

A high-affinity interaction between NusA and the *rrn nut* site in *Mycobacterium tuberculosis*

Kristine B. Arnvig^{*†}, S. Pennell[‡], B. Gopal^{*§}, and M. J. Colston^{*¶}

Divisions of ^{*}Mycobacterial Research and [†]Protein Structure, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, United Kingdom

Edited by Sankar Adhya, National Institutes of Health, Bethesda, MD, and approved April 26, 2004 (received for review February 24, 2004)

The bacterial NusA protein enhances transcriptional pausing and termination and is known to play an essential role in antitermination. Antitermination is signaled by a nut-like cis-acting RNA sequence comprising *boxB*, *boxA*, and *boxC*. In the present study, we demonstrate a direct, specific high-affinity interaction between the *rrn* leader nut-like sites and the NusA proteins of *Mycobacterium tuberculosis* and *Escherichia coli*. This NusA–RNA interaction relies on the conserved region downstream of *boxA*, the *boxC* region, thus demonstrating a key function of this element. We have established an *in vivo* assay for antitermination in mycobacteria and use this to show that the *M. tuberculosis rrn* nut-like site enhances transcriptional read-through of untranslated RNA consistent with an antitermination signal within this site. Finally, we present evidence that this NusA–RNA interaction affects transcriptional events further downstream.

NusA (N-using substance A) is a highly conserved transcription factor involved in transcriptional pausing, termination, and antitermination (reviewed in ref. 1). The protein is also thought to be essential for the increased elongation rate associated with *rrn* transcription and to be a constituent of the *rrn* antitermination complex (2, 3).

The *rrn* antitermination complex ensures that transcription of rRNA is not terminated by Rho despite the nascent transcripts being untranslated and thus naked (4). The complex is believed to include RNA polymerase (RNAP), Nus factors A, B, E, and G, ribosomal protein S4, and possibly one or more additional cellular components (5–10). The antitermination complex is assembled after transcription of a nut-like site (hereafter referred to as *nut* site for brevity) consisting of *boxB*, *boxA*, and *boxC* within the *rrn* leader region (1, 4). However, it has been shown that *boxA* is sufficient to obtain antitermination (7, 11). It is not entirely clear how the assembly proceeds, but evidence suggests that binding of a NusB/NusE heterodimer to *boxA* RNA may play a part, whereas the exact roles of NusA, *boxB*, and *boxC* remain unclear (6, 7, 11, 12).

Bacterial NusA proteins harbor three RNA-binding domains: S1, K homology 1 (KH1), and KH2 as well as an N-terminal domain that interacts with the RNAP (13–15). KH domains, like ribosomal protein S1 domains, are found in a variety of nucleic acid-binding proteins, often in multiple copies within one protein (16). Structural evidence suggests that KH domains interact sequence-specifically with four to five consecutive nucleotides in single-stranded regions (17–20). Comparable structures of S1 domains in complex with nucleic acids are not available. Based on the NusA structure of *Thermotoga maritima*, Worbs *et al.* (15) proposed that the three RNA-binding domains in the NusA protein act together to form an “extended RNA binding surface” consistent with the ability to bind RNA with high affinity and specificity. NusA interacts with *λboxAB* (21) as well as with mRNA (22). Although bacterial NusA proteins display a high degree of conservation, there is variability in the C-terminal region. Thus, in addition to the four domains present in the *Mycobacterium tuberculosis* protein, the *Escherichia coli* protein has a C-terminal domain, which in the absence of accessory factors masks one or more of the RNA-binding domains (14, 21).

M. tuberculosis harbors a single, nonredundant *rrn* operon, the regulation of which therefore comprises an obvious target for new therapeutic agents. The *M. tuberculosis rrn* leader and spacer regions contain sequences with homology to the *E. coli rrn* antitermination motifs, *boxB*, *boxA*, and *boxC*, with similar characteristics and organization (11, 23). The region spanning *boxA* and *boxC* shows 100% sequence conservation between *rrn* leader regions of different species of mycobacteria, suggesting an essential function, whereas the *boxB* element, comprising a stem-loop structure, has no strict sequence conservation between species or between *rrn* operons. The motifs are all found within the first half of *M. tuberculosis rrn* leader sequence (23). In this study we investigated interactions between recombinant *M. tuberculosis* NusA and the *M. tuberculosis rrn* leader region *in vitro* and identified a sequence-specific, high-affinity interaction between the *rrn nut* site and NusA. A similar interaction was observed between the *E. coli* NusA protein and the *E. coli rrnG* leader region. Using RNA footprinting, we identified a number of nucleotides within the *M. tuberculosis nut* site that seem to be involved in the interaction with NusA and subsequently identified a point mutation (U47A) in this region that interferes with NusA binding. We established an assay to measure antitermination *in vivo*. The assay indicated that the *M. tuberculosis rrn nut* site promotes antitermination similar to the *E. coli rrn nut* site. When the U47A mutation was introduced into the *M. tuberculosis nut* site, we observed a significant reduction in the transcriptional read-through of *lacZ*, suggesting that NusA binding affects the antitermination mechanism.

Materials and Methods

Plasmids and Strains. *E. coli* DH5 α was used for all cloning purposes, and *E. coli* BL21 DE3 pLysS was used for protein expression. *Mycobacterium smegmatis* mc²155 (24) was transformed by electroporation. The NusA clone has been described (13). Plasmid templates for electrophoretic mobility-shift assay (EMSA) probes were made by cloning the indicated fragments of the *M. tuberculosis rrn* leader region as *Hind*III fragments into pGEM3Z+ (Promega). The *E. coli rrnG* leader sequence nucleotides +1 to +64, according to the sequence in ref. 7, was cloned as an *Eco*RI–*Xba*I fragment into pGEM3Z+. To generate homogeneous RNA for footprinting a plasmid, pRBZ1/63, was made to obtain the *M. tuberculosis nut* site between a 5' hammerhead ribozyme and a 3' hepatitis δ ribozyme (25). pUC1198V, a derivative of pUC18T7Pst δ V (26) containing a T7 promoter and a 3' hepatitis δ ribozyme, was used as vector to generate the self-cleaving construct, pRBZ1/63. Two complementary oligonucleotides containing the 5' hammerhead ri-

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: RNAP, RNA polymerase; KH, K homology; EMSA, electrophoretic mobility-shift assay.

[†]To whom correspondence should be addressed. E-mail: karnvig@nimr.mrc.ac.uk.

[§]Present address: Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India.

[¶]Deceased February 20, 2003.

© 2004 by The National Academy of Sciences of the USA

bozyme and the *M. tuberculosis* *rrn* *nut* site [+1 to +63 relative to the transcription start point of the PCL1 promoter (tspPCL1)] with *Xba*I–*Pst*I overhangs were annealed and cloned into pUC119δV. The 5′ hammerhead sequence was designed to direct the proper folding and subsequent cleavage of the transcript, which depends on the 5′ end of the desired final product (see ref. 26 for details). For the *in vivo* antitermination assay we made a plasmid, p414_{UT}, derived from pEJ414 (27), which is an integrating mycobacterial vector with a promoterless *lacZ* gene. To obtain an untranslated *lacZ* transcript we changed pEJ414 to p414_{UT} by site-directed mutagenesis using the QuikChange kit (Stratagene) with mutating primers according to manufacturer instructions. Thus the forward primer has the sequence 5′-GCTAAAATGGATGATCCCGTAGTTTAAACAACGTCGTGACTGGG-3′ (with the stop codons underlined). The reverse primer is complementary to the forward primer. pKA103 (P_{rrn}-*leader*_{wt}-*lacZ*) was made by cloning the *M. tuberculosis* *rrn* promoters, leader region, and first 30 nt of the mature 16S (–194 to +219 relative to tspPCL1) into p414_{UT} as a *Spe*I–*Hind*III fragment into the *Xba*I and *Hind*III sites of p414_{UT}. pKA104 (P_{rrn}-*leader*_{U47A}-*lacZ*) was made by site-directed mutagenesis of pKA103 by using mutating primers; the sequence of the forward primer is 5′-GCGTGTGTTTGGAGAACAACAATAGTGTGTTTGGTAAGC-3′ (with the U47A substitution underlined). The reverse primer is complementary to the forward primer. pKA107 (P_{rrn}-*nut*_{wt}-*lacZ*) was made by cloning the *M. tuberculosis* *rrn* promoters and *nut* site (–194 to +63 relative to tspPCL1) into p414_{UT} as a *Spe*I–*Hind*III fragment into the *Xba*I and *Hind*III sites of p414_{UT}. pKA108 (P_{rrn}-*nut*_{U47A}-*lacZ*) was made by site-directed mutagenesis of pKA107 by using the same mutating primers as for pKA104. pKA109 was made by cloning the *M. tuberculosis* *rrn* promoter region (–194 to +1 relative to tspPCL1) into p414_{UT} by using the same sites as described above. Templates for *lacZ* probes were made by cloning 5′ and 3′ PCR fragments of *lacZ* (from p414_{UT}) into pGEM3Z+ as *Hind*III–*Eco*RI fragments. The 5′ insert (*lacZ* 1–125) was made with primers 5′-ATATAAGCTTATGGATGATCCCGTAGTTAAACAACG-3′ and 5′-TATAGAATCCGATCGGTGCGGGCCTCTTCGC-3′ in a standard PCR with p414_{UT} as template and *Pfu* polymerase (Stratagene). The 3′ insert (*lacZ* 2778–3057) was made accordingly with primers 5′-AAGCTTCCGAGCGAAAACGG-3′ and 5′-TTTTTGACACCAGACCAACTG-3′.

In Vitro Transcription and RNA Preparation. For probes, 1 μg of template DNA was linearized and used in a standard *in vitro* transcription reaction with 50 μCi (1 Ci = 37 GBq) of [α -³²P]UTP 800 Ci/mmol (Amersham Pharmacia) by using SP6 (or T7 in the case of *rrnG*) according to manufacturer instructions (Roche, Gifp-Oberfrick, Switzerland). After transcription, the template was digested for 30 min with 10 units of RNase-free DNaseI (Roche). The RNA was purified by using RNeasy columns (Qiagen, Valencia, CA), counted in a scintillation counter, and the concentration was determined from the activity. The ribozyme construct was linearized with *Pvu*II and transcribed in a standard large-scale T7 unlabeled reaction. Transcription of pRBZ1/63 produces a 382-nt transcript, which undergoes autocatalytic cleavage to release the 63-mer RNA. The RNA species were separated on a 15% sequence gel, and the desired fragment was purified from the gel by eluting overnight into PAGE elution buffer (0.5 M ammonium acetate/10 mM magnesium acetate/1 mM EDTA/0.1% SDS) at 4°C. The eluted RNA was filtered through a 0.2-μm sterilizing filter, ethanol-precipitated, and resuspended in water. A standard labeling reaction with 10 μg of RNA and [γ -³³P]ATP (Amersham Pharmacia) and T4 polynucleotide kinase (Promega) was performed. Because of degradation of RNA in the labeling reaction, a second gel purification step, analogous to the first, was

performed. The RNA was diluted to 100,000 cpm/μl and incubated with the indicated enzymes to identify regions of protein interaction.

Protein Expression and Purification. The NusA protein was expressed and purified essentially as described (13). The purity of the protein (>95%) was confirmed by SDS/PAGE and UV spectroscopy and subsequently used in EMSAs. The *E. coli* protein was subjected to an additional heparin purification step: the protein was dialyzed into 50 mM Tris/50 mM NaCl, pH 7.0. After recovery, the protein was added to the heparin column (Hi-trap, Amersham Pharmacia) and washed with 50 mM Tris/50 mM NaCl, pH 7.0 and eluted with 1 M NaCl. After dialysis into 50 mM Tris/300 mM NaCl, pH 7.0, the protein was concentrated and used in EMSAs.

EMSAs. Binding of NusA to the RNA was performed in 10-μl reactions in 20 mM Hepes, pH 7.5/0.2 mM EDTA/10 mM (NH₄)₂SO₄/1 mM DTT/15 mM MgCl₂/2 μg/ml tRNA/50 μg/ml BSA/0.25% Triton X-100/10% glycerol/20 units of RNasin (Promega) and the indicated amount of KCl. Samples were incubated for 15 min at room temperature before loading onto a native 8% acrylamide gel and run at 35 mA for 2.5 h in 1× TBE (100 mM Tris/100 mM boric acid/5 mM EDTA, pH 7.0). After drying, the RNA was visualized on a Storm860 scanner.

RNA Footprinting. Samples with NusA were incubated for 5 min with 250 nM NusA at room temperature before adding the enzymes. RNase T1 (specific for unpaired G residues) was obtained from Roche. T1 buffer (200 mM NaCl/10 mM MgCl₂/10 mM Tris, pH 7.5) was mixed with 100,000 cpm of RNA and 10 μg of tRNA in a 50-μl volume. T1 stocks (1 μl) of 0.1, 0.2, or 0.4 units/μl were added, and the samples were incubated on ice for 20 min. Reactions were stopped by adding 150 μl of ethanol, and samples were stored overnight at –20°C. RNase CV1 (specific for double-stranded regions and stacked bases) was obtained from Ambion (Austin, TX), and the procedure was exactly as described by the manufacturer using 100,000 cpm of RNA. Samples were incubated overnight at –20°C and precipitated by centrifugation at 13,000 × *g* for 20 min. After washing in 96% ethanol, the samples were dried in a desiccator and mixed with 5 μl of formamide loading buffer (Ambion) and separated on a 15% acrylamide sequencing gel. An alkaline hydrolysis ladder was prepared by drying 300,000 cpm of RNA with 10 μg of tRNA in a desiccator and resuspending in 3 μl of Na₂CO₃/NaHCO₃ (25 mM each, mixed in a 2:18 ratio). Samples were incubated at 95°C for 1 min and kept on ice before adding 5 μl of formamide loading buffer. After electrophoresis, the gel was fixed for 15 min in 10% acetic acid/25% methanol/2% glycerol, dried, and exposed to autoradiography.

RNA Isolation and Northern/Slot Blotting. The strains were inoculated directly from 7H11 plates into 50 ml of Dubos broth (27). Cultures were grown overnight to an OD of 0.65 < A₆₀₀ < 0.9 before harvest. RNA was isolated by means of the FastRNA kit (Q-Biogene, Cambridge, U.K.) according to manufacturer instructions. Northern and slot blots were prepared according to standard procedures. For Northern blotting, doublet samples of 5 μg of total RNA were loaded on a 1.5% denaturing agarose gel and electrophoresed at 75 V/cm for 3 h. After electrophoresis, the gel was cut in half; one half was stained in ethidium bromide, and the other half was transferred to a nylon membrane by using upward capillary transfer. The membrane was air-dried, and the RNA was cross-linked to the membrane by UV irradiation. For slot blots, three individual cultures were made from each strain, and samples were loaded in duplicate with 5 μg of RNA per slot. Each blot was made in quadruplicate, two of which were probed with 5′ *lacZ* and two with 3′ *lacZ*, and the bands were quantified

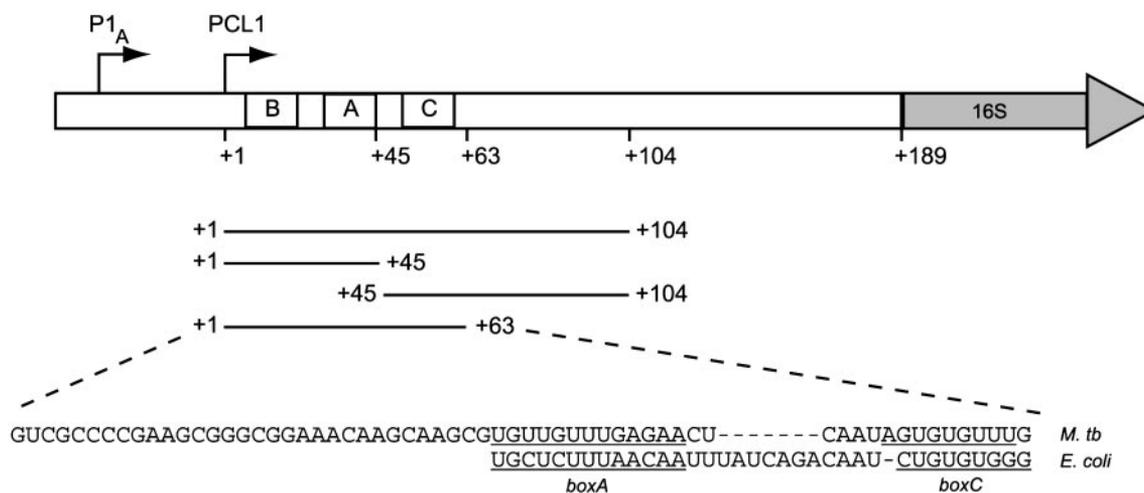


Fig. 1. The *M. tuberculosis* *rrnA* leader and spacer region. The transcription start points of the two promoters, P1_A and PCL1, are indicated with arrows, and the position of *boxB* (+2 to +19), *boxA* (+33 to +45), and *boxC* (+52 to +59) are indicated. The probes used in this study correspond to PCL1 transcripts and are shown with their 5' and 3' boundaries. The sequence of the *M. tuberculosis* (*M. tb*) *rrn* nut-like site (+1 to +63) are shown, and the *boxA* and *boxC* elements are underlined and aligned to the corresponding *E. coli* sequences.

on a Storm860 scanner. The 3'/5' ratio was calculated for each sample and normalized to the average values for KAM107, which had the highest read-through. The unnormalized average values for KAM107 were 388,076 cpm (5' probe) and 208,174 cpm (3' probe).

Results

A Specific Interaction Between NusA and the *rrn* Leader Region. The *M. tuberculosis* NusA protein is highly homologous to the *E. coli* NusA protein. One major difference between the two proteins is that the C-terminal domain, which in *E. coli* seems to mask the RNA-binding domain(s), is absent from the *M. tuberculosis* protein, which therefore should enable the *M. tuberculosis* protein to bind RNA without the need for accessory factors (13, 14). To investigate whether this indeed was the case, we used EMSAs to identify putative interactions between *M. tuberculosis* NusA and the *M. tuberculosis* *rrn* leader region.

Initially we investigated the interaction between recombinant *M. tuberculosis* NusA and the first half of the leader region, i.e., nucleotides +1 to +104 of the PCL1 transcript, which includes the putative antitermination motifs *boxB*, *boxA* and *boxC* (Fig. 1). The region was transcribed *in vitro*, and the transcripts were used as probes in EMSAs to determine whether NusA interacts and forms stable complexes with the *rrn* leader. The results, shown in Fig. 2A, demonstrate a clear interaction between *M. tuberculosis* leader RNA and NusA at protein concentrations as low as 5 nM (Fig. 2A, lane 2). Complex formation is significantly stronger with the sense transcript, and only a faint band is observed with the antisense probe (Fig. 2A, lanes 6–8). The addition of purified *M. tuberculosis* RNAP α subunit to the reaction had no effect on complex formation (data not shown). These results demonstrate that the *M. tuberculosis* NusA protein binds to the first half of the *rrn* leader *in vitro* with high affinity. Moreover, the binding does not require accessory factors such as RNAP α or λ N, which are required for RNA binding of the *E. coli* NusA protein (14, 21).

NusA Binds to the *M. tuberculosis* *nut* Site. To determine more accurately which part of the *rrn* leader interacts with NusA, we divided the 104-nt transcript into two shorter, contiguous probes: one comprising nucleotides +1 to +45, which includes *boxB* and *boxA*, and the other comprising nucleotides +45 to +104, which includes *boxC* (Fig. 1). The probes were used in EMSAs as

described before, but neither of the probes formed distinct complexes with NusA at protein concentrations up to 50 nM (data not shown), which suggests that the observed high-affinity interaction relies on regions in both halves of the 104-nt leader.

We therefore tested whether the *M. tuberculosis* *nut* RNA, i.e.,

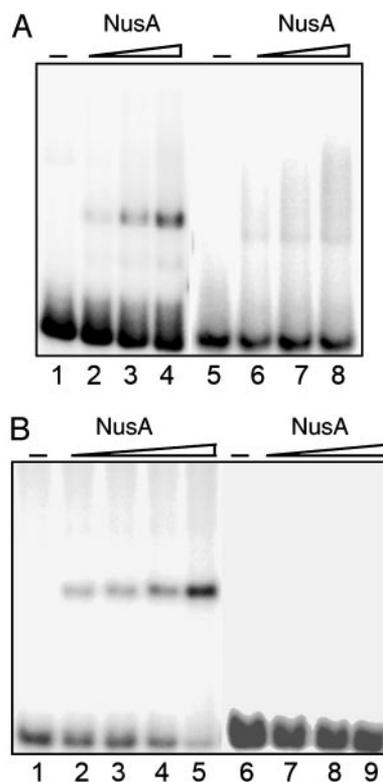


Fig. 2. Interactions between *M. tuberculosis* NusA and the *rrn* leader region. (A) EMSA with the first half of the *rrn* leader region (+1 to +104) as probe. Lanes 1–4, 1 nM sense RNA; lanes 5–8, 1 nM antisense RNA. NusA concentrations are 5, 10, and 20 nM in 200 mM KCl. (B) EMSA with 1 nM *nut* RNA (+1 to +63). Lane 1, sense RNA; lanes 2–4, as described for lane 1 with 0.5, 1, 5, or 10 nM NusA; lane 5, antisense RNA; lanes 6–8, as described for lane 5 with 10, 25, 50 nM NusA, all in 1 M KCl.

boxes *B*, *A*, and *C* of the *rrn* leader, would be sufficient for the interaction. Thus, we made a probe consisting of nucleotides +1 to +63 of the PCL1 transcript and used it in EMSAs (shown in Fig. 2*B*). This figure clearly demonstrates that the *M. tuberculosis nut* RNA forms distinct complexes with NusA. Complex formation is observed at protein concentrations as low as 0.5 nM NusA, indicating binding with very high affinity (at 1 M KCl). The antisense transcript forms no visible complexes with NusA at protein concentrations up to 50 nM, indicating that the interaction between NusA and RNA is specific for boxes *B*, *A*, and *C*. These results demonstrate that *M. tuberculosis rrn nut* RNA comprises a high-affinity and high-specificity binding site for NusA.

Nucleotides of the *rrn* Leader Involved in NusA Interaction. With the aim of obtaining information about specific nucleotides involved in the NusA–RNA interaction, we decided to investigate the interaction by footprinting of the NusA–RNA complex. To obtain large amounts of RNA without vector sequences, the *M. tuberculosis nut* site was cloned into a ribozyme-flanked construct, pRBZ1/63. Binding of NusA to the resulting RNA was confirmed by an EMSA, which indicated that the buffers used for the digests reduced the affinity of the NusA protein for the RNA. We therefore increased the NusA concentration for the footprinting experiments. The RNA, alone or in complex with NusA, was subjected to limited digestion with RNases T1 and CV1 and separated by denaturing electrophoresis. Although no cleavage sites were fully protected by the presence of the NusA protein, some changes in reactivity, seen as changes in band intensity, were observed. The most prominent changes were observed in the *boxC* region (i.e., +46 to +63) with RNase CV1 at positions C46, C48, and A49 and RNase T1 at positions G55 and G57 (Fig. 3*A*). According to an MFOLD computer model (28) of the *nut* RNA, these positions are located around a noncanonical U·U base pair in the third stem of the structure shown in Fig. 3*B*. Because noncanonical base pairs are often involved in sequence-specific RNA–protein interactions (29), we decided to alter the U·U base pair to an A·U base pair to determine whether it would interfere with the NusA binding. The theoretical structure of the mutant probe was tested with MFOLD, and it matched the wild type, i.e., a three-stem–loop configuration as the principal option, although the stability of the predicted structure was higher because of the A·U base pair. The mutant probe, U47A, was *in vitro* transcribed as described before and used in an EMSA (shown in Fig. 4). The figure illustrates how the single point mutation affects the mobility as well as complex formation. Thus, the free U47A transcript moves slightly faster than the free wild-type transcript (compare lane 1 with lane 4), consistent with a more compact structure. More importantly, the U47A mutation eliminated complex formation (compare lanes 2 and 3 with lanes 5 and 6), thereby confirming the requirement for the *boxC* region in the NusA–RNA interaction.

Antitermination in *M. tuberculosis*. To determine the putative antitermination effect of the *M. tuberculosis nut* site as well as the significance of the NusA–RNA interaction *in vivo*, we made a series of plasmids to measure antitermination in mycobacteria. The mycobacterial Rho protein has not been characterized, and hence no Rho-dependent terminators have been identified in mycobacteria. However, the *Micrococcus luteus* Rho protein, which is highly homologous to the mycobacterial Rho proteins, recognizes the Rho-dependent terminator, *tiZ1*, located ≈200 bp into the *E. coli lacZ* gene (30). Assuming, therefore, that the *M. smegmatis* Rho protein would recognize at least one of the Rho-dependent *lacZ* terminators, we designed a *lacZ*-based construct for measuring antitermination in mycobacteria. The basic construct (KAM109) is an integrating vector harboring the two *M. tuberculosis rrn* promoters upstream of the *lacZ* gene. To

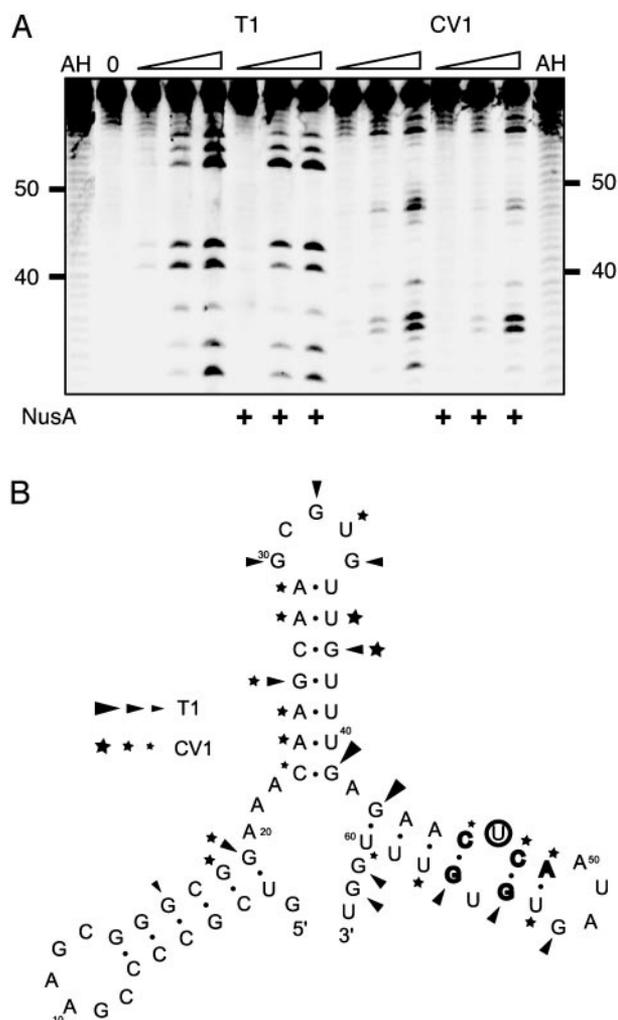


Fig. 3. Footprinting *M. tuberculosis nut* RNA. (A) Reactions with RNases T1 (0.1, 0.2, and 0.4 units) and CV1 (10^{-1} , 10^{-2} , and 10^{-3} units) in the region from +30 to +63 with and without 250 nM NusA. AH denotes alkaline hydrolyses ladder and 0 is RNA only. (B) Structure for nucleotides +1 to +63 predicted with the MFOLD program. Letters in bold indicate positions that were most affected by NusA, and U47 is circled. Regions cut with T1 and CV1 are indicated by triangles and stars, respectively, where the size of the symbol indicates the degree of cleavage.

prevent translation so as to mimic the *rrn* antitermination scenario we introduced stop codons in all three reading frames within the first 10 codons of *lacZ*. Plasmids containing the *M. tuberculosis rrn* full-length leader (KAM103) or *nut* site (KAM107) in their native positions relative to the *rrn* promoters, and thus comprising the 5' end of the *lacZ* transcripts, were made. All plasmids were transformed into *M. smegmatis* mc²155 (24), and total RNA from exponentially growing cultures was isolated. Initially we compared the termination pattern of the constructs by Northern blotting. The RNA was separated by denaturing electrophoresis and hybridized to a probe complementary to the 5' end of *lacZ* (nucleotides 1–125). The blot in Fig. 5 demonstrates that the band pattern differs significantly between the three strains. Note that, because of the different 5' leaders, KAM103- and KAM107-derived transcripts will be 219 and 63 nt longer than the corresponding KAM109 transcripts. However, these differences are obviously most pronounced in the smallest species and negligible in the larger transcripts. There is no signal in the lane containing RNA isolated from the control strain, mc²155, indicating that all hybridization is *lacZ*-specific

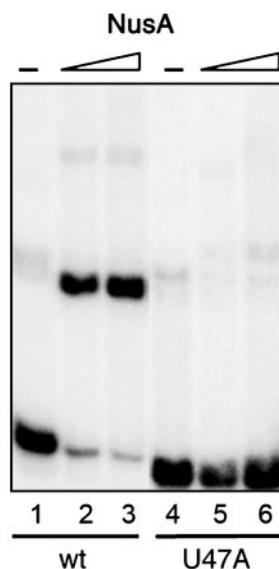


Fig. 4. EMSA with the mutated U47A *nut* RNA probe in 1 M KCl. Lane 1, 1 nM wild-type (wt) *nut* RNA; lanes 2 and 3, as described for lane 1 with 10 and 25 nM NusA; lane 4, 1 nM U47A *nut* RNA; lanes 5 and 6, as described for lane 4 with 10 and 25 nM NusA.

(data not shown). The blot displays *lacZ*-derived distinct transcripts ranging from ≈ 0.2 to >3.0 kb in size, the latter consistent with a full-length *lacZ* transcript. The large transcript is most prominent in the strain with the *nut* insert (KAM107; Fig. 5, lane 2) indicating that in this strain, a significant proportion of the elongation complexes persist throughout the entire length of the *lacZ* gene. At the same time, it is evident that most of the *lacZ* transcripts isolated from the leaderless strain (KAM109) are <1.3 kb (Fig. 5, lane 3), whereas the ethidium bromide-stained sample (Fig. 5, lane 4) demonstrates the integrity of the KAM109-derived RNA. This result suggests that the majority of elongation complexes in KAM109 are terminated within the first half of the *lacZ* gene. Together, the results indicate that transcription is terminated in several places within the untranslated *lacZ* gene in *M. smegmatis*. Moreover, some of these termination events can be alleviated by introducing an *rrn* leader or *rrn nut* site into the 5' end of the transcript. The proportion of the smallest transcript, which is ≈ 0.2 kb in size, is highest in the

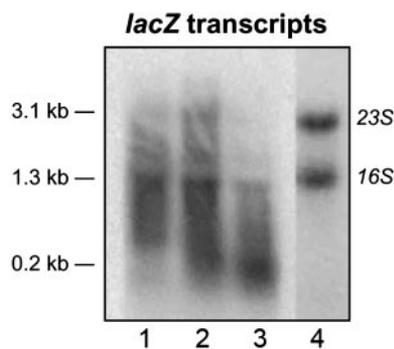


Fig. 5. Northern blot of untranslated *lacZ* expressed in *M. smegmatis*. The blot was probed with a probe complementary to nucleotides 1–125 of *lacZ*. Lane 1, KAM103 (P_{rrn} -leader_{wt}-*lacZ*); lane 2, KAM107 (P_{rrn} -*nut*_{wt}-*lacZ*); lane 3, KAM109 (P_{rrn} -*lacZ*); lane 4, KAM109, stained with ethidium bromide. Indicated transcript sizes are approximate and calculated from the positions of 16S and 23S RNA.

Table 1. *lacZ* read-through

Strain	Insert	Read-through (% of max)
KAM103	Leader	86
KAM104	Leader (U47A)	78
KAM107	<i>nut</i>	100
KAM108	<i>nut</i> (U47A)	51
KAM109	None	14

The 3'/5' ratio of *lacZ* RNA determined by slot blotting is shown. The values are expressed as percentages of the maximum value obtained with KAM107 (P_{rrn} -*nut*_{wt}-*lacZ*). Raw counts for KAM107 are given in *Materials and Methods*.

leaderless construct, and it therefore seems highly likely that this transcript arose from termination within the *tiZ1* region.

To get a more quantitative measure of the extent of *lacZ* read-through, we performed slot blots with RNA from the same strains. The membranes were probed with two radioactively labeled probes complementary to the 5' and 3' ends of *lacZ*, and the extent of hybridization was determined. The transcriptional read-through was subsequently calculated as the 3'/5' ratio, normalized to the maximal value, which was obtained with a *nut* insert (KAM107; Table 1). The results confirm the findings from the Northern blot, that KAM107 has the highest 3'/5' ratio, whereas the lowest ratio is found in KAM109, indicating a 7-fold higher frequency of transcriptional termination in the leaderless construct (Table 1). KAM103, containing a full-length leader in front of *lacZ*, displays a 3'/5' ratio that is 86% of maximum, indicating that the isolated *nut* site is better at promoting read-through than the full-length leader. We subsequently introduced the U47A mutation into the *boxC* region to determine whether this mutation would affect the extent of read-through. The results demonstrate that the U47A mutation has a dramatic effect when introduced into the minimal *nut* insert, reducing the read-through by approximately half, whereas the effect is marginal when the same mutation is introduced into the full-length leader (Table 1). These results suggest that the U47A mutation, which interferes with NusA binding, affects transcriptional events further downstream.

The NusA-*rrn* Leader Interaction Is Not Specific for *M. tuberculosis*.

The NusA protein and the *rrn nut* site display a high degree of conservation among bacteria (11, 13). To determine whether the NusA-RNA interaction is likewise conserved, we tested the *E. coli* NusA protein and the *E. coli rrnG nut* site (nucleotides +1 to +64) in our assay. The *E. coli* NusA protein, unlike the *M. tuberculosis* protein, contains a C-terminal domain, which masks the RNA-binding region (14). We therefore used a truncated version consisting of amino acids 1–416 of the *E. coli* protein, which still contains all three RNA-binding domains (21). The protein was expressed and purified by using similar methodology to that used for the *M. tuberculosis* proteins and used in EMSAs. Fig. 6 illustrates that the *E. coli* NusA416 forms complexes with the *rrnG* leader region as well as with the *M. tuberculosis nut* RNA. The *E. coli* NusA416 protein also binds to the *M. tuberculosis* antisense probe, whereas no complexes were observed with a nonrelated 60-mer RNA species (data not shown). Together, the results demonstrate that the truncated *E. coli* NusA protein binds various *nut*-like RNA species with high affinity, whereas the specificity is significantly lower than that of the *M. tuberculosis* NusA.

Discussion

In this study we have demonstrated a high-affinity interaction between the NusA protein and the *rrn nut* sites of *M. tuberculosis* and *E. coli* and thus established an additional function of this versatile transcription factor. Moreover, we have shown that the

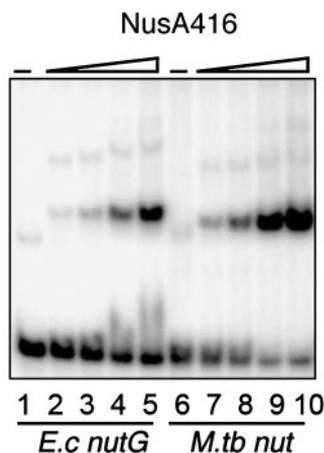


Fig. 6. EMSA with *E. coli* NusA416 and the *E. coli* (*E.c*) *rrnG* leader *nut* RNA (nucleotides +1 to +64; lanes 1–5) and *M. tuberculosis* (*M.tb*) *nut* RNA (+1 to +63; lanes 6–10). Lane 1, *E. coli rrnG*; lanes 2–5, *rrnG nut* RNA with 5, 10, 25, and 50 nM NusA416; lane 6, *M. tuberculosis nut* RNA; lanes 7–9, *M. tuberculosis nut* RNA with 5, 10, 25, and 50 nM NusA416. All lanes contain 1 nM RNA and 1 M KCl.

M. tuberculosis nut site promotes transcriptional read-through, i.e., antitermination, and that a Rho-dependent terminator within the *E. coli lacZ* gene seems to be recognized by the *M. smegmatis* Rho protein. We have shown that the otherwise elusive *boxC* region is essential for the NusA–RNA interaction and subsequently identified a single nucleotide (U47) within this region that is crucial for the interaction. Mutating U47 to A reduces NusA binding and transcriptional read-through downstream of the *nut* site. Because the mutated residue lies well outside the *boxA* region, which is known to bind NusB in *E. coli*, we believe that the reduced read-through is caused by the missing NusA–RNA interaction and that this interaction therefore plays a part in the *rrn* antitermination mechanism. The fact that the effect of the U47A mutation seems context-dependent, i.e., has a larger effect in an isolated *nut* site than in a full-length leader,

could have a number of explanations. Because the RNA-binding domains of NusA are of the S1 and KH type, we believe that the protein binds to single-stranded regions. It is possible that within an isolated *nut* site, the U47A mutation prevents NusA binding by stabilizing secondary RNA structures, such as that shown in Fig. 3. The mutated *nut* site in the context of the larger leader region may allow for alternative structures, some of which could expose the region(s) involved in the NusA interaction. Alternatively, it is possible that there are residual low-affinity interactions between NusA and the mutated *nut* site, which can be stabilized in the context of the entire leader region but not by the isolated *nut* site. We have demonstrated that the NusA–RNA interaction extends to the C-terminally truncated *E. coli* NusA416 protein. However, the *E. coli* protein seems more promiscuous than *M. tuberculosis* NusA in our assay, which could mean that, *in vivo*, the *E. coli* NusA protein is a more versatile RNA-binding protein than the *M. tuberculosis* protein and that its function relies on its ability to bind with high affinity to a variety of RNA species. In line with this explanation, it is possible that certain, as-yet-unidentified features are shared between the probes that lead to complex formation with the *E. coli* NusA416. Alternatively, the lack of selectivity may be a consequence of the C-terminal truncation, which could modify the specificity of the protein either directly or indirectly by removing the dependence of accessory factors known to interact with the C-terminal domain, e.g., RNAP α or λ N (14, 21).

The results presented here, together with the phylogenetic distance between *E. coli* and *M. tuberculosis*, indicate that the interaction between NusA and *nut* RNA is likely to be universal among bacteria and to play a significant role in bacterial *rrn* antitermination. Interfering with this interaction, therefore, may be a means of obstructing rRNA synthesis and hence presents a potential target for therapeutic agents.

We thank Cathy Squires, Finn Werner, Guy Dodson, Ian Taylor, Ian Brierley, and Andres Ramos for helpful comments. We thank Elaine Davis for pEJ414, Jack Greenblatt's laboratory for the *E. coli* NusA416 clone, and Sandra Searles for pUC18T7*Pst* δ V. This work was supported by the Medical Research Council and the European Community for Research, Technological Development, and Demonstration Activities Fifth Framework Program (Contract EU-Cluster QLK2-2000-01761).

- Richardson, J. P. & Greenblatt, J. (1996) in *E. coli and Salmonella: Cellular and Molecular Biology*, ed. Neidhardt, F. C. (Am. Soc. Microbiol., Washington, DC), Vol. 1, pp. 822–848.
- Vogel, U. & Jensen, K. F. (1995) *J. Biol. Chem.* **270**, 18335–18340.
- Vogel, U. & Jensen, K. F. (1997) *J. Biol. Chem.* **272**, 12265–12271.
- Condon, C., Squires, C. & Squires, C. L. (1995) *Microbiol. Rev.* **59**, 623–645.
- Li, J., Horwitz, R., McCracken, S. & Greenblatt, J. (1992) *J. Biol. Chem.* **267**, 6012–6019.
- Nodwell, J. R. & Greenblatt, J. (1993) *Cell* **72**, 261–268.
- Squires, C. L., Greenblatt, J., Li, J. & Condon, C. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 970–974.
- Torres, M., Condon, C., Balada, J.-M., Squires, C. & Squires, C. L. (2001) *EMBO J.* **20**, 3811–3820.
- Torres, M., Balada, J. M., Zellars, M., Squires, C. & Squires, C. L. (2004) *J. Bacteriol.* **186**, 1304–1310.
- Zellars, M. & Squires, C. L. (1999) *Mol. Microbiol.* **32**, 1296–1304.
- Berg, K. L., Squires, C. & Squires, C. L. (1989) *J. Mol. Biol.* **209**, 345–358.
- Mason, S. W. & Greenblatt, J. (1991) *Genes Dev.* **5**, 1504–1512.
- Gopal, B., Haire, L. F., Gamblin, S. J., Dodson, E. J., Lane, A. N., Papavinasundaram, K. G., Colston, M. J. & Dodson, G. (2001) *J. Mol. Biol.* **314**, 1087–1095.
- Mah, T. F., Li, J., Davidson, A. R. & Greenblatt, J. (1999) *Mol. Microbiol.* **34**, 523–537.
- Words, M., Bourenkov, G. P., Bartunik, H. D., Huber, R. & Wahl, M. C. (2001) *Mol. Cell* **7**, 1177–1189.
- Grishin, N. V. (2001) *Nucleic Acids Res.* **29**, 638–643.
- Braddock, D. T., Louis, J. M., Baber, J. L., Levens, D. & Clore, G. M. (2002) *Nature* **415**, 1051–1056.
- Braddock, D. T., Baber, J. L., Levens, D. & Clore, G. M. (2002) *EMBO J.* **21**, 3476–3485.
- Liu, Z., Luyten, I., Bottomley, M. J., Messias, A. C., Houngrinou-Molango, S., Sprangers, R., Zanier, K., Kramer, A. & Sattler, M. (2001) *Science* **294**, 1098–1102.
- Lewis, H. A., Musunuru, K., Jensen, K. B., Edo, C., Chen, H., Darnell, R. B. & Burley, S. K. (2000) *Cell* **100**, 323–332.
- Mah, T. F., Kuznedelov, K., Mushegian, A., Severinov, K. & Greenblatt, J. (2000) *Genes Dev.* **14**, 2664–2675.
- Liu, K. & Hanna, M. M. (1995) *J. Mol. Biol.* **247**, 547–558.
- Kempell, K. E., Ji, Y. E., Estrada, I. C., Colston, M. J. & Cox, R. A. (1992) *J. Gen. Microbiol.* **138**, 1717–1727.
- Snapper, S. B., Melton, R. E., Mustafa, S., Kieser, T. & Jacobs, W. R., Jr. (1990) *Mol. Microbiol.* **4**, 1911–1919.
- Price, S. R., Ito, N., Oubridge, C., Avis, J. M. & Nagai, K. (1995) *J. Mol. Biol.* **249**, 398–408.
- Jovine, L., Hainz, T., Oubridge, C. & Nagai, K. (2000) *Acta Crystallogr. D* **56**, 1033–1037.
- Papavinasundaram, K. G., Anderson, C., Brooks, P. C., Thomas, N. A., Movahedzadeh, F., Jenner, P. J., Colston, M. J. & Davis, E. O. (2001) *Microbiology* **147**, 3271–3279.
- Zuker, M. (1989) *Science* **244**, 48–52.
- Hermann, T. & Westhof, E. (1999) *Chem. Biol.* **6**, R335–R343.
- Nowatzke, W. L. & Richardson, J. P. (1996) *J. Biol. Chem.* **271**, 742–747.