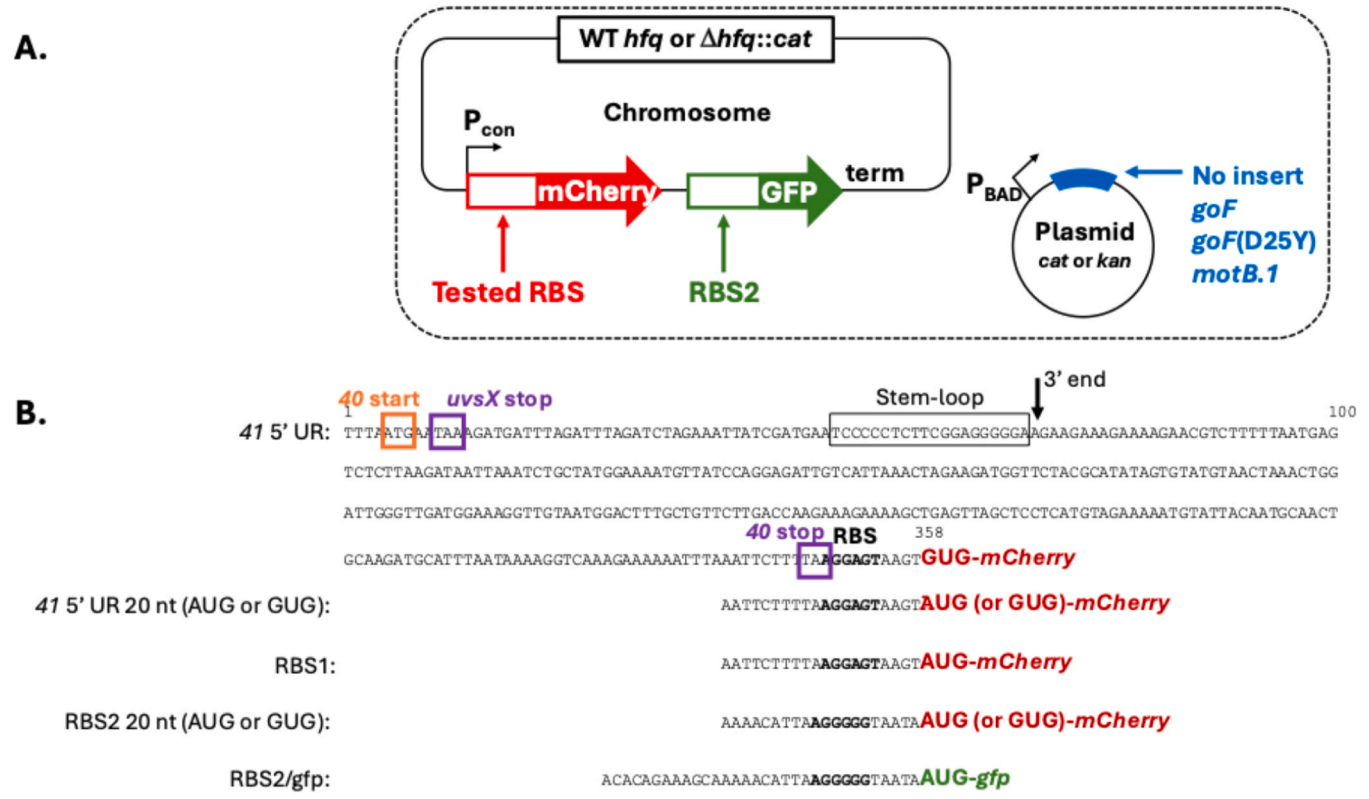

Fig. 5. Fluorescence reporter assay. (A) Schematic of the translation reporter system modified from [31]. The chromosome contains a strong synthetic constitutive promoter P_{con} that drives transcription of *mCherry* and *gfp* until reaching the terminator (term) and either WT *hfq* or a deletion of *hfq* replaced by the chloramphenicol-resistant gene *cat* (Δ*hfq*::*cat*). The region upstream of *mCherry* contains one of the sequences shown in (B), while the region upstream of *gfp* contains RBS2 shown in (B). The open reading frame (ORF) in the *41* 5'UR construct begins with a 'GUG' since this is the start of the gene *41* ORF. Expression of *goF*, *goF* (D25Y), or *motB.1* arises from a plasmid containing the arabinose-inducible P_{BAD} promoter and either a *cat* or kanamycin-resistant (*kan*) gene. **(B)** Sequences. In the *41* 5'UR, the start codon for the upstream T4 gene *40* is boxed in orange, the stop codons for T4 gene *uvsX* and gene *40* are boxed in purple, the position of the 3' end observed *in vivo* in a T4 infection is indicated, and the stem-loop immediately upstream of the 3' end is boxed in black. The ribosome binding site (RBS) present in each sequence is shown in bold, the AUG (or GUG) starts for *mCherry* and *gfp* are in red and green, respectively.

/uninduced) for *mCherry* and for *gfp* 2 h post-induction using pBAD33, pBAD33-*goF* or pBAD33-*goF*(D25Y). As expected, this analysis indicated that the steady-state level of *mCherry* RNA was higher when GoF(D25Y) was present (Fig. 7B).

Taken together, these results indicate that GoF(D25Y) increases mCherry fluorescence seen with *41* 5' UR-*mCherry* by increasing the stability of the *mCherry* RNA, which in turn leads to a higher level of steady-state *mCherry* RNA.

2.5. Hfq also affects the level of mCherry post-transcriptionally

Hfq, the major bacterial RNA chaperone, is known to bind certain RNAs, targeting them for degradation [33]. Given that the N-terminal regions of the T4 LSm-like fold proteins are predicted to be structurally similar to Hfq, we investigated if GoF(D25Y) increased the stability of *mCherry* RNA by interfering with an Hfq-mediated process. For example, the LSm-fold of Hfq could bind the RNA, leading to destabilization, while the LSm-like fold of GoF(D25Y) could bind the RNA to protect it from Hfq. To test this hypothesis, we generated a Δ*hfq*::*cat-sacB* in both the *41* 5'UR and the RBS1 strains. Thus, the chromosomal *hfq* gene was deleted and replaced by the *cat* gene, providing chloramphenicol resistance. We also modified the pBAD33 plasmids to replace the *cat* gene with kan^r (Fig. 6A). Using this system, we again tested the mCherry and GFP fluorescence as before.

In the absence of induction or in the presence of the vector, we observed a significant increase in mCherry fluorescence in both the Δ*hfq* *41* 5'UR-*mCherry* and the Δ*hfq* RBS1-*mCherry* strains relative to the hfq^+ strain while no effect was observed on the level of GFP (Fig. S8). Thus, Hfq does affect the mCherry RNA post-transcriptionally, perhaps by targeting it for degradation. Despite this fact, when we normalized for this difference (Fig. 8) we still observed a similar, strong effect of GoF (D25Y) with *41* 5'UR-*mCherry* (Fig. 6C). Again the levels of GoF, GoF (D25Y), and MotB.1 were high in the Δ*hfq* strains after induction (Fig. 6C and D). We conclude that Hfq does affect the level of the *mCherry* RNA, but the increase that we observe with GoF(D25Y) is not related to a process that involves Hfq.

2.6. Modeling of GoF and GoF(D25Y) in complex with RNA using AlphaFold3

Taking advantage of the latest feature in AlphaFold3, we examined models of GoF^{NTD} and $GoF(D25Y)^{NTD}$ in complexes with various RNA constructs [*41* 5'UR 20 nt, RBS1, RBS2 (20 nt), and RBS2 (Fig. 5B)]. We found that all the $GoF(D25Y)^{NTD}$-RNA complexes had higher ipTM scores (interface predicted template modeling, a measure of the accuracy of the predicted relative positions of the entities in the complex) than the WT GoF^{NTD}-RNA complexes, implying that $GoF(D25Y)^{NTD}$ is more likely to interact with those RNAs than WT GoF^{NTD} (Table S5 and Fig. S9). Although many complexes had ipTM scores below the threshold of 0.6, all the models predicted that it is the loop regions in the NTD that interact with RNA.

To investigate the specific effect of the D25Y substitution, we performed pairwise comparison of the models of WT GoF^{NTD} and GoF $(D25Y)^{NTD}$ on Dali Server and observed the models in ChimeraX. As shown in Fig. 8A-C, this analysis indicates that the mutation would introduce conformational changes and a misalignment in the loop

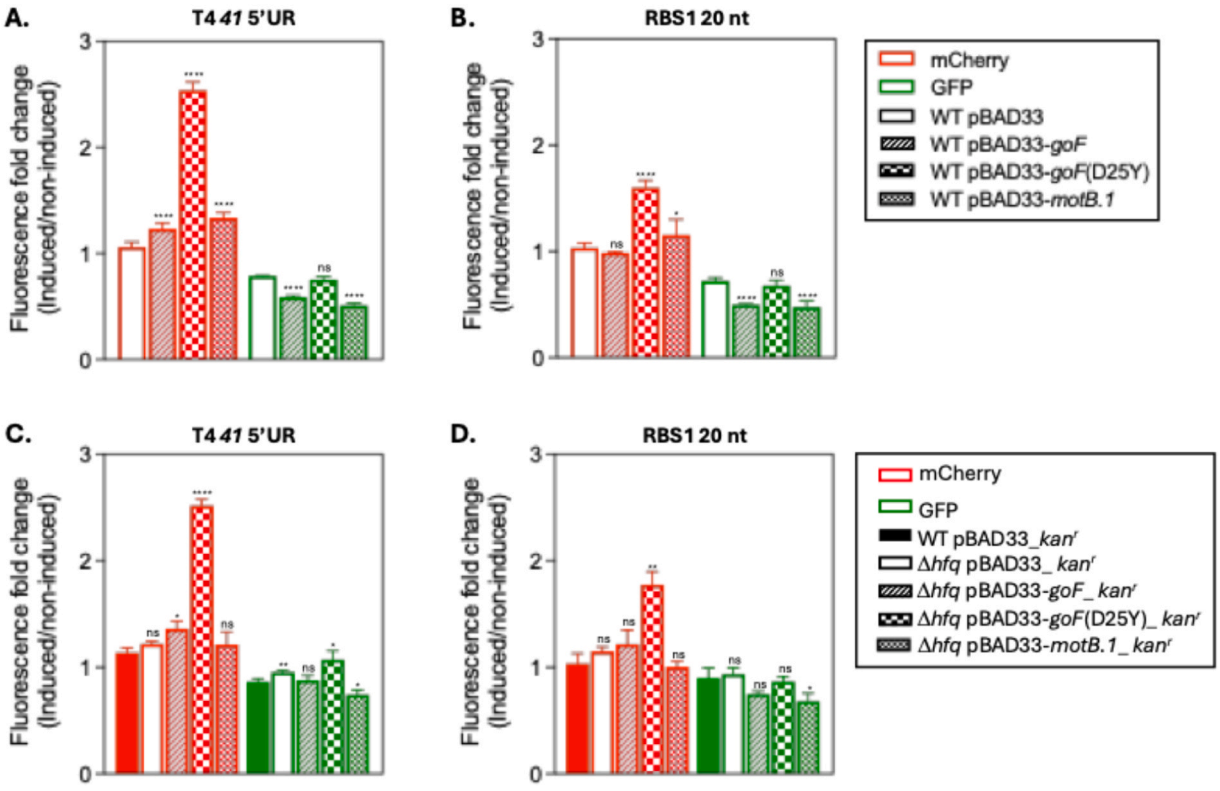

Fig. 6. GoF(D25Y) markedly increases mCherry fluorescence in the *41* 5'UR-*mCherry* construct in the presence or absence of Hfq. Relative levels of mCherry (red) and GFP (green) fluorescence (induced /uninduced) are shown for the *41* 5'UR (A), RBS1 (B), Δ*hfq 41* 5'UR (C), or Δ*hfq* RBS1 (D) strains containing the vector (pBAD33) or the pBAD33 plasmids expressing *goF*, *goF*(D25Y), or *motB.1* at 2 h post-induction. Identification of the plasmids is given in the box. Values were obtained from at least three biological replicates with standard deviations indicated. Statistical analysis was conducted by an unpaired two-tailed Student's t-test. [Not significant, ns, ($P > 0.05$); *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$].

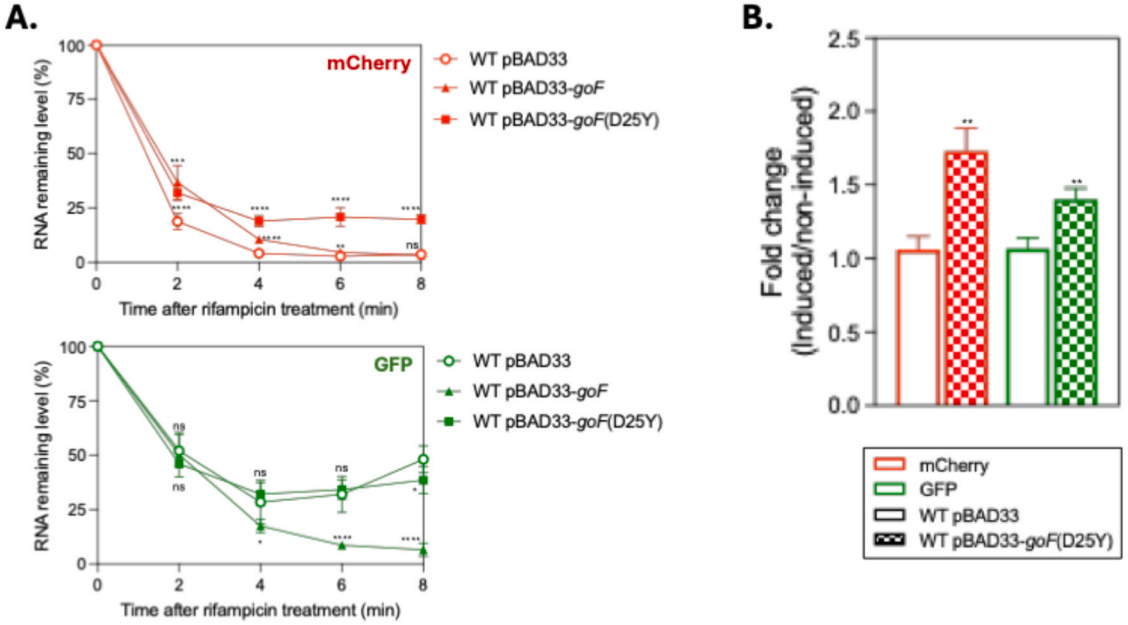

Fig. 7. GoF(D25Y) increases the half-life of *mCherry* RNA in the *41* 5'UR-*mCherry* construct. (A) The % of *mCherry* or *gfp* RNA remaining in the *41* 5' UR-*mCherry* strain containing either pBAD33, pBAD33-*goF*, or pBAD33-*goF*(D25Y) 2 h post-induction after transcription was stopped by the addition of rifampicin for the time indicated. (B) The relative level (induced/uninduced) of *mCherry* and *gfp* RNA in the *41* 5'UR-*mCherry* strain containing either pBAD33 or pBAD33-*goF*(D25Y) 2 h post-induction.

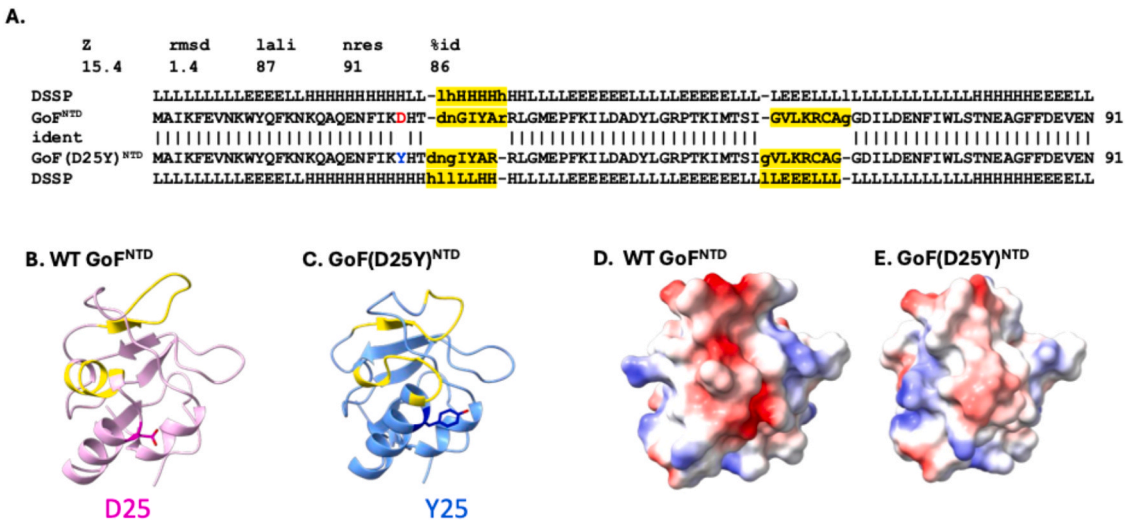

Fig. 8. AlphaFold3 models of GoF^{NTD} and GoF(D25Y)^{NTD}, their electrostatic potential, and the sequence alignment based on structural superposition by Dali pairwise comparison. (A) Results of Dali pairwise comparison between AlphaFold3 models of GoF^{NTD} and GoF(D25Y)^{NTD}, including the Z score, RMSD, number of aligned residues, % identity, and sequence alignment, which is based on the structural superposition. The secondary structure definitions by DSSP (H for helix, E for sheet, and L for coil) are shown above or below the amino acid sequences. Aligned and identical residue pairs are matched by vertical lines and are in uppercase while the unaligned regions are in lowercase and highlighted in yellow in both the sequence alignment and the structural models in (B) and (C) below. (B) and (C) AlphaFold3 models of WT GoF^{NTD} and GoF(D25Y)^{NTD}, respectively. GoF^{NTD} is in pink with the D25 side chain in magenta, while GoF(D25Y)^{NTD} is in corn blue with the Y25 side chain in darker blue. In both structures, the misalignment in the loop regions is shown in yellow. (D) and (E) The Coulombic electrostatic potential (ESP) of GoF^{NTD} and GoF(D25Y)^{NTD}, respectively calculated by UCSF ChimeraX [69], with red for negative potential through white to blue for positive potential. Structures are shown in the identical orientation as in (B) and (C).

regions (shown in yellow), resulting in an RMSD of 1.4 angstroms between the two models. Calculating the Coulombic electrostatic potential, we found that such sequence and structural changes also affect the electrostatic potential on the protein surface, making the surface more positive on GoF(D25Y)^{NTD} than on WT GoF^{NTD} (Fig. 8D, E). Thus, AlphaFold3 predicts that the single D→Y substitution within GoF results in a more global structural change. We speculate that this change might result in RNA binding or in a function that can affect RNA stability indirectly.

3. Discussion

Approximately one third of the T4 genome is expressed from early genes within the first minute of infection [1]. Many of these early genes are uncharacterized, nonessential under normal laboratory conditions, and are thought to be involved in phage takeover of the *E. coli* host cell. However, identification of their functions has been hampered by the lack of homologs outside the phage world in the protein sequence database [3]. Thus, the standard approach of finding a sequence homolog or motif to direct one toward possible functions is not applicable for many of these genes.

One such gene is T4 *goF*, which encodes a 16.7 kDa protein [11]. Although the function of GoF is unknown, a mutant GoF(D25Y) has been shown to increase the levels of both the T4 gene *41* RNA [13] and 41 protein [12] under certain conditions. It was first postulated that GoF (D25Y) did this by being a T4 anti-terminator of transcription [6–8]. However, other work suggested that GoF(D25Y) worked post-transcriptionally [15, 34, 35].

Using a tandem fluorescence translational reporter assay with a reporter *mCherry* gene [31], we have shown that GoF(D25Y) works post-transcriptionally, rather than transcriptionally (Fig. 6). This increase is most striking when using the entire 358 nt gene *41* 5' UR, but we also see a modest increase with other upstream untranslated regions (Fig. 6A). Furthermore, we show that GoF(D25Y) works by increasing the half-life of the RNA (Fig. 7A). Although Hfq is known to play a major role in RNA degradation [33], we find that the effect of GoF(D25Y) does not involve Hfq (Fig. 6C). We conclude that GoF(D25Y) either protects the RNA by directly binding to it or it works indirectly by interfering with nucleases and/or by increasing translation, which can also increase RNA stability [36]. Additional work is needed to distinguish among these possibilities.

We do not see a strong post-transcriptional effect using either WT GoF or MotB.1, indicating that the D25Y mutant imparts a significant change in GoF activity. We speculate that the aspartic to tyrosine substitution within the predicted LSm-like fold of GoF generates a gain of function mutation that imparts the ability to increase the stability of the RNA. Understanding how gain of function happens is of utmost importance, as such mutations are involved in dynamic biological processes, such as evolution [37,38] and tumor progression [39]. Recent work indicates that these mutations need to be considered in the context of structure, not sequence [40]. However, previous studies have shown that predicting the structural and functional effects of a point mutation are challenging [41–43]. Furthermore, only a limited number of CASP (Critical Assessment of Structure Prediction) targets of wild type and mutants have been evaluated to assess how well the prediction programs perform [44]. Consequently, the dramatic conformational changes introduced by the D→Y substitution in GoF, which predicts more positive electrostatic potential on the protein surface, taken together with our supporting experimental observation, are encouraging. Of course, structure determination and direct evidence of function is needed to confirm this prediction. Going forward, the evaluations of additional mutant structures and protein-nucleotide complexes in CASP should aid the development of better programs.

4. Materials and methods

4.1. Strains and phage

E. coli B strains were NapIV [BE strain; *hsdR*$_k$ *hsdMk*$^+$ *hsdSk*$^+$ *thi*] [45], NapIV *supD* [45], and BL21 [*fhuA2 [lon] ompT gal [dcm] ΔhsdS*] [46]. *E. coli* K12 TOP10 [*F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL(Str*^{R}*) endA1 nupG*] was from Thermo Fisher Scientific (Waltham, MA), *E. coli* DH5α [*F' φ80lacZΔM15 Δ(lacZYA-argF)U169 recA1 endA1 hsdR17(*r_K^-*,* m_K^+*)*

phoA supE44 λ⁻thi-1 gyrA96 relA1] [47,48] was used for construction of T4 deletion phages, and MG1655 (*F⁻ λ⁻ ilvG⁻ rfb-50 rph-1*) was a gift of M. Cashel (National Institutes of Health). The fluorescent reporter strain RS41, containing the T4 *41* 5'UR region (from −358 to −1 relative to the translation start site of T4 gene *41*), was constructed by modifying the strain JC1283 [31] through recombineering [49,50] using a PCR product amplified from T4 DNA with primers CP26–40-F and mCh-40-R (Table S6), followed by selection on M63 glycerol sucrose agar [31]. The other fluorescent reporter strains [with 20 nt of T4 *41* 5'UR region or RBS2 (20 nt) having a start codon of AUG or GUG] were constructed similarly using PCR products amplified from the oligonucleotides s41-AUG, s41-GUG, StrRBS-AUG, or StrRBS-GUG with primers CP26-amp-F2 and mCherry-amp-R2 (Table S6). Sequencing by Psomagen (Rockville, MD, USA) using primers lacI-f and JC129 (Table S6) verified the gene reporter fusions. The previously constructed strains JC1245 (wild type *hfq*) and JC1263 (Δ*hfq*) [31], contain RBS1, a synthetic consensus ribosome-binding sequence, upstream of *mCherry*. To obtain an Δ*hfq* version of the *41* 5'UR reporter strain, Δ*hfq*::*cat-sacB* was introduced into RS41 via P1 transduction, using the P1 lysate from AZ234 [31]. Strains were transformed with the various pBAD33 or pBAD33_*kan*^{r} plasmids as indicated.

Amber mutations in *goF*, *motB.1*, and *frd.2* were recombined into WT T4D$^+$ using a modification of a previously described procedure [51]. Briefly, BL21(DE3) [52] cultures containing pTE103-*goF*^{am}, pTE103-*motB.1*^{am}, or pTE103-*frd.2*^{am} were infected with WT T4 at a multiplicity of infection of ~1. The resulting lysate contained WT T4 as well as the amber mutant phage that was generated by recombination between WT T4 and the plasmid. Lysates were titered on *E. coli* NapIV *supD,* whose amber suppressing mutation incorporates serine (the wild-type amino acid) at the amber stop codon. Phage plaque lifts on nitrocellulose were hybridized with 5'-^{32}P labeled oligonucleotides (GoF-amber, GoF-WT, MotB.1-amber, MotB.1-WT, Frd.2-amber, Frd.2-WT, Table S6) containing either the WT or the mutant amber sequence. Single plaques that hybridized specifically with the mutant sequence were grown in NapIV *supD* and sequenced (Psomagen) to confirm the presence of the mutation.

Construction of the T4 Δ*goF* and T4 Δ*motB.1* phages was performed as described previously [53,54]. Briefly, WT T4 phage was added to *E. coli* DH5α that had been previously transformed with the spectinomycin-resistant CRISPR-Cpf1 spacer plasmids and the kanamycin-resistant pET28b donor plasmids and then incubated for 7 min at 37° C. LB top agar (3 mL, containing 0.7 % agar and 50 μg/mL spectinomycin) was added and the solution was poured onto LB plates supplemented with 50 μg/mL spectinomycin. After incubation overnight at 37° C, single plaques were transferred into 1.5 mL microcentrifuge tubes containing LB broth and used to infect *E. coli* DH5α. This process was repeated three times to ensure purity. To confirm the deletion of *goF* or *motB.1* in the T4 genome, single plaques were subjected to DNA sequencing. One microliter of phage lysate was denatured at 95° C for 7 min and used as a template for PCR with Taq DNA polymerase (New England Biolabs, Ipswich, MA, USA) and primers GoF-con-F/R or MotB.1-con-F/R (Table S6). The amplified DNA fragments were purified using a QIAquick PCR Purification Kit (Qiagen, Hilden, Germany), and the sequences were confirmed by Sanger sequencing (Psomagen, Rockville, MD, USA).

4.2. DNA

Unless otherwise indicated, gene synthesis, plasmid construction, and DNA sequencing of the constructed plasmids were performed by GenScript (Piscataway Township, NJ, USA). For pBAD33-*gof* and pBAD33-*motB.1, goF* and *motB.1*, respectively, were cloned downstream of the arabinose-inducible promoter P_{BAD} between the *Kpn*I and *Sal*I sites of the pACYC-based expression vector pBAD33 [55]. In pBAD33-*goF* (D25Y) the codon for GoF residue D25 was mutated from GAC to TAT. His_6-tagged derivatives were constructed similarly with the native stop codon being replaced with a C-terminal His_6-tag (5'caatccggctctcatcaccatcaccatcactaa3').

pTE103-*goF*^{am}, pTE103-*motB.1*^{am}, and pTE103-*frd.2*^{am} were generated by cloning the following mutated bacteriophage T4 sequences (NC_000866.4) surrounding and including *goF*, *motB.1*, and *frd.2*, respectively, between the *Bam*HI and *Sal*I sites of pTE103 [56]: for *goF*, 5881–6300 with the 5'GG3' at 6094–6095 mutated to 5'CT3'; for *motB.1*, 7800–7441 with the 5'AG3' at 7609–7610 mutated to 5'CT3'; for *frd.2*, 146281–146640 with the 5'AG3' at 146482–146483 mutated to 5'CT3'. In each case this sequence change converted the first serine codon in the gene to a TAG (amber) stop codon.

To replace the chloramphenicol resistance gene (*cat*) with the kanamycin resistance gene (*kan*^{r}) in pBAD33, pBAD33-*goF*, pBAD33-*goF* (D25Y), and pBAD33-*motB.1* (generating pBAD33_*kan*^{r}, pBAD33-*goF*_*kan*^{r}, pBAD33-*goF*(D25Y)_*kan*^{r}, and pBAD33-*motB.1*_*kan*^{r}, respectively), *E. coli* MG1655 harboring the plasmid pSIM6 [57], which expresses the lambda red recombinase upon induction at 42° C, was transformed with pBAD33, pBAD33-*goF*, pBAD33-*goF*(D25Y), and pBAD33-*motB.1*. At mid-log phase (OD_{600}~0.5), cells were shifted from 32° C to 42° C, grown for another 15 min, and then electroporated with a DNA fragment (1 μg) containing the *kan*^{r} gene. This fragment was obtained by PCR generation of a *kan*^{r} DNA cassette using the plasmid pNW129 [58] and primers Kan-cat-F and Kan-cat-R (Table S6) followed by treatment with *Dpn*I (New England Biolabs, Ipswich, MA, USA) to eliminate the template plasmid. After electroporation with the DNA, cells were grown at 37° C for 1 h with shaking (250 rpm) and then incubated at room temperature overnight. The cells harboring recombinant plasmids were selected on LB with 40 μg/mL kanamycin.

To construct the kanamycin-resistant donor plasmids pET28b-*goF* and pET28b-*motB.1* for making the T4 Δ*goF* and T4 Δ*motB.1* mutants, the regions spanning −149 to + 1 and 150 bp downstream of *goF* or 150 bp upstream and downstream of *motB.1*, respectively, were cloned into the *Bam*HI and *Bgl*II sites of pET28b (EMD Millipore, Burlington, MA, USA). Spectinomycin-resistant CRISPR-Cpf1 spacer plasmids targeting *goF* or *motB.1* were constructed using the CRISPR-LbCas12a/SpCas9 system [54] and the *goF* spacers, goF 1 and goF 2, or the *motB.1* spacers, motB.1 1 and motB.1 2, respectively (Table S6).

4.3. Phylogeny and AlphaFold analyses

To generate the phylogenetic tree, five runs of PSI-BLAST [59] sequence profile searches were performed for GoF, MotB.1, and Frd.2 against the NCBI ClusteredNR database. From this analysis, 210 unique sequences were collected and aligned with MUSCLE [60]. Multiple sequence alignment was then used to construct the tree with Maximum Likelihood in MEGA11 [61] and displayed in iTOL [62] v6.0.

Protein structures were predicted for GoF, MotB.1, and Frd.2 using AlphaFold2 [63] and AlphaFold3 [64,65], and 2D topology diagrams were generated by PDBsum Generate [66]. To generate the dendrogram, a Dali server exhaustive search [67] in PDB25 was performed to find proteins similar to the AlphaFold2 models of GoF^{NTD} residues 1–91, $MotB.1^{NTD}$ residues 1–82, and $Frd.2^{NTD}$ residues 1–88. A collection of 55 unique similar proteins with Dali Z-scores greater than 3 together with the AlphaFold2 models of GoF^{NTD}, $MotB.1^{NTD}$, $Frd.2^{NTD}$ and the structure of Hfq residues 1–65 were then used to do an All against All structural comparison on the Dali server. The dendrogram was derived by average linkage clustering of the structural similarity matrix Dali Z-scores, and the domain structural classification of the 55 hits was further examined in the CATH [68] database. The pairwise comparison of GoF^{NTD} and $GoF(D25Y)^{NTD}$ was also performed on the Dali server.

The GoF dimer and GoF-RNA models were obtained by submitting protein and RNA sequences to the AlphaFold3 server. UCSF ChimeraX [69] was used to examine the models and generate the figures. The Coulombic electrostatic potential was calculated in ChimeraX with coulombic command.

4.4. Pull-down assay and mass spectrometry analyses

An overnight culture of TOP10 containing pBAD33-*goF-his*$_6$, pBAD33-*goF*(D25Y)-*his*$_6$, or pBAD33-*motB.1-his*$_6$ was grown in LB containing 25 μg/mL chloramphenicol and 0.025 % (*w/v*) glucose (Mallinckrodt Chemicals, Phillipsburg, UK) at 37° C. Cells were collected by centrifugation, washed with LB to remove glucose, diluted to an OD_{600}~0.1 in LB containing 25 μg/mL chloramphenicol, and then grown at 37° C with shaking (250 rpm) to an OD_{600}~0.3. Production of GoF-His$_6$, GoF(D25Y)-His$_6$, or MotB.1-His$_6$ was induced by the addition of 0.2 % (*w/v*) arabinose (Sigma-Aldrich, MO, USA) for 2 h, cells were harvested by centrifugation at 5,000 x g for 15 min, resuspended in ice-cold lysis buffer [20 mM HEPES (pH 8.5), 50 mM NaCl, 1 mM EDTA, 0.01 % (*v/v*) Triton X-100], and lysed by sonication until the OD_{600} was reduced ~3-fold. The clarified supernatant lysate was obtained after centrifugation at 15,000 x g for 30 min. An aliquot containing ~ 1 mg protein was mixed with 50 μL of Dynabeads His-Tag beads (Thermo Fisher Scientific) previously equilibrated with lysis buffer. After rotation for 30 min at 4° C, the beads were isolated by centrifugation, washed 4 times with cold lysis buffer (300 μL each), and the protein was eluted after rotation for 5 min at 4° C in lysis buffer containing 300 mM imidazole (50 μL).

After SDS-PAGE in the absence of β-mercaptoethanol, protein bands of interest obtained with the GoF-His$_6$ protein were excised, dehydrated in acetonitrile for 10 min, dried under a vacuum to remove the acetonitrile, reswelled at 4° C for 45 min in a buffer containing 50 mM $(NH_4)_2CO_3$ and trypsin (Promega, Madison, WI, USA; 200 ng/gel slice), and then incubated overnight at 37° C for tryptic digestion. After centrifugation, the supernatant was collected, and the tryptic peptides were extracted three times with 50 % acetonitrile containing 20 mM $(NH_4)_2CO_3$ and 5 % formic acid at room temperature. The peptide samples were combined and concentrated in a SpeedVac. Peptides were analyzed by RP-nano LC-ESI-MS/MS by the Seoul National University National Instrumentation Center (NICEM, Seoul, South Korea) using a Thermo Fisher Scientific Quadrupole-Orbitrap instrument (Thermo Fisher Scientific) equipped with a Dionex U 3000 RSLCnano HPLC system and an Exploris 240 mass spectrometer. Fractions were reconstituted in solvent A [water/acetonitrile (98:2 *v/v*), 0.1 % Formic acid] and then injected into the LC-nano ESI-MS/MS system. Samples were first trapped on an Acclaim PepMap 100 trap column (100 μm x 2 cm, nanoViper C18, 5 μm, 100 Å, Thermo Fisher Scientific, part number 164564) that was washed for 6 min with solvent A at a flow rate of 4 μL/min and then separated on a PepMap RSLC C18 column (75 μm x 15 cm, nanoViper C18, 3 μm, 100 Å, Thermo Fisher Scientific, part number ES900) at a flow rate of 300 nL/min. The LC gradient was run from 2 % to 8 % solvent B (Solvent B: 99.9 % acetonitrile and 0.1 % Formic acid) for 10 min, then from 8 % to 30 % for 55 min, followed by 90 % solvent B for 4 min, and finally 2 % solvent B for 20 min. (Dilutions of solvent B were made in water.) Xcaliber™ software version 4.4 (Thermo Fisher Scientific) was used to collect MS data. The Orbitrap analyzer scanned precursor ions with a mass range of 350–1800 *m/z* with 60,000 resolutions at *m/z* 200. Mass data were acquired automatically using proteome discoverer 2.5 (Thermo Fisher Scientific). The results were correlated against the known databases [Taxonomy ID: 2681598 for bacteriophage T4; for *E. coli* TOP10, the sequence of every protein in *E. coli* TOP10 was downloaded from the NCBI protein database].

4.5. Growth curves and protein gels

Cultures, inoculated with a single colony, were grown overnight at 37° C with shaking in MOPS EZ Rich Defined Media (MERDM broth, Teknova, Hollister, CA, USA) containing 0.025 % (*w/v*) glucose and 25 μg/mL chloramphenicol (or 25 μg/mL chloramphenicol plus 40 μg/mL kanamycin for Δ*hfq* strains).

The next day, cells from overnight cultures were collected by centrifugation, washed twice with MOPS EZ Rich Defined Media, used to inoculate MOPS EZ Rich Defined Media containing 25 μg/mL chloramphenicol, and grown at 37° C with shaking to OD_{600}~ 0.3. Cultures were then split in half with one portion receiving L-arabinose at a final concentration of 0.2 % (*w/v*) and the other receiving an equal volume of water. Aliquots taken before and after the addition of water/arabinose were used to monitor growth, as measured by OD_{600}. For SDS-PAGE analyses, 1 mL of each culture was collected 2 h post-induction and centrifuged at 16,000 x g for 1 min. Pellets were resuspended in 1 x Laemmli sample buffer (Bio-Rad) for a final concentration of 0.008 OD_{600} units/μl, heated at 95° C for 5 min, and vortexed vigorously. Protein samples (10 μl) were separated by SDS-PAGE on Any kD™ Mini-Protean® TGX™ gels, visualized by staining with AcquaStain protein gel stain (Bulldog Bio, Portsmith, NH), and destained with distilled water.

4.6. Phage efficiency of plating and burst size assays

To obtain phage titers for T4 *goF*^{am}, T4 *motB.1*^{am}, and T4 *frd.2*^{am}, NapIV and NapIV *supD* cultures were grown to mid-log phase (OD_{600}~0.4–0.5) and then placed on ice. Cell aliquots (0.2 mL) and phage (0.1 mL) were mixed with 2.5 mL melted, warm (45° C) top agar (10 g bactoagar per liter of LB media) and then poured onto solid, bottom LB plates (15 g bactoagar per liter of LB media, 25 mL per plate). Plates were incubated at 37° C overnight, and plaques were counted.

To determine burst sizes, *E. coli* TOP10 was grown to mid-log phase, harvested by centrifugation, and resuspended in 20 mL of fresh LB broth (OD_{600} ~0.5). A 0.1 mL aliquot of WT T4, Δ*goF*, or Δ*motB.1* was added at an MOI of 0.01 and allowed to adsorb for 5 min at room temperature. During subsequent incubation of the resuspension at 37° C with shaking (250 rpm), samples were taken at 5 or 10 min intervals for 40 min. Cell aliquots (0.2 mL) and phage (0.1 mL) were mixed with 2.5 mL melted, warm (45° C) top agar and then poured onto solid, bottom LB plates. Plates were incubated at 37° C overnight. Burst size was calculated by dividing the average number of plaques from the post-burst time points (35–40 min) by the average number of plaques from the pre-burst time points (0–20 min). The latent period represents the time between phage adsorption and host cell lysis, measured by the point at which phage progeny are released.

4.7. Fluorescent reporter assay

To quantify fluorescence signals, overnight cultures of *E. coli* cells were grown in MOPS EZ Rich Defined Media containing 25 μg/mL chloramphenicol and 0.025 % glucose at 37° C with shaking at 250 rpm. Cells were collected by centrifugation, washed with MOPS EZ Rich Defined Media to remove glucose, and diluted to an OD_{600}~0.1 in 195 μL MOPS EZ Rich Defined Media with 25 μg/mL chloramphenicol. After addition to 96-well microplates, cells were grown at 37° C for 1 h with shaking at 250 rpm and then induced by the addition of 0.2 % (*w/v*) L-arabinose (final concentration at time = 0). The OD_{600} and mCherry/GFP fluorescence were measured every 20 min using a TECAN microplate reader (Tecan, Männedorf, Switzerland; mCherry: 560 ± 20 nm excitation wavelength, 610 ± 20 nm emission wavelength; GFP: 475 ± 20 nm excitation wavelength, 520 ± 20 nm emission wavelength; optimal gain).

The expression of *goF*(D25Y) led to a modest inhibition of growth in the strains, which was not observed with the vector or with the expression of WT *goF* or *motB.1*. Consequently, fluorescent values were normalized to correct for any growth difference by dividing the fluorescent values by the OD_{600} at the time of the reading. The ratio of the normalized mCherry or GFP fluorescence observed with the induced culture *vs.* that observed with the uninduced culture was used as the read-out. At least three biological replicates were performed. Statistical significance was determined using an unpaired two-tailed Student's *t*-test. [Not significant, ns, ($P > 0.05$); *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$].

4.8. Measurements of RNA level and RNA decay

Overnight cultures of *E. coli* cells were grown in MOPS EZ Rich Defined Media containing 25 μg/mL chloramphenicol and 0.025 % glucose at 37° C with shaking at 250 rpm. Cells were harvested by centrifugation, washed twice with MOPS EZ Rich Defined Media to remove glucose, and diluted to an OD_{600}~0.1 in 45 mL of fresh MOPS EZ Rich Defined Media containing 25 μg/mL chloramphenicol. The cells were grown at 37° C with shaking to OD_{600}~ 0.3. The culture was split into two 20 mL portions, with one portion receiving L-arabinose at a final concentration of 0.2 % (*w/v*) and the other receiving an equal volume of water. Both cultures were incubated at 37° C with shaking. At 2 h post-induction, 1 mL of culture was collected and mixed with a one-tenth volume of a stop solution (10 % saturated phenol in ethanol) and chilled rapidly. The culture was then pelleted by centrifugation at 16,000 x g for 1 min, placed on dry ice, and stored at −80° C until RNA extraction. This represented a 0 min sample for the RNA decay assay. Rifampicin was then added to the L-arabinose treated cultures at a final concentration of 500 μg/mL. After rifampicin treatment, the cells were harvested at various time points (2, 4, 6, and 8 min) and processed in the same manner as the 0 min samples. Total RNA was isolated using Method II as previously described [14].

cDNA was generated from 200 ng of RNA using random hexamers (d(N), New England Biolabs, 1 nM) as primers and SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific). RT-qPCR was performed using a CFX96 Touch real-time detection system (Bio-Rad). *rpoD* was used as the internal control, and SsoAdvanced SYBR green Supermix (Bio-Rad) was used as the fluorescent dye. Each reaction was performed in a 96-well microtiter PCR plate, with 1 μL of cDNA at 1.5 ng/μl and primers (0.5 μM each; mCherry-RT-F/R, GFP-RT-F/R, rpoD-RT-F/R; Table S6) targeting the *mCherry*, *gfp*, and *rpoD* genes. The amplification conditions were as follows: denaturation (95° C for 30 s), amplification and quantification (40 cycles of 95° C for 15 s and 55° C for 15 s), a melting curve program (65°–95° C with a heating rate of 0.5° C per second and continuous fluorescence measurement), and a cooling step to 65° C. Data were analyzed using the CFX manager software (Bio-Rad).

For RNA level analysis at 2 h post-induction, the quantitation cycle for the gene of interest (Cq_{gene}) and for the reference gene (Cq_{rpoD}) was determined, and the ΔCq was calculated as $Cq_{gene} - Cq_{rpoD}$. The ΔΔCq was calculated as ΔCq for the induced condition minus ΔCq for the uninduced condition, and the fold change (FC) was determined using $2^{-\Delta\Delta Cq}$.

For RNA decay analysis, ΔCq was calculated for each time point as $Cq_{gene} - Cq_{rpoD}$. The ΔΔCq for each time point was determined by subtracting the ΔCq at 0 min for the same strain, and the RNA remaining level (%) was calculated as $2^{-\Delta\Delta Cq}$ x 100.

Funding

This work was funded by the Intramural Research Program of the National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases (J.K., O.S., N.L., D.H.), by the National Cancer Institute (C.-H.T.), and by a grant from the Ministry of Science and ICT (MSIT) through the National Research Foundation (NRF) of Korea (NRF-RS-2023–00212000; B.S.).

Author Statement

The authors declare no relevant conflicts of interest in relation to this work. All authors have read and approved the final manuscript and agree to its submission to this journal.

CRediT authorship contribution statement

Hinton Deborah: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Bokyung Son:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Livingston Natalie:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis. **Chin-Hsien Tai:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation. **Oliver Stearns:** Writing – review & editing, Methodology, Investigation, Formal analysis. **Jennifer Patterson-West:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jinshil Kim:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare no competing interest.

Acknowledgments

We thank Nadim Majdalani (NCI), Susan Gottesman (NCI), and Mike Cashel (NICHD, NIH) for strains and advice, Hernan Lorenzi (NIDDK, NIH) and Andrew Millard (University of Leicester) for bioinformatics help, Jingen Zhu and Venigalla Rao (The Catholic University of America) for help making T4 Δ*goF* and T4 Δ*motB.1*, Ethan Pham and Virginia Rosas (NIDDK, NIH) for help with transformations and growing phage, and Leslie Knipling (NIDDK, NIH) for useful discussions. We are especially grateful to Aravind Iyer (NCBI, NIH), who made the initial observation of a possible similarity among the NTDs of GoF, MotB.1, and Frd.2. Molecular graphics and analyses were performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311. This work utilized the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2025.05.028.

References

[1] Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Ruger W. Bacteriophage T4 genome. Microbiol Mol Biol Rev 2003;67(1):86–156.
[2] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215(3):403–10.
[3] Hatfull GF. Dark matter of the biosphere: the amazing world of bacteriophage diversity. J Virol 2015;89(16):8107–10.
[4] Geiduschek EP, Kassavetis GA. Transcription of the T4 late genes. Virol J 2010;7:288.
[5] Hinton DM. Transcriptional control in the prereplicative phase of T4 development. Virol J 2010;7(1):289.
[6] Stitt BL, Revel HR, Lielausis I, Wood WB. Role of the host cell in bacteriophage T4 development. II. Characterization of host mutants that have pleiotropic effects on T4 growth. J Virol 1980;35(3):775–89.
[7] Caruso M, Coppo A, Manzi A, Pulitzer JF. Host–virus interactions in the control of T4 prereplicative transcription. I. *tabC* (*rho*) mutants. J Mol Biol 1979;135(4):959–77.
[8] Takahashi H. Genetic and physiological characterization of *Escherichia coli* K12 mutants (*tabC*) which induce the abortive infection of bacteriophage T4. Virology 1978;87(2):256–65.
[9] Chiurazzi M, Pulitzer JF. Characterisation of the bacteriophage T4 *comC alpha 55.6* and *comCJ* mutants. A possible role in an antitermination process. FEMS Microbiol Lett 1998;166(2):187–95.
[10] Takahashi H, Yoshikawa H. Genetic study of a new early gene, *comC-alpha*, of bacteriophage T4. Virology 1979;95(1):215–7.
[11] Sanson B, Uzan M. Sequence and characterization of the bacteriophage T4 *comC alpha* gene product, a possible transcription antitermination factor. J Bacteriol 1992;174(20):6539–47.

[12] Stitt BL, Mosig G. Impaired expression of certain prereplicative bacteriophage T4 genes explains impaired T4 DNA synthesis in Escherichia coli *rho* (*nusD*) mutants. J Bacteriol 1989;171(7):3872–80.
[13] Hinton DM. Altered expression of the bacteriophage T4 gene *41* (primase-helicase) in an *Escherichia coli rho* mutant. J Biol Chem 1989;264(24):14440–6.
[14] Hinton DM. Transcript analyses of the *uvsX-40-41* region of bacteriophage T4. Changes in the RNA as infection proceeds. J Biol Chem 1989;264(24):14432–9.
[15] Sozhamannan S, Stitt BL. Effects on mRNA degradation by *Escherichia coli* transcription termination factor Rho and pBR322 copy number control protein Rop. J Mol Biol 1997;268(4):689–703.
[16] Lekontseva NV, Stolboushkina EA, Nikulin AD. Diversity of LSM family proteins: Similarities and differences. Biochemistry 2021;86(1):S38–49.
[17] Weichenrieder O. RNA binding by Hfq and ring-forming (L)Sm proteins: a trade-off between optimal sequence readout and RNA backbone conformation. RNA Biol 2014;11(5):537–49.
[18] Mura C, Randolph PS, Patterson J, Cozen AE. Archaeal and eukaryotic homologs of Hfq: a structural and evolutionary perspective on Sm function. RNA Biol 2013;10 (4):636–51.
[19] Santiago-Frangos A, Woodson SA. Hfq chaperone brings speed dating to bacterial sRNA. Wiley Interdiscip Rev RNA 2018;9(4):e1475.
[20] Updegrove TB, Zhang A, Storz G. Hfq: the flexible RNA matchmaker. Curr Opin Microbiol 2016;30:133–8.
[21] Dos Santos RF, Andrade JM, Pissarra J, Deutscher MP, Arraiano CM. Hfq and RNase R mediate rRNA processing and degradation in a novel RNA quality control process. mBio 2020;11(5).
[22] Cowan J, D'Acci K, Guttman B, Kutter E. Gel analysis of T4 prereplicative proteins. In: Karam J, editor. Molecular Biology of Bacteriophage T4. Washington, DC: American Society of Microbiology; 1994. 520–T7.
[23] Son B, Patterson-West J, Arroyo-Mendoza M, Ramachandran R, Iben JR, Zhu J, et al. A phage-encoded nucleoid associated protein compacts both host and phage DNA and derepresses H-NS silencing. Nucleic Acids Res 2021;49(16):9229–45.
[24] Homyk Jr T, Weil J. Deletion analysis of two nonessential regions of the T4 genome. Virology 1974;61(2):505–23.
[25] Teyra J, Huang H, Jain S, Guan X, Dong A, Liu Y, et al. Comprehensive analysis of the human SH3 domain family reveals a wide variety of non-canonical specificities. Structure 2017;25(10):1598–610. e3.
[26] Lau CK, Ishida H, Liu Z, Vogel HJ. Solution structure of *Escherichia coli* FeoA and its potential role in bacterial ferrous iron transport. J Bacteriol 2013;195(1):46–55.
[27] Radman K, Crnolatac I, Bregovic N, Matosevic ZJ, Fernanades PA, Merunka D, et al. Conformational change induced by binding of Mn^{2+} ions activates SloR transcription factor in *Streptococcus mutans*. Int J Biol Macromol 2025;290:138828.
[28] Pohl E, Holmes RK, Hol WG. Crystal structure of the iron-dependent regulator (IdeR) from *Mycobacterium tuberculosis* shows both metal binding sites fully occupied. J Mol Biol 1999;285(3):1145–56.
[29] Rodriguez GM, Voskuil MI, Gold B, Schoolnik GK, Smith I. ideR, An essential gene in *Mycobacterium tuberculosis*: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. Infect Immun 2002;70(7):3371–81.
[30] Kozbial Das D, Axelrod P, Miller HL, McMullan MD, Krishna D, Abdubek SS, et al. IA. Crystal structure of a novel Sm-like protein of putative cyanophage origin at 2.60 A resolution. Proteins 2009;75(2):296–307.
[31] Chen J, To L, de Mets F, Luo X, Majdalani N, Tai CH, et al. A fluorescence-based genetic screen reveals diverse mechanisms silencing small RNA signaling in *E. coli*. Proc Natl Acad Sci USA 2021;118(27).
[32] Andrzejewska A, Zawadzka M, Pachulska-Wieczorek K. On the way to understanding the interplay between the RNA structure and functions in cells: a genome-wide perspective. Int J Mol Sci 2020;21(18).
[33] Park S, Prevost K, Heideman EM, Carrier MC, Azam MS, Reyer MA, et al. Dynamic interactions between the RNA chaperone Hfq, small regulatory RNAs, and mRNAs in live bacterial cells. eLife 2021;10.
[34] Washburn RS, Jin DJ, Stitt BL. The mechanism of early transcription termination by Rho026. J Mol Biol 1996;260(3):347–58.
[35] Washburn RS, Stitt BL. *In vitro* characterization of transcription termination factor Rho from *Escherichia coli rho*(*nusD*) mutants. J Mol Biol 1996;260(3):332–46.
[36] Dreyfus M. Killer and protective ribosomes. Prog Mol Biol Transl Sci 2009;85: 423–66.
[37] Jensen RA. Enzyme recruitment in evolution of new function. Annu Rev Microbiol 1976;30:409–25.
[38] Jayaraman V, Toledo-Patino S, Noda-Garcia L, Laurino P. Mechanisms of protein evolution. Protein Sci 2022;31(7):e4362.
[39] Li Y, Zhang Y, Li X, Yi S, Xu J. Gain-of-function mutations: an emerging advantage for cancer biology. Trends Biochem Sci 2019;44(8):659–74.
[40] Sotomayor-Vivas C, Hernandez-Lemus E, Dorantes-Gilardi R. Linking protein structural and functional change to mutation using amino acid networks. PLOS One 2022;17(1):e0261829.
[41] Elofsson A. Progress at protein structure prediction, as seen in CASP15. Curr Opin Struct Biol 2023;80:102594.
[42] Pak MA, Markhieva KA, Novikova MS, Petrov DS, Vorobyev IS, Maksimova ES, et al. Using AlphaFold to predict the impact of single mutations on protein stability and function. PLOS One 2023;18(3):e0282689.
[43] Garrido-Rodriguez P, Carmena-Bargueno M, de la Morena-Barrio ME, Bravo-Perez C, de la Morena-Barrio B, Cifuentes-Riquelme R, et al. Analysis of AlphaFold and molecular dynamics structure predictions of mutations in serpins. PLOS One 2024;19(7):e0304451.
[44] Kryshtafovych A, Montelione GT, Rigden DJ, Mesdaghi S, Karaca E, Moult J. Breaking the conformational ensemble barrier: ensemble structure modeling challenges in CASP15. Proteins 2023;91(12):1903–11.
[45] Nelson MA, Ericson M, Gold L, Pulitzer JF. The isolation and characterization of TabR bacteria: hosts that restrict bacteriophage T4 *rII* mutants. Mol Gen Genet 1982;188:60–8.
[46] Studier FW, Moffatt BA. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol 1986;189(1):113–30.
[47] Hanahan D. Techniques for transformation of *E. coli*. In: Glover DM, editor. DNA Cloning. 1. Oxford: IRL Press; 1985. p. 109–35.
[48] Woodcock DM, Crowther PJ, Doherty J, Jefferson S, DeCruz E, Noyer-Weidner M, et al. Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. Nucleic Acids Res 1989; 17(9):3469–78.
[49] Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL. An efficient recombination system for chromosome engineering in *Escherichia coli*. Proc Natl Acad Sci USA 2000;97(11):5978–83.
[50] Majdalani N, Chattopadhyay M, Keller C, Gottesman S. Lack of polyamines leads to cotranslational degradation of the general stress factor RpoS in *Escherichia coli*. J Biol Chem 2023;299(8):104943.
[51] Sharma M, Hinton DM. Purification and characterization of the SegA protein of bacteriophage T4, an endonuclease related to proteins encoded by group I introns. J Bacteriol 1994;176(21):6439–48.
[52] Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzym 1990;185:60–89.
[53] Tao P, Wu X, Tang WC, Zhu J, Rao V. Engineering of bacteriophage T4 genome using CRISPR-Cas9. ACS Synth Biol 2017;6(10):1952–61.
[54] Liu Y, Dai L, Dong J, Chen C, Zhu J, Rao VB, et al. Covalent modifications of the bacteriophage genome confer a degree of resistance to bacterial CRISPR systems. J Virol 2020;94(23).
[55] Guzman LM, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol 1995;177(14):4121–30.
[56] Elliott T, Geiduschek EP. Defining a bacteriophage T4 late promoter: absence of a "-35" region. Cell 1984;36(1):211–9.
[57] Datta S, Costantino N, Court DL. A set of recombineering plasmids for gram-negative bacteria. Gene 2006;379:109–15.
[58] Bonocora RP, Caignan G, Woodrell C, Werner MH, Hinton DM. A basic/ hydrophobic cleft of the T4 activator MotA interacts with the C-terminus of *E.coli* sigma70 to activate middle gene transcription. Mol Microbiol 2008;69(2):331–43.
[59] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25(17):3389–402.
[60] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004;32(5):1792–7.
[61] Tamura K, Stecher G, Kumar S. MEGA11: Molecular evolutionary genetics analysis version 11. Mol Biol Evol 2021;38(7):3022–7.
[62] Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 2021;49(W1):W293–6.
[63] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596(7873): 583–9.
[64] Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Addendum: accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 2024;636(8042):E4.
[65] Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 2024; 630(8016):493–500.
[66] Laskowski RA, Jablonska J, Pravda L, Varekova RS, Thornton JM. PDBsum: Structural summaries of PDB entries. Protein Sci 2018;27(1):129–34.
[67] Holm L, Laiho A, Toronen P, Salgado M. DALI shines a light on remote homologs: one hundred discoveries. Protein Sci 2023;32(1):e4519.
[68] Sillitoe I, Bordin N, Dawson N, Waman VP, Ashford P, Scholes HM, et al. CA. CATH: increased structural coverage of functional space. Nucleic Acids Res 2021; 49(D1):D266–73.
[69] Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, et al. UCSF ChimeraX: tools for structure building and analysis. Protein Sci 2023;32(11): e4792.
[70] Basquin Sauter C, Suck J. D. Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from *Escherichia coli*. Nucleic Acids Res 2003;31(14):4091–8.