Characterization of an Unusual Rho Factor from the High G + C Gram-positive Bacterium Micrococcus luteus*

(Received for publication, September 18, 1995, and in revised form, November 2, 1995)

William L. Nowatzke and John P. Richardson†

From the Department of Chemistry, Indiana University, Bloomington, Indiana 47405

A transcription termination factor (Rho) was purified from the Gram-positive bacterium Micrococcus luteus, and the complete gene sequence was determined. The M. luteus Rho polypeptide has 690 residues, which is 271 residues more than its homolog from Escherichia coli. Most of the additional residues compose a highly charged, hydrophilic segment that is inserted in a nonconserved region between two conserved regions of the RNA-binding domain of the known Rho homolog proteins. This segment extends from residues 49 to 311 and includes a stretch of 238 residues that contain no hydrophobic side chains. Biochemical studies indicate that the M. luteus protein is very similar to E. coli Rho in terms of its RNA-dependent NTPase activity and its sensitivity to the Rho-specific inhibitor bicyclomycin. However, the M. luteus protein has a less stringent RNA cofactor specificity. It also acts to terminate RNA transcription with E. coli RNA polymerase on the λ cro DNA template, but at much earlier termination stop points than those recognized by E. coli Rho. Thus, the M. luteus protein functions as a true Rho factor, but with a different specificity than that of E. coli Rho. We propose that this altered specificity is consistent with its need to function on transcripts that have a high content of G + C residues.

The orderly expression of the genetic information in DNA segments into RNA molecules depends on the function of transcription terminators. In Escherichia coli, one mechanism of transcription termination is mediated in part by an essential protein factor called Rho (1). Rho factor from E. coli has been studied since its discovery nearly 25 years ago (2). The Rho monomer is a 47-kDa protein. However, Rho factor functions as a homohexamer (3) that can bind to a nascent transcript and mediate its release by actions on the transcription complex that are coupled to the hydrolysis of NTPs (1).

A recent phylogenetic study by Opperman and Richardson (4) comparing rho genes isolated from organisms from several of the major branches of bacteria suggests that Rho is ubiquitous throughout the bacterial domain. An unexpected result was discovered during the analysis of the rho gene from Micrococcus luteus, a Gram-positive soil bacterium that has an unusually high G + C DNA content (74%) (5). The M. luteus rho homolog was found to have an open reading frame encoding a protein that was homologous to E. coli Rho through a very long portion. However, the homology did not extend all the way through the RNA-binding domain toward the amino terminus of the protein. Because the region of homology starts in a segment that has an in-frame GTG codon preceded by a sequence that is a good match to a Shine-Dalgarno sequence, Opperman and Richardson proposed that translation began at that GTG codon to yield a 41,733-Da protein of 382 amino acids that is 52% identical (71% similar) to E. coli Rho. If this proposal were correct, the M. luteus Rho protein would be unusual in comparison with the other Rho homologs as it would lack a conserved part of its RNA-binding domain. Additionally, the protein would be ~30 amino acids smaller than any of the other predicted Rho factors that have been sequenced.

The data of Opperman and Richardson (4) were also consistent with an alternative hypothesis, namely that the M. luteus Rho polypeptide is much larger than the homologs from other organisms and includes a large region with a very unusual amino acid sequence. The DNA sequence determined in this work indicated that the open reading frame extended upstream for at least 160 amino acid residues. However, because the G + C content of that upstream region was ~78%, which is a value that is typical of intergenic spacer regions in M. luteus DNA (5), and because these upstream codons had a very unusual bias favoring Arg, Asp, Gln, and Gly residues and lacking hydrophobic residues, Opperman and Richardson argued that it was unlikely to be part of the coding region for the M. luteus Rho protein.

To resolve this issue, we purified Rho protein from M. luteus. Our studies show conclusively that the latter hypothesis is correct and demonstrate directly that an organism that is phylogenetically distinct from E. coli also has a factor that can cause the termination of RNA transcription.

EXPERIMENTAL PROCEDURES

Materials—Restriction enzymes, T4 DNA ligase, Vent DNA polymerase, and DNA polymerase I Klenow fragment were purchased from New England Biolabs Inc. Sequenase version 2.0 was purchased from U. S. Biochemical Corp. E. coli RNA polymerase was purchased from Epicentre Technologies Corp. E. coli Rho protein was provided by Lisa S. Stitt (Temple University). Bicyclomycin was obtained from Fujisawa Pharmaceutical Co. Ltd. (Osaka, Japan). NusG was a gift from Barbara Stitt (Temple University). Ribonuclease T1 was purchased from ICN Radiochemicals. Ribonucleotides were purchased from ICN Radiochemicals. Ribonucleotides, deoxyribonucleotides, and deoxyribonucleotides were purchased from Perkin-Elmer Cetus Corp. (Norwalk, CT). NusG was a gift from Barbara Stitt. Cloning vectors were obtained from Stratagene (La Jolla, CA) and Invitrogen (San Diego, CA).

ATPase Assay—ATPase activity was assayed colorimetrically as described by Lanzetta et al. (6). Typically, 50 ng of protein were mixed with 100 µl of assay solution (40 mM Tris-HCl, pH 7.7, 50 mM KCl, 10 mM MgCl₂, 1 mM ATP, and 10 µg/ml poly(C)). After 10 min at 37 °C, the
release of P<sub>i</sub> from ATP was detected by the addition of 10 mM of ATP to a mixture of 4.2 mM ammonium molybdate, 0.045% malachite green, concentrated flame photometer diluent (Bacharach, Inc.) (12.5:37.5:1) that had been premixed and filtered through Whatman No. 1 filter paper. Color development was quenched after 1 min by the addition of 100 µl of 34% citric acid. After 30 min at 20 °C, the absorbance at 660 nm was measured. One unit is defined as the amount that hydrolyzes 1 µmol of ATP/min. This assay was found to be quantitative from 1 to 15 nmol of P<sub>i</sub>.

Protein Sequencing—10 µg of the intact protein were bound to a strip of polyvinylidene difluoride membrane by direct adsorption (7). The N terminus was sequenced by William Lane at the Harvard University Microchemical Laboratory. To obtain N-terminal sequence information of a stable tryptic fragment, 16.8 µg of Rho protein was digested in the presence of 105 nm trypsin for 3 min at 37 °C in a total volume of 30 µl. Trypsin digestion was stopped by the addition of 10 µl of 4 × sample loading buffer (252 mM Tris-HCl, pH 6.8, 8% SDS, 50 mM glycine, 2% mercaptoethanol, and 0.004% bromophenol blue) followed by immersion of the mixture in a boiling water bath for 2 min. The products were separated by electrophoresis on an 8% SDS-polyacrylamide gel (8), and the protein bands were transferred to polyvinylidene difluoride membrane with a Pharmacia Biotech 2117 Multiphor II blotting apparatus. The transfer was carried out in Towbin buffer (25 mM Tris-HCl, pH 8.2, 192 mM glycine, 15% methanol) (9) for 30 min at 37 °C. After the addition of 40 µl of ATP reagent (10), ATPase was detected by Coomassie Blue staining. The band with an M<sub>r</sub> of 42,000 was removed, washed with double distilled H<sub>2</sub>O, air-dried, and stored at −20 °C. N-terminal sequencing was performed by Brian VanWuyckhuysen at the Protein Sequencing Facility at the University of Rochester Medical Center (Rochester, NY).

Isolation of the Upstream Portion of the M. luteus rho Gene—Genomic DNA was prepared from M. luteus as described by Wilson (10). The method employs the detergent sodium deoxycholate/SDS/Na<sub>2</sub>CO<sub>3</sub>/0.5 M NaCl/1 M KCl/1% Triton X-100 to remove proteins and polysaccharides. A typical yield from a saturated 100-ml culture was 0.7 µg of DNA.

To identify a 203-kbp<sup>1</sup> fragment containing the sequence encoding the N-terminal region of the M. luteus rho gene, 10 µg of genomic DNA were digested with various combinations of restriction endonucleases, and the products were separated by agarose gel electrophoresis. Fragments containing the desired rho gene segment were identified by Southern hybridization (11) with a 0.28-kbp HindIII/Smal fragment from the plasmid pMRLHOSK<sup>−</sup> (12), which had been radiolabeled to a high specific activity with <sup>125</sup>I (13).

1 It was determined that a BamHI/PstI double digest of M. luteus genomic DNA produced a ~2-2 kbp fragment that contained the desired portion of the M. luteus rho gene. To do this DNA fragment, 50 µg of M. luteus DNA were digested with BamHI and PstI, and the fragments were size-selected on a 1% agarose gel and ligated into pBluescript II SK<sup>−</sup> (Stratagene). Colonies of E. coli DH5<sup>α</sup> transformed with these ligation complexes were screened by nucleic acid hybridization (14) with the HindIII/Smal fragment used for Southern analysis.

DNA Sequencing—Both double-stranded and single-stranded DNA templates were utilized for sequencing reactions (15). Double-stranded templates were prepared as described in the Sequenase protocol manual (U.S. Biochemical Corp.). Single-stranded DNA was generated from pBluescript derivatives by infection with M13K07 helper phage (16). Reactions were carried out essentially according to the manufacturer’s instructions with the following modifications. 7-Deaza-dGTP was used in place of dGTP to reduce band compressions (17), and a dideoxy-dGTP/7-deaza-dGTP ratio of 1:15 and a dideoxy-CTP/dCTP ratio of 1:100 were used in the termination mixtures to enhance dG and dC signals.

RNA Transcription—A DNA fragment encoding the <i>λ cro</i> gene was prepared by amplification of the segment of the gene from residues 1–188 of <i>λ</i> DNA with Vent DNA polymerase. This plasmid contains a mutation of C to G at position 6 of the cro gene. Transcription complex formation was carried out according to Burns and Richardson (18) with slight modifications. In a reaction volume of 60 µl, 3 pmol of <i>λ</i> RNA polymerase were incubated with 3 pmol of DNA in transcription buffer (150 mM potassium glutamate, pH 7.8, 40 mM Tris-HCl, pH 7.8, 2 mM Mg<sub>2</sub>Cl<sub>2</sub>, 1 mM EDTA, 0.02% Nonidet P-40, 0.002% acetylated bovine serum albumin, 1% glyceral) for 3 min at 37 °C to form open complexes. The addition of 40 µl of NTP mixture (4 µM GTP, 4 µM ATP, 4 µM [α<sup>32</sup>P]UTP (350 nCi/pmml)) and incubation at 16 °C for 20 min resulted in the formation of stalled ternary complexes 24 residues downstream from the start point of transcription of AP<sub>r</sub> (A<sup>24</sup> complex). After the addition of 5 µl of rifampicin (1 mg/ml), the unincorporated NTPs were removed by ultrafiltration (20 min, 500 × g) with a Microcon-100 (Amicon, Inc.). The retentate (10 µl) was diluted to 200 µl with transcription buffer containing 36 units of RNasin (Promega). 10 µl of this diluted A<sup>24</sup> complex solution were used per reaction. After the addition of Rho factor, all four NTPs (including CTP) were added to 200 µM, and the 20-µl reaction mixture was incubated at 37 °C for 3 min. After the addition of 20 µl of 2 × stop mixture (20 mM EDTA, 0.1% SDS, 0.5 mg/ml RNA, 0.3 mg/ml proteinase K) and incubation at 37 °C for 20 min, the RNA was collected by ethanol precipitation and resuspended in 6 µl of loading dye (10 mM EDTA, 0.001% bromphenol blue, 0.001% xylene cyanol in 98% formamide). The entire sample was loaded on a 6% polyacrylamide gel (20 × 40 cm) containing 5% urea (14), and the RNA transcripts were separated by electrophoresis for 2 h at 30 watts.

RESULTS

Purification of an RNA-dependent ATPase from M. luteus—From preliminary studies, we found that the ATPase activity in partially fractionated extracts of M. luteus cells was stimulated by the addition of RNA homopolymers and that poly(C) was especially effective. Since E. coli Rho is an RNA-dependent ATPase that is strongly activated by poly(C) (19), we made use of an assay for ATP hydrolysis with poly(C) present for purification of a putative Rho factor from M. luteus. The purification procedure, which is described in detail elsewhere (20), involved chromatography of the crude extract on a Bio-Rex 70 cation exchange column (Bio-Rad), concentration of the pooled fractions containing poly(C)-dependent ATPase with a Centriprep-100 ultrafiltrator (Amicon, Inc.), and chromatography on heparin-Sepharose CL-6B resin (Pharmacia Biotech Inc.).

Analysis of the final fraction by electrophoresis on a 10% SDS-polyacrylamide gel (8) revealed that it consisted of a single polypeptide in >95% purity with an apparent M<sub>r</sub> of 95,000 (Fig. 1), which is approximately twice as large as the E. coli Rho polypeptide and also significantly larger than the M. luteus rho gene product (M<sub>r</sub> = 41,733) proposed by Opperman and Richardson (4).

A sample of this highly purified M. luteus ATPase, analyzed by the Microsequencing Facility at Harvard University, yielded an N-terminal amino acid sequence of TESTE, which is different from the sequence of MAGIL at the N terminus of the proposed M. luteus rho gene product (4). This sequence also did not match any other pentapeptide sequence in the segment of the M. luteus rho gene that had been sequenced prior to this work. However, a 42-kDa fragment generated from the 95-kDa protein by partial digestion with trypsin had the N-terminal sequence GRPGPEVDE, which did match a sequence located upstream of the previously proposed rho translational start site (12). Together, these results concerning the apparent size and
N-terminal sequence of the M. luteus RNA-dependent ATPase suggested that the M. luteus rho gene sequence reported by Opperman and Richardson (4) was not complete.

Location of the TESTE Sequence—The M. luteus rho gene identified in the previous study (12) was cloned as a 2.1-kbp SphI/SacI DNA fragment in the plasmid pMLRHOSK1. The remainder of the M. luteus rho gene was cloned as a 2-kbp BamHI/PstI DNA fragment into pBluescript II SK+ to create the plasmid pBN10. We intentionally chose a fragment that would overlap partially with the rho insert in pMLRHOSK1.

The DNA sequence of this fragment confirmed that the target DNA had been successfully isolated and contained 1200 base pairs of additional upstream M. luteus sequence. 111 base pairs (37 residues) upstream from the SphI site, we found a segment encoding the sought-after N-terminal pentapeptide sequence TESTE, which had been identified from microsequencing of the purified RNA-dependent ATPase. This result demonstrated that the RNA-dependent ATPase is indeed encoded by the gene identified by Opperman and Richardson (4). Although the open reading frame continues upstream for 108 base pairs from the codon of the first Thr residue, the next upstream codon is GTG, which is used as the start codon in 50% of the M. luteus genes (5), and is preceded by an excellent Shine-Dalgarno sequence (Fig. 2). We thus conclude that the mRNA transcript of the M. luteus rho gene starts at the GTG codon at position 1 and that the initiating Met residue is removed post-transcriptionally. The amino acid composition of the predicted protein from the completed gene sequence is in excellent agreement with the amino acid composition of the purified protein (Table I). The molecular mass of the M. luteus Rho protein is 74,957 Da; therefore, the protein runs anomalously (~95 kDa) in the Laemmli gel system (8) (Fig. 1).

Substrate Specificity—It has previously been demonstrated that the E. coli Rho protein is capable of catalyzing the hydrolytic conversion of any one of the four ribonucleoside triphosphates to the corresponding nucleoside diphosphate and Pi in the presence of an RNA cofactor such as poly(C) (21). This aspect of M. luteus was investigated by assaying for Pi release with each of the four different NTPs (Table II). M. luteus was able to utilize all four NTPs as well as ATP as substrates, but only when RNA was present. These results indicate that M. luteus Rho, like E. coli Rho, is an RNA-dependent NTPase. However, M. luteus Rho differs from E. coli Rho in having a significantly lower activity with CTP. The results also show that, on a molecular basis, M. luteus Rho catalyzed ATP hydrolysis at the same rate as E. coli Rho.

![Fig. 2. Nucleotide and predicted amino acid sequences of the M. luteus rho gene.](image)

| Amino acid | Protein analysis | Gene sequence |
|------------|-----------------|---------------|
| Asx | 13.2 | 13.0 |
| Gly | 12.3 | 12.4 |
| Ser | 5.8 | 6.5 |
| Gly | 11.3 | 10.7 |
| His | 1.2 | 1.0 |
| Arg | 12.0 | 12.6 |
| Thr | 6.3 | 4.3 |
| Ala | 11.3 | 10.7 |
| Pro | 4.7 | 4.6 |
| Tyr | 0.8 | 1.0 |
| Val | 5.3 | 5.6 |
| Met | 0.3 | 1.7 |
| Ile | 2.8 | 3.0 |
| Leu | 7.6 | 7.4 |
| Phe | 1.5 | 1.4 |
| Lys | 3.7 | 3.6 |
| Cys | 0 | 0 |
| Trp | 0.1 | 0.1 |
M. luteus Rho Has a Broader RNA Cofactor Specificity than E. coli Rho—To determine whether M. luteus Rho is significantly different from E. coli Rho with respect to its RNA-dependent ATPase activity, the rate of ATP hydrolysis of the two proteins was determined in the presence of various synthetic polynucleotide homopolymers (Table III). Both proteins were dependent on the presence of an RNA for ATP hydrolysis, and neither hydrolyzed ATP in the presence of poly(dc). Of the RNA polymers tested, poly(C) was the most effective activator for both proteins. M. luteus Rho, however, had appreciable activity with poly(A) as well as relatively high activity with poly(U) and measurable activity with poly(I). This was significantly different from E. coli Rho. Thus, the spectrum of RNA molecules that can activate ATP hydrolysis is greater for M. luteus Rho than for E. coli Rho.

To test whether the lower activity M. luteus Rho had with CTP was related to the RNA cofactor used, the rate of CTP hydrolysis was measured with poly(U) and was found to be ~30% of that with ATP (data not shown). Thus, the lower activity of M. luteus Rho with CTP was not a consequence of using poly(C) as an activator.

Bicyclomycin Inhibits the ATPase Activity of M. luteus Rho—Bicyclomycin is an antibiotic that has recently been shown to specifically inhibit E. coli Rho function (22). Mutants that exhibit bicyclomycin resistance contain mutations in the ATPase domain of the Rho protein (22). Because M. luteus Rho contains strong sequence homology in that conserved region, it was hypothesized that bicyclomycin would be an inhibitor of the M. luteus Rho protein. To investigate this, the standard ATPase assay was performed in the presence of increasing concentrations of bicyclomycin (Fig. 3). The results reveal that bicyclomycin is a potent inhibitor of M. luteus Rho. In the presence of 25 \( \mu \text{M} \) bicyclomycin, the lowest concentration tested, both the E. coli and M. luteus Rho proteins retain <30% of their poly(C)-dependent ATPase activity. Activity continues to decrease with increasing bicyclomycin concentrations and is nearly abolished at 200 \( \mu \text{M} \).

M. luteus Rho Terminates Transcription—To determine whether the RNA-dependent ATPase from M. luteus is a transcription termination factor, the purified protein was assayed for its effect on transcription of a \( \lambda \) cro template with E. coli RNA polymerase. This template has the well characterized Rho-dependent terminator tR\( \lambda \) (23). The control experiments show that starting with isolated complexes, incubation for 3 or 6 min in the absence of factors yielded a 372-nucleotide RNA (Fig. 4, lanes 4 and 5), the readthrough transcript from the promoter (P\( _{\lambda} \)) to the end of the template. When E. coli Rho factor was present at 28 \( \mu \text{M} \), a 3-min incubation yielded RNA molecules with ~290, ~312, and ~345 nt arising from termination at subsites I, II, and III (23), respectively, as well as the 372-nucleotide readthrough transcript (Fig. 4, lane 6). The overall termination efficiency was ~50% for this condition. The addition of the E. coli termination cofactor NusG, which has been shown to cause Rho-dependent termination at sites upstream of tR\( \lambda \) in vitro (24), yielded the expected pattern (Fig. 4, lane 7). When M. luteus Rho was used at the same concentration (28 \( \mu \text{M} \), a new set of discrete RNA molecules was formed with sizes in the range of 90–280 nucleotides (Fig. 4, lane 9), indicating that it caused termination at points well upstream from those used by E. coli Rho. The addition of E. coli NusG had only a small effect in enhancing the yield of smaller transcripts with M. luteus Rho (Fig. 4, lane 10). This effect of M. luteus Rho was not specific to tR\( \lambda \). Similar results were obtained with another E. coli Rho-dependent terminator, t45Z1, an intragenic terminator in the lacZ gene. M. luteus Rho terminated transcription at points earlier than E. coli Rho or E. coli Rho assayed in combination with NusG.2 This ability of M. luteus Rho to give rise to smaller RNA molecules during transcription of the cro template was completely blocked when 200 \( \mu \text{M} \) bicyclomycin was present in the reaction mixture (Fig. 4, lane 12), showing that this inhibitor of the ATPase activity of M. luteus Rho also inhibits its termination function.

To show that the smaller transcripts were the result of M. luteus Rho action as a transcription termination factor rather than as a ribonuclease, transcripts synthesized in the absence of Rho factor were subsequently incubated with M. luteus Rho (Fig. 4, lane 11). Although a small amount of a 145-nucleotide RNA appeared, the fact that no other RNA molecules appeared that had the same sizes as the products made when M. luteus Rho was present cotranscriptionally rules out the possibility that they were generated by a ribonuclease activity. Because very few of the transcripts were extended to the size of the readthrough RNA when M. luteus Rho was present during transcription, the overall efficiency of termination within the
Two lines of evidence suggest that the 145-nucleotide RNA arose as a result of a contamination of M. luteus Rho with a ribonuclease. First, the extent of appearance of the 145-nucleotide RNA was higher with other, less pure preparations of the M. luteus ribonuclease. When the function of M. luteus Rho was inhibited by bicyclic mycin (Fig. 4, lanes 12 and 13). A comparison of the distribution of transcripts in reaction mixtures lacking Rho that had been quenched at 2, 5, and 8 s after initiation (Fig. 4, lanes 1–3) with the distribution of the terminated transcripts (lanes 9 and 10) indicated that, as with E. coli Rho, the preferred positions for termination stop points were at the positions where RNA polymerase naturally pauses. However, with M. luteus Rho, the termination occurred at pause sites that were farther upstream than the pause sites that were used as the termination points by E. coli Rho.

**DISCUSSION**

We have isolated a transcription termination factor from M. luteus that is phylogenetically related to transcription termination factor Rho from E. coli. Although rho homologs have been identified from several different phylogenetic branches of bacteria (4, 25–28), this is the first demonstration that an organism that is distantly related to E. coli actually expresses its rho homolog gene. Although M. luteus Rho is similar to E. coli Rho in having a broad NTP substrate specificity, in its turnover number with poly(C) as a cofactor, and in its sensitivity to inhibition with bicyclic mycin, it differs in having a less stringent RNA cofactor specificity and in its specificity of termination during transcription of a coliphage gene with E. coli RNA polymerase. We have also found that M. luteus Rho differs from E. coli Rho in containing an extended insertion of very unusual sequence and likely structure within its RNA-binding domain.

M. luteus belongs to the phylogenetic branch called the high G + C Gram-positive group. The G + C content of its DNA is ~74% (5). In contrast, the G + C content of E. coli DNA is only 50%. In its function, the Rho factor of E. coli acts by binding to the nascent transcript at regions of the RNA called rut (rho utilization site). Although rut sequences lack a consensus (1), they do have certain specific, defining characteristics; they have little base-paired secondary structure (1) and usually have a compositional bias that is high in C residues and low in G residues (29). Because of their high G + C content, the RNA molecules in M. luteus are likely to have more extensive base pairing than the RNA molecules in E. coli. Thus, M. luteus Rho has likely been adapted to use a rut site that has more extensive base pairing than is typical for a rut site in E. coli. Evidence in support of this hypothesis is our finding that M. luteus Rho caused termination of transcription at a site located well before the rut site used by E. coli Rho on the λ cro gene template. The RNA encoded by the upstream region of λ cro forms extended base-paired secondary structures (30), thus making it unavailable as a rut site for E. coli Rho. This inter-

**Fig. 4**. *M. luteus* Rho terminates transcription. Ternary transcription complexes stalled on a λ cro template were prepared and elongated as described under “Experimental Procedures.” Lanes 1–5 show the distribution of RNA transcripts after elongation for 2 s, 5 s, 8 s, 3 min, and 6 min, respectively. Samples in lanes 6–13 were incubated for 3 min with the following additions: lane 6, 28 nM *E. coli* Rho; lane 7, 28 nM *E. coli* Rho and 25 nM NusG; lane 8, none, then for an additional 3 min with 28 nM *E. coli* Rho; lane 9, 28 nM M. luteus (Mlu) Rho; lane 10, 28 nM M. luteus Rho and 25 nM NusG; lane 11, none, then for an additional 3 min with 28 nM M. luteus Rho; lane 12, 28 nM M. luteus Rho and 200 μM bicyclic mycin; lane 13, none, then for an additional 3 min with 28 nM M. luteus Rho and 200 μM bicyclic mycin. The nucleotide lengths of the RNAs indicated at the right were determined by transcribing the cro template using RNA chain-terminating analogs (data not shown). RT, readthrough.

**Fig. 5**. Schematic representation of the *M. luteus* and *E. coli* Rho polypeptides. The *M. luteus* and *E. coli* Rho polypeptides have been drawn to scale. The relative positions of amino acid insertions are compared with *E. coli* Rho and are indicated by diverging lines. The black area in *M. luteus* Rho represents the 263-amino acid segment between Ile286 and Gly322, and the gray area the 10-amino acid segment between Lys364 and Gln375. The *E. coli* RNA-binding and ATP-binding domains are indicated by arrows below. The amino acids are numbered from the open reading frame.
hydrophobic side chain (excluding Pro and Ala residues). Since in a stretch of 238 residues, there are no amino acids with a C-terminal part rich in Arg, Asp, Gly, and Asn residues. Also, part of the insertion sequence is rich in Ala residues, while the residues instead of 14 amino acids with very little phylogenetic conservation. With its insert, M. luteus Rho has 263 residues instead of 14 in this putative loop region. The first part of the insertion sequence is rich in Ala residues, while the C-terminal part is rich in Arg, Asp, Gly, and Asn residues. Also, in a stretch of 238 residues, there are no amino acids with a hydrophobic side chain (excluding Pro and Ala residues). Since patterns of polar and nonpolar residues are important in the formation of ordered \( \beta \)-stranded and \( \alpha \)-helical secondary structures (32) and since hydrophobic residues have a major role in the formation of ordered tertiary structures for globular domains (33), we predict that this very hydrophilic segment of the protein will be randomly coiled, lacking a defined secondary structure. Indeed, when the insert sequence was analyzed for secondary structure (PHDsec Secondary Structure Prediction Program, EMBL, Heidelberg, Germany) (34–36), ~80% was predicted to exist as a loop. However, this segment has approximately an equal number of positively and negatively charged residues and might form an unprecedented, ordered structure consisting of many salt bridges.

The sequences of two other rho genes from this same group of organisms have recently become available: the genes from Streptomyces lividans\(^3\) and Mycobacterium leprae (GenBank\(^\text{TM}\) accession number U15186). The open reading frames of these genes predict Rho proteins with 706 and 610 residues, respectively. With both, the major part of the additional residues over the ~420 that are typical of Rho homologs from other phylogenetic groups start after Ile\(^{48}\) in S. lividans Rho and after Ile\(^{49}\) in M. leprae Rho and end before Gly\(^{396}\) and Gly\(^{397}\), respectively. Thus, S. lividans Rho has 263 and M. leprae Rho has 162 residues between these landmark residues. Like M. luteus Rho, S. lividans Rho has a major part that is very rich in Arg, Asp, Gly, and Glu residues, but is different in having many Glu residues instead of many Asn residues. The M. leprae sequence is also rich in polar residues. Like the M. luteus Rho insert, these sequences are very deficient in hydrophobic residues. These observations suggest that the presence of a polar, random-coiled structure insert is a conserved feature of the Rho proteins in these organisms that have a very high G + C content. However, in spite of the similar features, the three known Rho RNA-binding domain insertion sequences did not reveal any obvious phylogenetic relatedness. It will be of great interest to learn how the presence of a structurally unordered subdomain can help these Rho factors contend with their nascent transcripts to cause termination.

M. luteus Rho also contains another smaller insertion sequence that runs from Lys\(^{364}\) to Gin\(^{375}\) (Fig. 5; see Ref. 5). It is between two phylogenetically conserved residues in the RNA-binding domain corresponding to Glu\(^{108}\) and Arg\(^{119}\) in E. coli Rho. The S. lividans and M. leprae Rho homologs have insertions of three and six amino acids in that position, respectively. Like the large upstream insertion sequence, these lack amino acids with hydrophobic side chains.

Acknowledgments—We thank Barbara Stitt and Lislott Richardson for generously providing NusG and E. coli Rho, respectively. We also thank Fujisawa Pharmaceutical Co. Ltd. for providing bicyclomycin. We appreciate the insightful advice offered by Christopher Burns and thank Colin Ingham and Iain Hunter (University of Glasgow) for communicating the sequence of S. lividans rho prior to publication.

REFERENCES

1. Platt, T., and Richardson, J. P. (1992) in Transcriptional Regulation (McKnight, S. L., and Yamamoto, K. R., eds) pp. 365–388, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

2. Roberts, J. W. (1969) Nature 224, 1168–1174

3. Finger, L. R., and Richardson, J. P. (1982) J. Mol. Biol. 156, 203–219

4. Opperman, T., and Richardson, J. P. (1994) J. Bacteriol. 176, 5033–5043

5. Ghahremani, T., Muto, A., and Osaka, S. (1989) J. Mol. Biol. 209, 383–395

6. Lanzetta, P. A., Alvarez, L. J., Reina, P. S., and Candia, D. A. (1979) Anal. Biochem. 100, 95–97

7. Hugh, T. E. (1989) in Techniques in Protein Chemistry (Speicher, D. W., ed) pp. 28–30, Academic Press, Inc., San Diego, CA

8. Laemmli, U. K. (1970) Nature 227, 680–685

9. Towlson, H., Stahlhein, T., and Gordon, J. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 4350–4354

10. Wilson, K. (1989) in Current Protocols in Molecular Biology (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K., eds) pp. 2.1.1–2.4.5, John Wiley & Son, Inc., New York

11. Southern, E. M. (1975) J. Mol. Biol. 98, 503–517

12. Opperman, T., and Richardson, J. P. (1993) A Phylogenetic Comparative Analysis of the Predicted Amino Acid Sequence of Transcription Termination Factor Rho. Ph.D. thesis, Indiana University

13. Feinberg, A. P., and Vogelstein, B. (1983) Anal. Biochem. 132, 6–13

14. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

15. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463–5467

16. Vieira, J., and Messing, J. (1987) Methods Enzymol. 153, 3–11

17. Barr, I., Thayer, R. M., Layborn, P., Najarian, R. C., Seela, F., and Tolan, D. R. (1986) BioTechniques 4, 428–432

18. Burns, C. M., and Richardson, J. P. (1995) Proc. Natl. Acad. Sci. U.S.A. 92, 4738–4742

19. Lowery, C., and Richardson, J. P. (1977) J. Biol. Chem. 252, 1381–1385

20. Nishizato, K. W., Richardson, L. V., and Richardson, J. P. (1996) Methods Enzymol., in press

21. Lowery, C., and Richardson, J. P. (1977) J. Biol. Chem. 252, 1375–1380

22. Ziewlka, A., Kohn, H., and Widger, W. R. (1993) Biochemistry 32, 3564–3570

23. Lau, L. F., Roberts, J. W., and Wu, R. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 6171–6175

24. Li, J., Mason, S. W., and Greenblatt, J. (1993) Genes Dev. 7, 161–172

25. Tilly, K., and Campbell, J. (1993) Nucleic Acids Res. 21, 1040

26. Pinkham, J. L., and Platt, T. (1983) Nucleic Acids Res. 11, 3531–3545

27. Quirk, P. G., Dunkley, E. A., Lee, P., and Kruilwich, T. A. (1993) J. Bacteriol. 175, 647–654

28. Miloso, M., Limauro, D., Dallafiora, P., Ivilinelli, A., Galletta, A., Eriti, B., and Bruni, C. B. (1993) J. Bacteriol. 175, 8030–8037

29. Alifano, P., Rivellini, F., Limauro, D., Bruni, C. B., and Carloni, M. S. (1991) Cell 64, 553–563

30. Faus, I., and Richardson, J. P. (1990) J. Mol. Biol. 212, 53–66

31. Morgan, W. D., Bear, D. G., and von Hippel, P. H. (1984) J. Biol. Chem. 259, 8964–8971

32. Bowie, J. U., and Sauer, R. T. (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 2152–2156

33. Kauzman, W. (1959) Adv. Protein Chem. 14, 1–63

34. Rost, B., and Sander, C. (1993) J. Mol. Biol. 232, 584–599

35. Rost, B., Sander, C., and Schneider, R. (1994) Comput. Appl. Biosci. 10, 53–60

36. Rost, B., and Sander, C. (1994) Proteins Struct. Funct. Genet. 19, 55–72

\(^3\) C. Ingham and I. Hunter, personal communication.