Highly Divergent RfaH Orthologs from Pathogenic Proteobacteria Can Substitute for *Escherichia coli* RfaH both In Vivo and In Vitro

Heather D. Carter, Vladimir Svetlov, and Irina Artsimovitch*

Department of Microbiology, The Ohio State University, Columbus, Ohio 43210

Received 18 November 2003/Accepted 14 January 2004

The transcriptional enhancer protein RfaH positively regulates production of virulence factors in *Escherichia coli* and *Salmonella enterica* serovar Typhimurium via a cis element, *ops*. Genes coding for RfaH orthologs were identified in conceptually translated genomes of bacterial pathogens, including *Vibrio* and *Yersinia* spp. We cloned the *rfaH* genes from *Vibrio cholerae*, *Yersinia enterocolitica*, *S. enterica* serovar Typhimurium, and *Klebsiella pneumoniae* into *E. coli* expression vectors. Purified RfaH orthologs, including the most divergent one from *V. cholerae*, were readily recruited to the *E. coli* transcription elongation complex. Postrecruitment stimulation of transcript elongation appeared to vary with the degree of similarity to *E. coli* RfaH. *V. cholerae* RfaH was particularly defective in reducing downstream pausing and termination; this defect was substantially alleviated by an increase in its concentration. When overexpressed episomally, all of the *rfaH* genes complemented the disruption of the chromosomal copy of the *E. coli* gene. Thus, despite the apparently accelerated divergent evolution of the RfaH proteins, the mechanism of their action is conserved well enough to make them transcriptionally active in the *E. coli* system.

Bacterial RNA polymerase (RNAP) carries out transcription of the entire genome and is one of the most regulated cellular enzymes. During synthesis of a single RNA molecule, RNAP transitions through a series of mechanistically and conformationally distinct stages: recruitment, initiation, elongation, and termination (37). At each of these stages RNAP activity is regulated by a host of cis- and trans-acting signals, although historically it was the regulation of entry into the catalytic cycle (recruitment and initiation) that received the most attention. In the cell, RNAP cannot splice individual RNA molecules nor can it initiate synthesis of each RNA chain de novo in the absence of a promoter, and thus RNAP must remain associated with its product for thousands of nucleotide addition cycles. However, RNAP movement along the template is not monotonous as RNAP can pause or terminate transcription at distinct sites (37); some of these sites play regulatory roles, whereas others exist as an emergent property of the energetics of transcription (13). In combination these sites determine not only the rate of transcription elongation but also, at least in certain cases, the folding pathways of the RNA molecules (19, 28).

Elongating RNAP may release the nascent RNA chain either stochastically (with low frequency) or at specific sites, terminators (with comparatively high frequency ranging from 5 to >99%). Numerous bacterial protein factors, acting alone or as a part of large macromolecular complexes, regulate transcription termination, either negatively or positively (26, 38). While some of these factors act upon a single terminator, others are able to exert a processive antitermination effect that persists over thousands of nucleotides and allows bypass of several tandem terminators (38); these mechanisms are particularly important during transcription of long operons since as the length of the transcript increases so does the probability of termination. Regulation of expression of the phage λ genome (30, 32) and the nontranslated rRNA operons (9) by processive antitermination has been studied in great detail.

A protein factor may inhibit RNA chain release at a terminator through stabilization of the transcription elongation complex (TEC) against dissociation (antiterminator) or by reducing pausing and thus increasing the kinetic barrier to termination (elongation enhancer). In either case, a protein factor can upregulate the expression of the distal genes in the operon without affecting the rate(s) of recruitment/initiation. Two λ phage antiterminators, N and Q proteins, are thought to utilize both antitermination mechanisms (31, 42), whereas bacterial transcription elongation factor RfaH may act predominantly as an elongation enhancer (3).

RfaH controls expression of extracytoplasmic components in *Escherichia coli* and related bacteria, including lipopolysaccharides (LPS), hemolysin, capsules, etc. (reviewed in reference 6). Although expression of the genes encoding these extracytoplasmic components is not an essential cell function, it has a profound effect on bacterial pathogenicity—disruptions of the *rfaH* gene give rise to viable but avirulent mutants (23). RfaH regulatory function depends on a cis element called *ops* for “operon polarity suppressor” (6). RfaH increases expression of the downstream portions of *ops*-containing operons in vivo without affecting transcription initiation (6). In vitro, RfaH significantly enhances the overall elongation rate and suppresses pausing but only modestly inhibits termination, suggesting that its effects could be the strongest at the specific sites where termination is kinetically limited.

RfaH was first identified genetically as a regulator of the expression of LPS transferases in *Salmonella enterica* serovar Typhimurium and *tra* functions in *E. coli* (7, 40, 41). Conceptual translation of microbial genomes and National Center for Biotechnology Information BLAST-powered alignments (1)
allowed us to tentatively identify RfaH orthologs in many proteobacteria (Fig. 1A). Even among proteobacteria, the sequence divergence is quite significant, and only the E. coli protein was studied in vitro. No conclusive evidence of transcriptional activity of any other RfaH ortholog has been reported, and for most no direct indication of any regulatory role exists except for their sequence similarity to E. coli RfaH. In addition, RfaH is also homologous to NusG, a general elongation factor, and the percent identity approaches that of distant RfaH orthologs. Although, like RfaH, NusG also increases the RNA chain elongation rates in vitro (8) and in vivo (43), its other effects on transcription are opposite to those of RfaH: NusG increases Rho-dependent termination (36), reduces pausing at the ops site but does not affect pausing at the hairpin-dependent sites (2), participates in formation of macromolecular antitermination complexes (22, 34), and lacks the

FIG. 1. RfaH-NusG family. (A) Phylogenetic tree of RfaH and NusG homologs was built with the MegAlign module of the DNASTAR software (DNASTAR, Madison, Wis.), using the CLUSTAL W protocol (11). (B) Alignment of RfaH proteins from E. coli (RfaH_ECOLI), V. cholerae (RfaH_VCHOL), Y. enterocolitica (RfaH_YENTE), S. enterica serovar Typhimurium (RfaH_STYHI), and K. pneumoniae (RfaH_KPNEU) with E. coli NusG (NusG_ECOLI) and the most distant RfaH ortholog, the TaA protein from M. xanthus (TaA_MXANT). The height and darkness of the bars indicate the degree of conservation among protein residues, with the black bars corresponding to the invariant amino acids. Identity to the E. coli RfaH (percent) is shown on the right.
apparent sequence specificity of the ops-dependent RfaH. Thus, especially for distant RfaH homologs, a question whether these proteins act similarly to RfaH or NusG, or if they even function as transcription regulators, must be addressed.

The amino acid sequences of the RfaH orthologs appear to diverge to a greater degree than that of their paralogs, NusG (Fig. 1). It is not yet clear why the rate of protein evolution among the RfaH members of the NusG-RfaH superfamily is far greater than among their paralogs (one possible explanation is that nusG is in many instances an essential gene, whereas rfaH is not), but taken together with analysis of codon preferences (data not shown) these observations argue against horizontal transfer of rfaH genes. Instead, as more genomic information becomes available, the RfaH family of proteins emerges as rapidly evolving, proteobacterium-specific paralogs of NusG, with the rfaH gene arising most likely through an early duplication of nusG. The polypeptide sequence of the E. coli RfaH does not reveal any discernible structural motifs except the N-terminal coiled coil absent in most RfaH orthologs, whereas the statistically significant KOW (15) and NGN (29) motifs have not been characterized biochemically.

In contrast, the ops sites found in the same operons in different bacteria are remarkably conserved in sequence (Fig. 2). We have previously shown that E. coli RfaH can be recruited in vitro to the TEC through direct contact with the ops sequence exposed on the nontemplate DNA strand on the surface of RNAP (3). Conservation of the RfaH recruitment site and the multistep action of this transcription factor prompted us to investigate the in vitro activity of previously uncharacterized heterologous RfaH proteins in the E. coli-derived system. More specifically, we set out to determine which, if any, of the divergent RfaH orthologs can still be recruited to the ops site and what their effects are on elongation rate, pausing, and termination by the E. coli RNAP. To this effect, we cloned rfaH genes from the genomes of S. enterica serovar Typhimurium, Yersinia enterocolitica, Vibrio cholerae, and Klebsiella pneumoniae, expressed and purified the resulting proteins, and subjected them to a panel of single-round transcription assays, developed for E. coli RfaH. We also tested the ability of these proteins to substitute for the rfaH gene in vivo. Here we report that, despite substantial sequence divergence, all of the assayed RfaH orthologs demonstrate characteristic activity in vitro and are capable of complementing an rfaH mutation in vivo.

FIG. 2. ops site conservation. Homologous operons encoding LPS components in E. coli, V. cholerae (VCHOL), Y. enterocolitica (YENTE), and K. pneumoniae (KPNEU). Individual ORFs are represented by gray arrows. Locations of ops sites are indicated with vertical black bars, and the corresponding sequences are shown on the right.
linear DNA template and 50 nM RNA polymerase in 20 to 100 μl of transcription buffer (20 mM Tris - HCl, 20 mM NaCl, 10 mM MgCl2, 14 mM 2-mercaptoethanol, 0.1 mM EDTA [pH 7.9]). To halt the elongation complexes after addition of G37, transcription was initiated in the absence of UTP, with ApU at 150 μM, ATP and GTP at 2.5 μM, CTP at 1 μM, with 32P derived from [α-32P]CTP (3,000 Ci/mmol). Halted complexes were formed for 15 min at 37°C and stored on ice prior to use as described previously (16).

Single-round pause assays. Halted complexes were formed in 50 μl of transcription buffer. Transcription was restarted by addition of nucleotides (at concentrations indicated in the figure legends) and rifampicin to 25 μg/ml. When present, RfaH proteins were added prior to the nucleotides. Samples were removed at the times shown in the figures and after a final 5-min incubation with 250 μM concentration of each NTP (Chase) and quenched by addition of an equal volume of STOP buffer (10 M urea, 20 mM EDTA, 45 mM Tris-borate [pH 8.3]). Pause half-life (the time during which half of the complexes reenter the elongation pathway) was determined by nonlinear regression analysis as described previously (16).

Termination assays. Halted [32P]CTP-labeled elongation complexes were prepared in 20 μl of transcription buffer with 40 nM linear DNA templates and 50 nM RNA polymerase. Elongation was restarted by addition of NTPs (10 μM UTP and 200 μM ATP, CTP, and GTP) and rifampicin at 25 μg/ml. Reaction mixtures were incubated at 37°C for 15 min, and reactions were stopped by addition of an equal volume of STOP buffer. RNA products were analyzed on 5% denaturing gels, and the termination efficiencies were determined as described previously (4).

Sample analysis. Samples were heated for 2 min at 90°C and separated by electrophoresis in denaturing acrylamide (19:1) gels (7 M urea, 0.5 × Tris-borate-EDTA) of various concentrations (5 to 20%). RNA products were visualized and quantified using a Molecular Dynamics (Piscataway, N.J.) phosphorimaging system, ImageQuant software, and Microsoft Excel (16).

Hemolytic assays. Hemolytic assays were performed as described previously (18, 21) with the following modifications. E. coli strains WAM 1925 (lacZA145 rfaE-Tn 5 l.8) and WAM1931 (lacZA145) were obtained from R. Welch (17). Plasmids with the rfaH gene under control of the Pm promoter (listed in Table 1) were cotransformed with the reporter plasmid pSF4000 into WAM1925 or WAM1931. Overnight cultures were inoculated at 1:100 into fresh Legionaria-Bertani medium supplemented with chloramphenicol at 30 mg/ml and ampicillin at 100 mg/ml, grown for 45 min at 37°C, and then induced by addition of IPTG to 1 mM. Cultures were grown to an optical density at 600 nm of 0.8 to 1.2. Cells were pelleted by centrifugation, and the supernatant was incubated with 2% sheep erythrocytes (defibrinated sheep blood; Fisher, Pittsburgh, Pa.) in hemolysis buffer (20 mM CaCl2, 150 mM NaCl) at 37°C for 30 min. The hemolytic activity measured was in the linear range for different ratios of supernatant to erythrocytes in hemolysis buffer. Erythrocytes were then removed by centrifugation, and absorbance readings at 543 nm were taken from the supernatant by spectrophotometry. The relative amount of hemolysis was calculated as the optical density at 600 nm. For plate assays, cultures were streaked onto blood agar plates (5% defibrinated sheep blood in Luria-Bertani agar) with 30 mg of chloramphenicol/liter, 100 mg of ampicillin/liter, and 1 mM IPTG and incubated at 37°C.

### RESULTS

**Experimental setup.** RfaH-like open reading frames (ORFs) are found in genomes of proteobacteria (Fig. 1A). We reasoned that it should be possible to test whether RfaH homologs from these bacterial species could function in a heterologous E. coli-based in vitro transcription system, since currently this is the only system that allows rigorous dissection of the transcription elongation process. In particular, given the high conservation of the ops sequence (Fig. 2), we postulated that RfaH orthologs should all be able to bind to the E. coli TEC paused at the ops site and, depending on their degree of divergence from the E. coli RfaH, could also regulate transcription elongation after escape from the ops site. To study the effect of RfaH proteins in vitro, we used DNA templates that feature a strong T7A1 promoter for E. coli RNAP and an rfaQ ops site centered around position 37, followed by different transcribed regions. rfaQ belongs to the LPS biosynthesis operon and is implicated genetically in LPS modification, but its exact function is unknown; the mechanism of pausing at the rfaQ ops site and the effect of RfaH on transcription through this site have been characterized in detail (2, 3).

We selected RfaH proteins from four species of proteobacteria, V. cholerae, Y. enterocolitica, S. enterica serovar Typhi murium, and K. pneumoniae, for comparative analysis with E. coli RfaH. These proteins span a rather wide range of sequence identity relative to the E. coli ortholog, from 88% (S. enterica serovar Typhimurium) to 43% (V. cholerae); for comparison, the β subunit of RNAP, a likely target for RfaH, displays 99 and 87% identity to the respective species pairs.

Selected RfaH proteins were expressed in E. coli and purified to apparent homogeneity using the same procedures as those for E. coli RfaH. In order to eliminate potential concerns about the variable refolding efficiency, we purified only the soluble fraction of expressed RfaH homologs, which accounts for approximately 5% of the total IPTG-dependent expression product (the balance being the insoluble material of the inclusion bodies [data not shown]). The same ORFs were subcloned into a lower-expression pTrc99 plasmid to test their ability to complement the rfaH-null mutation in vivo.
Recruitment of RfaH orthologs to the ops site in vitro.

We have previously shown that E. coli RfaH is recruited to the E. coli RNAP paused at the ops site through sequence-specific binding to the nontemplate DNA strand (3). The immediate consequence of this binding is a significant delay of RNAP escape from the ops pause site: C45 transcripts are readily elongated by RNAP in the absence of RfaH (Fig. 3, leftmost lane) but persist significantly longer when RfaH is present. Addition of E. coli RfaH increases the half-life of the C45 TEC from 18 s to 162 s on a pIA416-derived template (Fig. 3, second lane). This effect likely results from the blocking of RNAP translocation imposed by the persistent RfaH-ops contact. Addition of RfaH orthologs from S. enterica serovar Typhimurium, Y. enterocolitica, V. cholerae, and K. pneumoniae to the E. coli TECs also led to a shift of the ops-dependent pausing from position U43 to C45 and increased the dwell time at position C45 (C45 t1/2), delaying TEC escape from this site to essentially the same degree as the E. coli RfaH (Fig. 3). All RfaH proteins were present at a concentration equal to that of the TEC. The observation that even the more than 50% divergent factors from E. coli and V. cholerae perform similarly in this assay supports our hypothesis that the recruitment of RfaH to TEC is mediated primarily by the interactions between RfaH and the well-conserved ops site.

Antitermination effects of RfaH orthologs in vitro. We have previously proposed a two-step mode of E. coli RfaH action (3): (i) at the recruitment step RfaH binds to the ops-paused TEC, delaying its escape from the pause site; (ii) once recruited to the TEC, RfaH modifies RNAP by a yet unknown mechanism to effectively increase the rate of elongation and

FIG. 3. Recognition of the ops site by RfaH proteins. (Top) Transcript generated from the T7A1 promoter on a linear pIA416 DNA. The template transcription start site (+1), ops (boxed), T_hly terminator, and transcript end (run-off) are indicated. (Bottom) Halted G37 TECs were formed with E. coli RNAP and challenged with NTPs (20 μM GTP and 150 μM ATP, CTP, and UTP) and rifampin at 25 μg/ml in the absence (leftmost panel) or presence of 50 nM RfaH proteins indicated below (abbreviations are as for Fig. 1). Aliquots were withdrawn at the times indicated above each lane, followed by the high-concentration NTP chase as above. Positions of the halted (G37), paused (U43, C45, and P1), terminated (T_hly), and runoff transcripts are indicated with arrows; transcripts above the runoff RNA were generated as a result of a template switch (10). Sizes of the 32P-labeled DNA markers used as molecular weight standards (MWM; pBR322 MspI digest) are indicated on the right.
make the TEC less prone to pausing and termination. For example, *E. coli* RfaH decreases pausing at certain sites (such as P1, Fig. 3) and allows RNAP to reach the end of the template faster: the runoff RNA can be clearly seen after a 20-s chase compared to 30 s in the absence of RfaH (compare two lanes on the left in Fig. 3). RfaH orthologs from *S. enterica* serovar Typhimurium, *Y. enterocolitica*, and *K. pneumoniae* behaved similarly, whereas *V. cholerae* RfaH did not appear to accelerate RNAP to the same extent (as judged by the appearance of the runoff RNA; Fig. 3).

In order to evaluate antitermination potential of the RfaH orthologs, we tested their effects on transcription through the intergenic T$_{hb}$ terminator from the J96 hly operon encoded on the pLA416 template. This signal was reported to be regulated by RfaH in vivo (14) and responds to RfaH in vitro (V. Svetlov and I. Artsimovitch, unpublished data). The termination efficiency at this terminator (fraction of RNA molecules that terminated at position 219) is 40%; those RNAP molecules that read through the terminator (60%) reached the half-maximum value almost twice as fast as the reaction mixture containing no RfaH. The RfaH proteins from *S. enterica* serovar Typhimurium and *K. pneumoniae* also increased the elongation rate indistinguishably from the *E. coli* RfaH (data not shown). In contrast, *V. cholerae* RfaH was essentially inactive in this assay. This apparent lack of stimulatory effect on elongation by the *V. cholerae*-derived ortholog is in good agreement with its decreased antitermination activity (Fig. 4).

Stimulatory effects of the *E. coli* RfaH on transcription require the presence of an *ops* site, both in vivo (6) and in vitro (3); the same is true for the RfaH orthologs studied here (data not shown).

**Regulation of hemolysin production in *E. coli* by proteobacterial rfaH genes.** The results of our in vitro studies indicate that, despite significant polypeptide sequence divergence, RfaH orthologs from different proteobacteria appear fairly conserved in their ability to be recruited to an *ops*-paused TEC and, to a lesser degree, stimulate the elongation rate and reduce pausing and termination. In order to extend these findings to regulatory events in vivo, we created a test system that allowed us to determine rfaH effects on expression and secretion of hemolysin (Fig. 6A). A host *E. coli* strain with a disrupted chromosomal copy of the rfaH gene was transformed with plasmid pSF4000, bearing the entire hemolysin (*hly*) operon under the control of *ops* (18, 39). Using plate and liquid hemolytic assays, we ascertained that this rfaH mutant strain...
was not hemolytic, while the wild-type rfaH+ strain was able to produce and secrete hemolysin (Fig. 6B and C).

In order to test the ability of orthologous rfaH genes to restore hemolytic activity of the rfaH mutant host, we cloned their respective ORFs from the pET28a-based expression vectors into the ptrc99 plasmid under control of the IPTG-inducible P_{trc} promoter (Fig. 6; Table 1). The resulting plasmids were then cotransformed with pSF4000 into the rfaH mutant strain, and RfaH expression was induced by IPTG. The hemolytic activity of RfaH was determined both on blood agar plates and in liquid medium. The blood agar plates show complementation of an rfaH deletion in E. coli with the rfaH from the organisms studied, as evidenced by clearing/hemolysis zones around the streaked-out colonies (Fig. 6B). The levels of he-
molysis seen on the agar plates are comparable to the level of hemolysis seen with *E. coli* containing the intact *rfaH* gene.

In order to obtain a more quantitative measure of the hemolytic activities of these strains and thus better assess the ability of different *rfaH* genes to complement the chromosomal disruption of the *E. coli* gene, we adopted a liquid hemolytic assay (18, 21). Expression of all RfaH orthologs in the *rfaH* mutant background led to hemolysis at levels comparable to that observed with the expressed *E. coli* RfaH. RfaH from *V. cholerae* induced 1.5-fold higher hemolysis than the *E. coli* RfaH, RfaHs from *S. enterica* serovar Typhimurium and *K. pneumoniae* induced hemolysis at similar levels, and RfaH from *V. cholerae* induced 1.2-fold less hemolysis than the *E. coli* RfaH. The levels of hemolysis from the overexpressed RfaH proteins are higher than the level of hemolysis seen with *E. coli* containing the chromosomal copy of the *rfaH* gene (2.2-fold for the vector-encoded *E. coli* RfaH), likely due to the difference in the concentration of RfaH in the cell. When overexpressed from this plasmid, the concentration of RfaH is at least 10-fold higher than in the *E. coli rfaH/H11001* cells, as determined by Western blotting (data not shown).

Postrecruitment defects in *V. cholerae* RfaH activity can be overcome at higher concentrations. Since, when overexpressed, all RfaH orthologs were able to substitute for the absent *E. coli* RfaH in vivo we reasoned that the reduced antitermination (Fig. 4) and elongation enhancement (Fig. 5)
by the RfaH orthologs, in particular the one from *V. cholerae*, in the *E. coli*-based in vitro transcription assay could be explained by the reduced affinity of the heterologous RfaH for *E. coli* RNAP due to a substantial divergence within the RfaH and RpoB/RpoC families. A corollary to this assertion is the hypothesis that such reduced affinity can be at least partially compensated for by increasing the concentration of one of the ligands.

To test this hypothesis, we added increasing amounts of a poorly performing (in this assay) RfaH from *V. cholerae* and determined its effect on pausing by the *E. coli* TEC, using a pIA349 template (3). In addition to the T7A1 promoter and *ops* site, this template featured a *P*_{his} pause site (Fig. 7). In the absence of RfaH, the TEC paused at both *ops* and *P*_{his} sites with half-lives of 28 and 79 s, respectively (Fig. 7). Addition of 40 nM *E. coli* RfaH delayed escape from the *ops* site by approximately fivefold (from 28 to 136 s) while reducing the pausing at the downstream *P*_{his} site more than twofold (*P*_{his} \( t_{1/2} \) dropped from 79 to 35 s). RfaH proteins from *S. enterica* serovar Typhimurium, *Y. enterocolitica*, and *K. pneu-

![Image](https://example.com/image.png)

**FIG. 7.** Increase in *V. cholerae* RfaH concentration suppresses its elongation defects. (Top) Transcript generated from the T7A1 promoter on a linear pIA349 DNA template. The transcription start site (+1), *ops* (boxed), *P*_{his} pause site, and transcript end are indicated. (Bottom) Halted G37 TECs were formed with *E. coli* RNAP and challenged with NTPs (20 \( \mu \text{M} \) GTP and 150 \( \mu \text{M} \) ATP, CTP, and UTP) and rifampin at 25 \( \mu \text{g/mL} \) in the absence (leftmost panel) or presence of proteins indicated below. Aliquots were withdrawn at the times indicated above each lane, followed by the high-concentration NTP chase as above. Positions of the halted (G37), paused (U43, C45, and *P*_{his}), and runoff transcripts are indicated with arrows. Pause half-lives at the C45 position and at the *P*_{his} site were determined to evaluate *ops* binding and the effect on elongation, respectively. Sizes of the molecular weight DNA markers (see Fig. 3) are indicated on the right.
moniae also reduced pausing at the P_his site more than twofold (data not shown). In contrast, assayed in the same conditions, the most divergent member of the RfaH family from V. cholerae affected pausing at the ops site similarly, increasing the pause half-life from 28 to 140 s, but had little (negative) to no effect on the P_his half-life (85 versus 79 s). We repeated the experiments with increasing concentrations of V. cholerae RfaH. Elevated to 100 and 250 nM, V. cholerae RfaH further increased delay in TEC escape from the ops site (from 140 to 180 and 300 s, respectively) but also reduced P_his pausing to a greater degree by decreasing the t_{1/2} from 85 s at 40 nM to 52 s at 100 nM and 49 s at 250 nM (Fig. 7). Notably, even the more than sixfold excess of V. cholerae RfaH did not reduce pausing at the P_his site to the level achieved by E. coli RfaH. These findings are consistent with our hypothesis that postrecruitment effects of RfaH depend on less-conserved (compared to RfaH-ops binding) interactions between RfaH and RNAP.

We also tested the effect of elevated concentrations of V. cholerae RfaH on T_{his} terminator readthrough. Consistent with the predictions of our hypothesis, an increase in the concentration of V. cholerae RfaH to 250 nM resulted in limited antitermination activity, reducing termination ~1.5-fold (data not shown). Thus, even when present at more than sixfold excess relative to the effective concentration of E. coli RfaH, its ortholog from V. cholerae was unable to fully restore the level of RfaH-dependent readthrough.

**DISCUSSION**

In this work we showed that orthologs of the E. coli transcription elongation factor RfaH from as evolutionary distant a bacterium as V. cholerae still exhibit characteristic effects on transcription by E. coli RNAP in vitro, whereas their cognate genes, when expressed episomally, can complement disruption of the chromosomal copy of E. coli rfaH.

RfaH proteins belong to a larger superfamily of NusG-like polypeptides, but unlike seemingly ubiquitous NusG proteins, RfaHs are found only among enterobacteria and related groups (Fig. 1; data not shown; 3). As evident from substantially greater branch lengths of the RfaH tree versus the NusG tree (Fig. 1), the RfaH polypeptide sequence evolved much faster than that of its paralog, which makes the ability of RfaH proteins to regulate transcription in heterologous systems reported here even more remarkable. The branching of the NusG-RfaH phylogenetic trees suggests that, restricted to proteobacteria, the rfaH gene arose from nusG through a duplication-divergence mechanism, in a way recycling a general elongation factor to create a specific regulator for a set of genes involved only in establishing pathogenicity and related functions. Some lineages possess, in addition to RfaH, other paralogs of NusG such as ActX of E. coli and Y. enterocolitica (Fig. 1); in others unambiguous placement of the distant NusG/RfaH homologs like AnfA1 (Serratia entomophila) or TaA (Myxococcus xanthus) is hindered by a scarcity of relevant genomic information. It does appear certain though that, despite their relatively small size, the proteins of the NusG/RfaH family possess enough versatility in their molecular architecture to perform functions of a general elongation factor (NusG), regulon-specific operon-polarity suppressor (RfaH), or single operon regulator (TaA).

We have previously formulated a two-step model of RfaH action. During the first step, RfaH is recruited to the ops-bound TEC through sequence-specific interactions with the non-template DNA strand; these interactions apparently have to be disrupted to allow RNAP escape from the ops site as RfaH induces a strong pause at position C45 (Fig. 3). In the second step, RNAP is shifted into a state in which it is less prone to pausing and termination (3). Although by analogy to other processive antitermination systems (38), we currently assume that, following its recruitment, RfaH remains associated with RNAP, the direct evidence for this retention is yet to be obtained. Existence of at least two distinct steps in RfaH action is supported by the data reported in this work, in particular by the behavior of RfaH from V. cholerae: when assayed in vitro with E. coli RNAP this ortholog was recruited to the elongation complex as well as E. coli RfaH but failed to stimulate transcription elongation to the same extent even at a six-times-greater concentration. This indicates that the recruitment and stimulatory steps can be uncoupled from each other, in this case by the changes in amino acid sequence accumulated since the branching of the E. coli and V. cholerae lineages. Presently we are investigating the feasibility of similar uncoupling of E. coli RfaH recruitment and postrecruitment events through limited radical mutagenesis of its functional signatures/domains.

We have proposed that RfaH has two separate structural modules, one that mediates the recruitment to the ops element and another that is responsible for the subsequent effects of RfaH on elongation, pausing, and termination. In agreement with this model, we observed an apparent disparity between a high degree of conservation for ops recruitment among divergent RfaH orthologs and greater variability of their effects on transcription by E. coli RNAP. We propose that the DNA-binding module of RfaH is more conserved than its RNAP-binding module(s) and that the disparity observed between recruitment and postrecruitment events in heterologous systems is due to a greater conservation of RfaH-ops interactions than those between RfaH and RNAP. While ops sequence divergence is limited and does not appear to vary with the evolutionary distance between different lineages (Fig. 2), RNAP subunits reveal a patchwork of conserved and variable segments (44). When amino acid sequence positions exhibiting greater and lesser degrees of variability are plotted onto three-dimensional structures of bacterial RNAP enzymes, more conserved residues appear to cluster in the interior of the RNAP, near the active site and architecturally important structures, whereas the least conserved residues tend to localize on or near the surface of the enzyme (unpublished observations; 44). Furthermore, extended lineage-specific insertions present in some bacteria also tend to form compact domains on the surface of the enzyme (5, 27). Although the exact location of the RfaH-binding site on the RNAP is not known, its initial recruitment occurs through the surface-exposed non-template strand of the ops site (3). Whether postrecruitment events take place in the same site or elsewhere on RNAP, contacts between RfaH and RNAP that mediate these events are more likely to involve the surface areas of both proteins rather than their interior. Thus, the reduced efficiency of evolutionarily distant orthologs in the E. coli transcription system can be
easily explained by the greater divergence of the interaction surfaces.

It is not known which regions of RfaH and NusG are responsible for various activities of these proteins. While the structure of RfaH is not yet known, two high-resolution X-ray structures of the *Aquifex aeolicus* NusG (12, 35) reveal that NusG proteins are composed of at least two separate domains. On the basis of the structure, the C-terminal KOW motif is proposed to mediate the nucleic acid interactions (35), whereas the determinants for the NusG interactions with other proteins could be located in the N-terminal PNP domain as well as in the KOW domain. In addition, the non-specific nucleic acid contacts could be mediated by a positively charged loop (in *E. coli* NusG) or a separate domain (in *A. aeolicus* NusG).

Since the putative protein contact sites do not overlap with the regions proposed to make contacts with the nucleic acids, NusG appears to be well designed to function as an adaptor mediating indirect protein-nucleic acid associations (35) that are likely necessary to explain its ability to stabilize macromolecular complexes, such as that of the Rho quaternary complex with the TEC (25). Presently, no biochemical data are available in support of these structural predictions, and the biological relevance of the contacts between NusG and the nucleic acids in vitro that have been reported (20) is unclear. However, the presence of two separate domains in NusG (and possibly in RfaH) could indicate that the recruitment and postrecruitment activities of these proteins could be indeed separable. The modest degree of similarity between RfaH and NusG is insufficient for building a good structural model for RfaH on the basis of the NusG structure. We are currently trying to obtain solution and X-ray RfaH structures in parallel with the ongoing mutagenesis of the *E. coli* RfaH.

Quantitative differences between elongation effects of various RfaH orthologs notwithstanding, they appear to employ the same mechanism of rendering *E. coli* RNAP less prone to pausing and termination. One line of evidence that argues in favor of this comes from partial recovery of in vitro activity through increased concentration of *V. cholerae* RfaH (Fig. 7). Another line of supporting evidence is provided by the in vivo complementation experiments, where all of the epistemally overexpressed *rfah* genes restored the hemolytic activity of the *E. coli* *rfah* mutant strain similarly (Fig. 6). RfaH orthologs studied here clearly displayed characteristic features of RfaH—delay of the TEC escape from the *ops* site (Fig. 3), antitermination at T*sho* (Fig. 4), reduction of pausing at the hairpin-dependent T*his* site (Fig. 7 and data not shown), and hemolysis (Fig. 6)—and thus appear to function like RfaH rather than like NusG, which demonstrates either no or opposite effects in these assays (2; data not shown), consistent with the presence of a bona fide *nusG* gene in each of these species (Fig. 1).

These findings also suggest that RfaH may play an important part in pathogenicity of *Vibrio* and *Yersinia* spp., similar to its role in *E. coli*. Deletion/disruption of the *rfah* gene in uropathogenic *E. coli* 536 leads to a substantial decrease in expression of hemolysin, the hemin receptor, and a number of other virulence factors, resulting in more than a 100-fold drop in the ability of the *rfah* mutant to colonize the urinary tract and an 82% decline in the associated mortality rate (23).

Mechanistic insights into the substantially conserved RfaH mode of action presented in this work should facilitate investigation of its largely overlooked role as a potential regulator of virulence genes in the infectious agents of cholera and bubonic plague and a number of other important pathogens, whose genomes encode RfaH-like factors (Fig. 1).

Since the surface-exposed lineage-specific insertions (5) in RNAP are utilized by specific transcription regulators such as Alc and T7 gp2 (24, 33), we considered a possible functional link between the *E. coli* lineage-specific insertions and the RfaH factor, both of which are found in proteobacteria. However, by testing *E. coli* RfaH with the RNAP variants missing one or more of these insertions, we established that the latter do not mediate RfaH effects on transcription (Svetlov and Artsimovitch, unpublished). We are currently in the process of elucidating the details of RfaH interactions with RNAP through chemical cross-linking and mutagenesis.

Together, the evolutionary history of RfaH factors and their activity in vitro and in vivo in *E. coli*-based assays are consistent with *rfah* arising from ubiquitous *nusG* genes early in the evolution of proteobacteria through a duplication-divergence mechanism before separation of *Vibrio* and *Escherichia* lineages rather than through horizontal transfer between existing groups. In each lineage RfaH appears to have evolved faster than its paralog, NusG, nevertheless retaining the ability to be recruited to heterologous RNAP and stimulate elongation downstream of the recruitment site.

**ACKNOWLEDGMENTS**

This work was supported by grants from the American Heart Association (0265013B) and the National Institutes of Health (GM67153).

**REFERENCES**


**Vol. 186, 2004 REGULATION OF RNA CHAIN ELONGATION BY RfaH ORTHOLOGS 2839**


