doi:10.1016/j.jmb.2005.03.042 *J. Mol. Biol.* (2005) **348**, 1039–1057

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

A Quantitative Description of the Binding States and *In Vitro* Function of Antitermination Protein N of Bacteriophage λ

Clarke R. Conant, Marc R. Van Gilst, Stephen E. Weitzel William A. Rees and Peter H. von Hippel*

Institute of Molecular Biology and Department of Chemistry University of Oregon, Eugene OR 97403, USA

The N protein of bacteriophage λ activates transcription of genes that lie downstream of termination sequences by suppressing transcription termination. N binds to specific (boxB) and non-specific sites on the transcript RNA and contacts RNA polymerase via *cis*-RNA looping, resulting in "antitermination" of transcription. To find the effect of N–boxB binding on antitermination, we quantitatively relate binding measurements made in isolation to *in vitro* antitermination activity. We measure binding of N to boxB RNA, non-specific single-stranded RNA, and non-specific double-stranded DNA fluorimetrically, and use an equilibrium model to describe quantitatively the binding of N to nucleic acids of *Escherichia coli* transcription elongation complexes. We then test the model by comparison with *in vitro* N antitermination activity measured in reactions containing these same elongation complexes. We find that binding of N protein to the nucleic acid components of transcription elongation complexes can quantitatively predict antitermination activity, suggesting that antitermination *in vitro* is determined by a nucleic acid binding equilibrium with one molecule of N protein per RNA transcript being sufficient for antitermination. Elongation complexes contain numerous overlapping non-specific RNA and DNA-binding sites for N; the large number of sites compensates for the low N binding affinity, so multiple N proteins are expected to bind to elongation complexes. The occupancy/activity of these proteins is described by a binomial distribution of proteins on transcripts containing multiple non-specific sites. The contribution of specific (boxB) binding to activity also depends on this distribution. Specificity is not measured accurately by measurements made in the presence and in the absence of boxB. We find that antitermination is inhibited by non-productive binding of N to non-specific sites on template DNA, and that NusA protein covers RNA sites on the transcript, limiting N access and activity. The activity and specificity of regulatory proteins that loop from high-affinity binding sites are likely modulated by multiple non-specific binding events; *in vivo* activity may also be regulated by the modulation of non-specific binding.

Keywords: antitermination; transcription regulation; enhancer binding; non-specific binding; binding specificity

**Corresponding author*

Present addresses: M. R. Van Gilst, Department of Cellular and Molecular Pharmacology, Box 0450, University of California at San Francisco, San Francisco, CA 94143-0450, USA; W. A. Rees, Department of Immunology, Corixa, 1124 Columbia St, Seattle, WA 98104, USA.

Abbreviations used: RNAP, RNA polymerase; ssRNA, single-stranded RNA; dsDNA, double-stranded DNA.

E-mail address of the corresponding author: petevh@molbio.uoregon.edu

Introduction

Bacteriophage λ N-dependent antitermination is an important simple model system for *cis*-regulatory interactions in transcription, both because genetic studies have clearly defined the regulatory pathways involved and because purified components can be obtained easily, permitting quantitative measurements *in vitro* of the interactions of the components of these regulatory pathways. To date, however, a quantitative and testable model connecting the binding of complex components to transcriptional activity has not been proposed. Such a model could provide insight into questions about the central role of N-RNA binding in λ development. Connecting binding and activity can also further the role of N antitermination as a general model for transcription by giving a detailed picture of regulator–transcription complex interactions that underlie specificity mechanisms.

The bacteriophage λ N protein activates expression of the delayed early genes by inducing read-through of transcription terminators that, in the absence of N, would prevent delayed early gene synthesis.[1–4] This process, known as antitermination, affects only terminators located in the early lambda operons. The terminator specificity of λN protein depends on the *nut* (N-utilization) DNA sequences,[5–12] which are located close to the early promoters. Transcription of *nut* produces the RNA regulatory elements boxA and boxB, which function as protein binding sites within the nascent transcript.[13,14] The N protein binds to a stem-loop structure formed by the boxB element of *nut*,[14–16] and acts *via* an RNA looping mechanism to contact elongating RNA polymerase (RNAP) at downstream terminators (C.R.C. *et al.*, unpublished results).[10,17] N-mediated antitermination *in vivo* also requires *Escherichia coli* host factors.[18–22] These proteins do not possess antitermination activity in the absence of N, but they do form stable complexes with N and RNAP,[10,23,24] and increase the activity of N at terminators far downstream from *nut*.[25,26] The NusA protein binds both N and RNAP.[27,28]

The boxA and boxB RNA sequences, through their ability to bind N protein and change the expression of downstream terminators, are an essential component of the genetic switch from lysogeny to lytic modes of development in the lambdoid phages. Binding of N to the RNA hairpin encoded by boxB is required for λ development, and is sufficient for antitermination activity when paired with promoters and terminators from other sources *in vivo*. The interaction of N with boxB is also phage-specific, as N proteins of related lambdoid phages, which are homologous, recognize only their cognate boxB sequences.[29,30] Recognition of boxB by N thus ensures the correct location, timing, and phage specificity of lambda development.

A generally accepted model postulates that the interaction of N with RNAP elongation complexes is sufficient to cause antitermination. However, the binding of N to RNAP is too weak to occur under physiological conditions. As a consequence, N protein must bind the nascent transcript, which elevates the local concentration of N in the vicinity of RNAP to levels that facilitate a productive N-RNAP interaction. Regulation occurs in three ways: *via* increased binding of N to the RNA transcript, more efficient RNA looping, or increased activity of N once fully bound to RNAP at terminators.

Evidence suggests that boxB is more than a simple high-affinity binding platform for N. Mutations in the boxB hairpin loop can be found that bind N with only two- to fivefold lower affinity than wild-type (WT), but lack antitermination activity *in vivo* and *in vitro*.[14] Similarly, the affinity of mutant N peptides for 2-aminopurine-substituted boxB sequences measured in isolation correlates poorly with *in vivo* antitermination function.[28] The antitermination defect of these mutants may indicate a requirement for NusA protein. If so, then both boxB and NusA must be present for NusA to increase antitermination, because NusA does not increase antitermination in the absence of boxB.[31,32] NusA does not act by increasing the binding of N to boxB either, as binding of N to boxB remains unchanged in the presence of NusA.[27,28] Rather, this model proposes that boxB forms a complex with NusA and N that is recognized specifically by RNAP; recognition increases the intrinsic ability of N protein, once bound to RNAP, to cause antitermination.

Other evidence suggests that boxB is not required for N function, and that anything that allows binding of N protein to transcript RNA can cause antitermination. Thus N alone can antiterminate in the absence of boxB and NusA *in vitro*, *via* nonspecific binding of N to the RNA transcript.[32] *In vivo*, the RNA-binding domain of N protein can be exchanged with the RNA-binding domains of related phages and still support antitermination, so long as the correct cognate phage boxB binding partner is present in the RNA transcript.[29] *In vivo*, antitermination appears to be sensitive to cellular N protein levels as well, as the phage specificity of antitermination is overcome by overexpression of N protein.[33] It appears that antitermination does not require a specific recognition of boxB by RNAP, and that the role of boxB is to increase binding of N to RNA transcripts.

Thus, boxB appears to both increase the delivery of N protein to RNA polymerase, and the intrinsic activity of N protein once bound to RNA polymerase. To find the relative importance these mechanisms, we have estimated how much N protein is bound to RNA transcripts, and how much activity results; the unit activity of N. We obtain this parameter in the presence and in the absence of boxB and, by difference, we measure the size of the boxB effect on binding and activity.

To do this, we relate the nucleic acid binding affinity of N for nucleic acids to *in vitro* transcription activity, resulting in a detailed and quantitative

picture of how N protein interacts with those elements common to all transcription complexes, RNA, DNA, and RNAP. In short, we measure binding of N to specific and non-specific single-stranded RNA (ssRNA) and double-stranded DNA (dsDNA) sites in isolation, estimate the number of specific and non-specific sites contained in elongation complexes used in activity assays, and calculate the binding of N to specific and non-specific sites in the RNA transcript and DNA template. We then directly compare the fraction of RNA transcripts that are bound by N protein with the fraction of RNA transcripts that read through termination sequences in activity assays.

This approach allows us to estimate separately the effect of each type of nucleic acid binding on antitermination activity, which would not be possible in direct measurements of N-elongation complex binding. We find that binding of N to non-specific sites on the template DNA results in the formation of inactive N–DNA complexes that inhibit antitermination by competing for free N protein with binding sites on the RNA transcript. We find that multiple N proteins bind to non-specific binding sites on RNA transcripts, and that the degree to which the boxB sequence increases both binding and antitermination depends crucially on the distribution of multiple complexes of N and non-specific RNA on nascent transcripts. We demonstrate that the standard measurement of binding specificity, in which activity is measured in separate reactions in the presence and in the absence of an enhancer, underestimates the effect of enhancer on binding and activity. We propose a quantitative definition of binding specificity that estimates accurately the increase in N binding and antitermination that results from the presence of boxB.

The non-specific binding-related activity that plays such a great role in the behavior of antitermination *in vitro* is barely observable *in vivo*, raising the question of how non-specific interactions are suppressed, and how specific reactions are enhanced *in vivo*. We present the results of experiments that suggest that the NusA protein covers non-specific binding sites on the RNA transcript and increases the ratio of specific to non-specific binding sites. We discuss this and other potential *in vivo* specificity mechanisms in the context of the equilibrium binding model presented here.

Results

Methodology

We relate fluorimetric measurements of N-nucleic acid binding to antitermination activity measured in transcription runoff assays. We base our relationship of binding and activity on the initial hypothesis that the binding of a single N molecule to a transcription elongation complex will cause maximal antitermination activity. We define maximal antitermination activity empirically, by measuring the change in the fraction of RNA transcripts that read through terminators due to addition of saturating amounts of N protein. As a first test of this hypothesis, we measure the activity that results from N binding to the template DNA that is contained within the transcription complexes, and find that our initial assumption must be modified, since N bound to the DNA template is found to be inactive in antitermination (see below). To define the various classes of N binding sites on the transcription complex we then measure the binding of N protein to non-specific RNA and DNA sites, and estimate the concentration of non-specific RNA and DNA-binding sites in the elongation complexes used to measure N activity. We calculate the concentration of N protein that is bound to transcript RNA and determine the fraction of RNA transcripts bound by N protein. We then predict antitermination activity by multiplying the fraction of transcripts bound by N and the fraction of terminator read-through that corresponds to maximal N activity. This allows direct comparison of the fraction of RNA transcripts bound by N with the fraction of elongation complexes that read through transcription terminators in our antitermination assays. This comparison is then used to examine the modified hypothesis that one N-RNA binding event produces maximal antitermination activity.

Determination of the maximum activity of N protein

We measure N activity by using transcription runoff assays to measure the fraction of elongation complexes that read through transcription terminators in the presence and in the absence of N protein. We then compare these values (which vary from ~15% in the absence of N to ~90% in the presence of N) with the binding of N protein to RNA transcripts (a fractional value, ranging from 0 to 1). To equate activity and binding, we measured the maximum change in activity due to addition of saturating amounts of N protein, and we set this value to unity. Thus, we propose that if the addition of saturating amounts of N protein causes the fraction of terminator read-through to increase from 15% to 90%, then complete binding of N protein to the RNA transcripts of elongation complexes will result in a 75% change in antitermination activity.

We have used these experiments to test the proposal that the boxB RNA sequence and NusA protein increase the intrinsic ability of N protein that is bound to RNAP to cause antitermination.[14,28] At a saturating concentration of N, we expect all elongation complexes to be fully bound; thus, we expect antitermination levels to reflect the intrinsic activity of N protein that is already bound to RNAP, and not binding of N to the RNA transcript. In addition, in these and all other experiments, we used transcription templates that express short RNA transcripts. As these transcripts can exhibit

nearly 100% antitermination,[26,32] any inhibitory effect of RNA looping on antitermination activity must be negligible.

N-dependent antitermination activity in the presence and absence of boxB and NusA was measured using three DNA templates that encode the λ pL promoter located upstream of the transcription terminator, tR′. Template pRB2, which encodes the lambda *nutL* boxB sequence, was titrated in the presence of NusA. Template pRB35, which encodes a deletion of boxB from pRB2, and template pRB2G1A, which encodes an N binding-deficient point mutation (*nutL44*) in the boxB sequence of pRB2, were transcribed in the absence of NusA (Figure 1(b)). Figure 1(a) shows the fraction of terminator read-through resulting from titration of the above three templates with N protein. Also shown in Figure 1(a) is a curve representing the predicted fraction of pRB35 RNA transcripts bound by N (see below). In all cases, a plateau of activity is reached at a high concentration of N, which we interpret to represent the maximum activity of N protein under saturating conditions. This activity is close to 100% for reactions with NusA and transcript-encoded boxB RNA, and 90% for reactions lacking NusA and boxB. In reactions with the boxB-encoding template, but no NusA, the maximum antitermination activity also achieves a plateau at 85–90% (data not shown). Thus most, but not all, of the activity of N protein is independent of NusA and boxB.

Figure 1(a) shows that the concentration of N required to reach a plateau of maximum activity is approximately 1 μM. Because this activity varies to a small extent with each transcription template used, we define maximum N activity independently for each transcription template as the change in antitermination activity due to addition of 1 μM N protein. This value is set (normalized) to unity to permit subsequent comparisons N of binding and antitermination activity.

Salt concentration-dependence of the binding of N to boxB RNA, to non-specific RNA, and to non-specific dsDNA

Previously, we used fluorescence spectroscopy to measure the binding of N to RNA by monitoring the change in the intrinsic tryptophan fluorescence of N as a function of RNA concentration.[16] Here, we employ the same methodology to determine the salt concentration-dependence of the specific and non-specific binding of N to boxB and non-specific RNA. In addition, we measure the salt dependence of N binding to non-specific dsDNA by monitoring the change in anisotropy of fluorescent-labeled dsDNA as a function of N concentration. Figure 2 displays a plot of the apparent binding constants of N to boxB, to a non-specific 32 nt ssRNA oligomer, and to a 12 bp dsDNA oligomer, as a function of monovalent cation concentration. The data are plotted as log K_a *versus* −log [monovalent salt] and analyzed using a

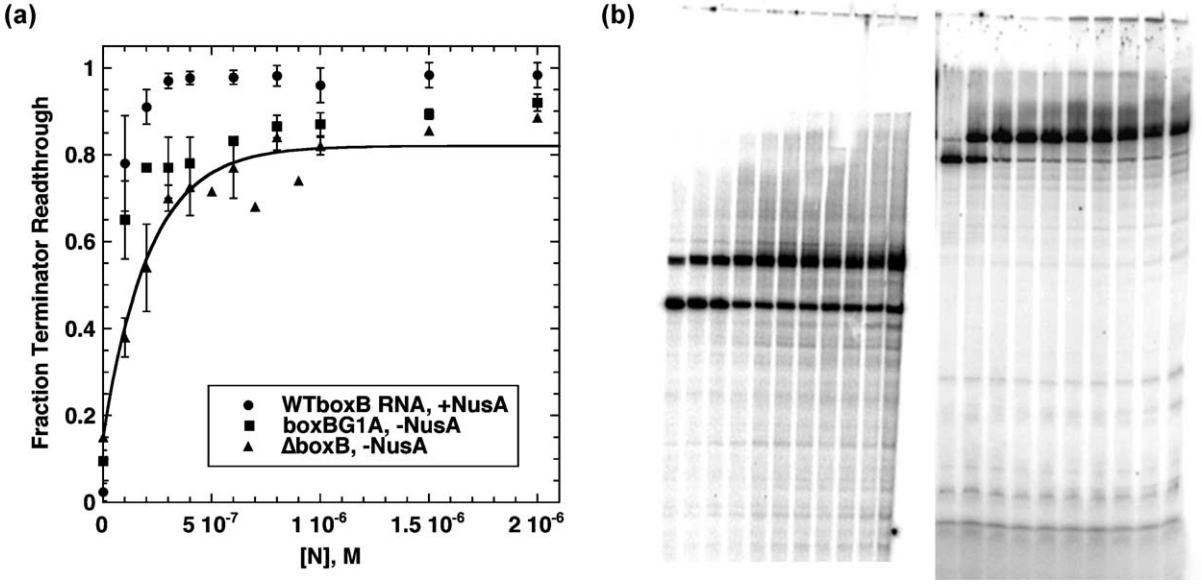

Figure 1. Measurement of N-protein-dependent terminator read through. (a) Fractional read through of intrinsic terminator tR′ as a function of added N protein in the presence and in the absence of NusA protein and boxB RNA. Circles, antitermination activity measured using template pRB2, encoding *nutL* (WT boxB RNA) in the presence of 120 nM NusA protein; squares, template pRB2G1A, encoding N binding-deficient boxB mutation *nutL44* (boxBG1A), no NusA; triangles, template pRB35, encoding a deletion of *nutL* (ΔboxB), no NusA. The curve fit represents the predicted fraction of pRB35 transcripts bound by one or more N proteins on the basis of measurements of N to non-specific RNA and DNA (see the text). Termination efficiency was determined as described in Materials and Methods. (b) *In vitro* transcription antitermination assays used to measure termination efficiency. Left, terminated (bottom) and runoff (top) RNAs produced by titration of pRB35 with N protein. Right, titration of pRB2 with N in the presence of 120 nM NusA protein.

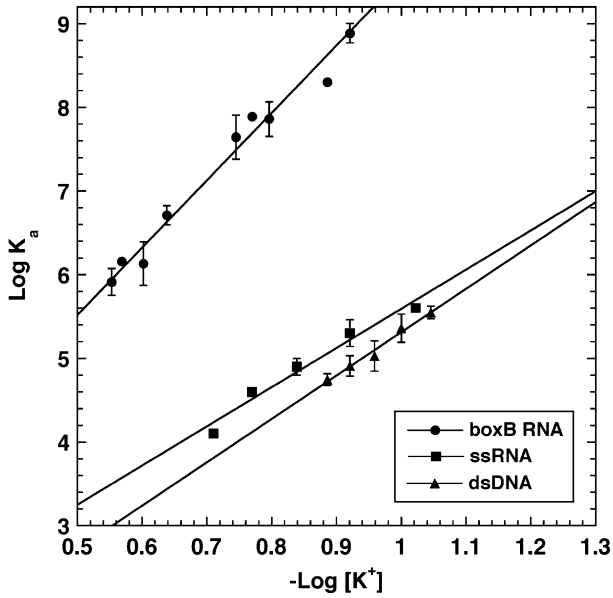

Figure 2. Log–log plot of the salt-dependence of N binding to boxB RNA, non-specific RNA and DNA. The K_a for N binding to boxB RNA (circles) was measured at monovalent cation concentrations of 120 mM, 130 mM, 160 mM, 170 mM, 180 mM, 230 mM, 250 mM, 270 mM and 280 mM. The K_a for N binding to non-specific RNA (squares) was measured at monovalent cation concentrations of 95 mM, 120 mM, 145 mM, 170 mM and 195 mM. The K_a for N binding to dsDNA (triangles) was measured at monovalent cation concentrations of 70 mM, 80 mM, 90 mM, 100 mM and 110 mM. Binding assays and data analysis are described in Materials and Methods.

linear fitting routine. The results are summarized in Table 1.

For boxB-containing RNA, the dependence of the association constant K_a on the concentration of monovalent cation ($\partial \log K_a / \partial \log[M^+]$) is 8.3 ± 0.03. For N binding to non-specific RNA and DNA, the data were best fit using a non-cooperative overlapping binding site model.[34] As Table 1 shows, for non-specific ssRNA, the salt concentration-dependence of N binding (4.7 ± 0.5, binding site size $11.5(\pm 1.6)$ nt) is somewhat lower than the salt dependence of N binding to non-specific dsDNA (5.2 ± 0.4, binding site size $17.3(\pm 1.9)$ nt or $8.7(\pm 1.0)$ bp). This outcome is consistent with the lower charge density of ssRNA relative to dsDNA.[35]

Table 1. Salt-dependence of N binding to RNA and DNA

	$\partial \log K_{obs} / \partial \log[K^+]$	$\log K$ (1 M)
boxB RNA	$8.28(\pm 0.03)$	$1.30(\pm 0.05)$
ssRNA	$4.7(\pm 0.5)$	$0.9(\pm 0.4)$
dsDNA	$5.2(\pm 0.4)$	$0.1(\pm 0.4)$

All dissociation constants were measured according to the procedures described in Materials and Methods.

We have used the linear fits of Figure 2 to calculate the affinity of N for boxB RNA, non-specific RNA and non-specific DNA at different concentrations of salt, and to calculate the concentration of N–RNA and N–DNA complexes formed upon binding of N to the nucleic acid sites of the elongation complexes used in activity assays.

Non-specific binding of N protein to DNA inhibits antitermination activity

The effect of N–DNA binding on antitermination activity is not known. We therefore measured antitermination activity in the presence of DNA templates of varying size. Three templates encoding the 248 nt pRB2 RNA transcript and 124 bp (pRB2), 324 bp (pRBD324), or 524 bp (pRB2D524) of non-specific dsDNA located upstream of the transcription start site (Figure 12) were transcribed in the presence of increasing concentrations of N protein. In the absence of N protein, extra non-coding DNA had no effect on termination. In the presence of N, antitermination was inhibited on templates that contain more non-specific DNA. The magnitude of the inhibition depends on the concentration of N. At low concentrations of N, the antitermination activity measured on transcription templates that contain 324 bp and 524 bp of "extra" non-coding DNA is inhibited by 20–30%, and by 40–50%, respectively, relative to the 124 bp template. At a high concentration of N, the activity measured with the 324 bp and 524 bp templates equals the activity measured on the 124 bp template at both 100 nM and 400 nM concentrations of N (Figure 3).

The dependence of this effect on the concentration of N suggests that inhibition is controlled by the binding equilibrium between free N protein and non-specific DNA-binding sites on the template (see Figure 2). We propose that N binding to non-specific DNA results in the formation of inactive N–DNA complexes†. In subsequent calculations of the activity that results from binding of N to elongation complexes, we assume that N protein bound to DNA is "silent" in terms of antitermination activity. We test this assumption by predicting N activity in the presence of competitor dsDNA (see below).

A model relating the binding of N to non-specific RNA and DNA, and to the boxB-independent antitermination activity of N protein

We attempt to predict the amount of N protein bound to non-specific sites in transcript RNA

† This notion that N–DNA binding inhibits antitermination by reducing the concentration of free λN is supported by experimental and theoretical evidence that shows that the short linear dsDNA molecules used in our experiments are too stiff to permit N–DNA complexes bound at non-specific sites to contact RNAP bound at termination positions by *cis*-looping, thus ruling out such looping events as contributing to N-dependent antitermination activity.[52]

produced by transcription antitermination assays. We assume that if one N molecule is bound non-specifically to the RNA transcript, then maximal antitermination will occur. For example, if the nascent transcript contains 100 RNA-binding sites, and if 1% of these sites are occupied by N, we predict that elongation complexes, on average, will contain one N protein molecule bound to the RNA transcript, resulting in 100% antitermination activity (subject to consideration of distributive processes, see below).

The total number of N binding sites in a reaction is the sum of the number of dsDNA sites and ssRNA sites. We therefore estimate the number of non-specific sites in the elongation complex as follows, using the *nut*- transcription template pWW31 as our antitermination test substrate. Transcription of pWW31 produces an RNA transcript whose length at termination positions is 175 nt. The binding site size of N on non-specific RNA is 11 nt; hence, we estimate there are 165 $(175-11+1)$[34] overlapping non-specific RNA-binding sites on the nascent transcript when RNA polymerase is located at the termination position†. Similarly, transcription template pWW31 contains 489 bp (978 nt) of DNA, and the binding site size of N on dsDNA is 17 nt (8.5 bp). As a consequence, we estimate that there are $978-17+1=962$ potential binding sites for N on the dsDNA of the pWW31 template.

The number of non-specific binding sites yields an estimate of the concentration of N binding sites present in a transcription reaction. For reactions containing 25 nM elongation complexes, the concentration of non-specific overlapping RNA-binding sites is 4.1 μM (25 nM elongation complexes ×165 non-specific RNA-binding sites), and the concentration of DNA sites is 24 μM (25 nM elongation complexes ×962 non-specific DNA-binding sites). Because the concentration of nucleic acid binding sites in antitermination reactions (∼28 μM) is 1000-fold higher than the concentration of RNA polymerase (25 nM), and because the affinity of N for non-specific RNA and DNA ($K_a \sim 10^7\ M^{-1}$ at 50 mM salt) is 50-fold higher than the estimated affinity of N for RNAP ($K_a \sim 10^5\ M^{-1}$),[36] we predict that N binds predominantly to RNA and DNA sites in antitermination reactions, and does not bind appreciably to RNAP. This allows expression of the equilibrium between N and elongation complexes as a mass balance, expressing the partition of N protein between non-specific RNA and DNA-binding sites in terms of N protein concentration:

$$[N_{tot}] = [N_F] + [N_{RNA}] + [N_{DNA}] \quad (1)$$

This expression can be written as a competitive binding equation with terms for the affinity and total concentration of RNA and DNA:

$$[N_{tot}] = [N_F] + \frac{K_{RNA}[N_F][RNA_{tot}]}{1 + K_{RNA}[N_F]} + \frac{K_{DNA}[N_F][DNA_{tot}]}{1 + K_{DNA}[N_F]} \quad (2)$$

where $[N_F]$, $[RNA_{tot}]$, and $[DNA_{tot}]$ are the concentrations of free N, total RNA and total DNA, respectively, and K_{RNA} and K_{DNA} are the association constants (in M^{-1}) for N binding to RNA and DNA sites at the levels shown in Figure 2. Equation (2) was solved numerically to determine the concentration of free N protein in antitermination reactions.

Because we find that N bound to DNA is inactive in antitermination, we use only N–RNA complexes to predict antitermination activity. Accordingly, values for free N determined using equation (2) are substituted into the N–RNA component of the mass balance to predict the concentration of N–RNA complexes, which is divided by the concentration of elongation complexes to obtain the ratio of N–RNA complexes to total transcript:

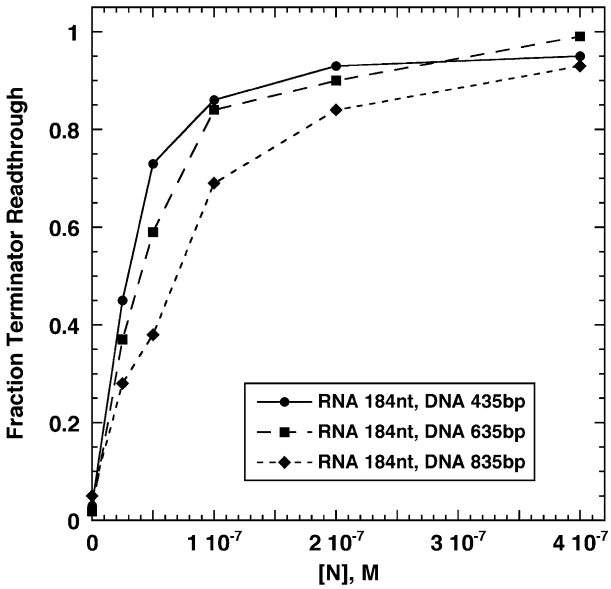

Figure 3. Non-specific DNA contained in transcription templates inhibits antitermination. Identical transcription templates were prepared with 260 bp (pRB2; circles), 584 bp (pRB2D324; squares), and 784 bp (pRB2D524; triangles) of non-coding template DNA flanking the transcription unit, and were titrated with the indicated amounts of N protein in the presence of 120 nM NusA.

† We note that, in principle, the formation of secondary structure in the RNA transcript could either increase or decrease N binding. To account for the possible presence of boxB-like secondary structures in the transcript, we have measured binding of N to the boxB sequence of bacteriophage λ21, which is structurally similar to λ boxB but does not share sequence homology. We find that N binds to this structure with the same affinity as to a non-specific RNA site (see Materials and Methods). This is supported by the observation that very minor modifications in the boxB hairpin sequence also drop the binding constant for such secondary structures to that of a non-specific RNA site.[16] As a consequence, the possible formation of boxB-like secondary structures in the transcript that have high binding affinity for N can be ignored in these calculations.

$$\frac{[N_{RNA}]}{[EC]} = \left(\frac{K_{RNA}N_F RNA_{tot}}{1 + K_{RNA}N_F}\right) \times \frac{1}{[EC]} \quad (3)$$

where [*EC*] is the concentration of elongation complexes (and transcripts) in the reaction.

Under conditions of tight binding, multiple N molecules bind to the same transcript. We assume that one N molecule bound to RNA causes maximal antitermination. It follows that the second, third, and other N molecules bound to the same RNA transcript cannot increase activity beyond this level. Thus, N molecules that share an RNA transcript are functionally redundant, and have a lower probability of causing antitermination than N molecules that are singly bound. Formation of these "extra" complexes reduces the concentration of free N protein available for binding to other RNA transcripts that are not bound by N, and so increases the fraction of elongation complexes that do not antiterminate. We account for this behavior by assuming a random binomial distribution of N-non-specific RNA binding events on RNA transcripts. Given a single site binding probability p derived from the fraction of RNA sites bound by N ([N–RNA sites]/[total RNA sites]), the probability that an RNA transcript containing k sites will have n molecules bound is given by:

$$P_p(n|k) = \frac{k}{n!(k-n)!} p^n (1-p)^{k-n} \quad (4)$$

The fraction of transcripts that contain one or more molecules of N is equal to unity minus the fraction that have no N:

$$a1 - P_p(0|k) = 1 - (1-p)^{k-0}$$

$$= 1 - (1 - ([N_{RNA}]/[RNA_{tot}]))^k \quad (5)$$

This gives the fraction of transcripts that are bound by one or more molecules of N protein, and that are therefore subject to N-mediated antitermination.

Non-specific binding of N to RNA and DNA predicts the boxB-independent antitermination activity of N

We test the assumption that a single N protein molecule bound to transcript RNA carries maximal antitermination activity by comparing the predicted fraction of RNA transcripts bound by N (equation (5)) with experimental measurements of antitermination activity. Figure 4 shows measurements of the change in terminator read-through due to addition of 400 nM N protein to reactions containing the *nut*-transcription template pWW31 at various concentrations of salt. Also shown are curves generated using the linear fits of Figure 2 and equation (5) to calculate the fraction of RNA transcripts that are bound non-specifically by one or more N protein molecules somewhere on the nascent RNA transcript as a function of salt concentration. A close correspondence of N binding and antitermination activity is observed over a wide range of salt concentration, supporting our assumption that one N protein molecule, bound to transcript RNA, is sufficient for maximal activity. These data demonstrate that the *in vitro* boxB-independent antitermination activity of N can be predicted by a simple N-nucleic acid binding equilibrium.

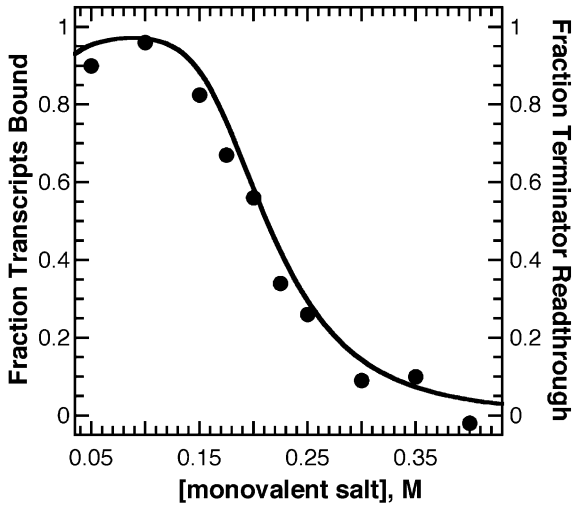

Figure 4. Comparison of predicted fraction of transcripts bound by N protein and boxB-independent antitermination activity as a function of salt concentration. The fraction of RNA transcripts bound by one or more N proteins (continuous line) as a function of salt concentration was calculated by inserting K_a values determined from the linear fit in Figure 2 into equation (5) using a total N concentration of 400 nM and the concentrations of non-specific RNA and DNA-binding sites that were present in the RNA transcript and DNA template (see Figure 12 and the text). The activity of N (black circles) was measured using transcription runoff assays to determine the change in termination efficiency due to addition of 400 nM N protein at various salt concentrations on Δ*nut* template pWW31. Antitermination levels were normalized to maximum change in termination efficiency due to addition of 1 μM N protein, where maximum terminator read-through due to the addition of 1 μM N protein at 50 mM potassium acetate is 67% and minimum terminator read-through in the absence of N (16–29%) was determined separately for each concentration of salt.

Figure 4 suggests that antitermination *in vitro* is determined by a competition between antitermination-competent RNA-binding sites and inactive DNA-binding sites for free N protein. Thus, increased N–RNA binding results in higher antitermination, and increased binding of N to DNA results in lower antitermination. Because the affinity of N for RNA and for DNA are similar, the partitioning of N between active RNA-binding sites and "silent" DNA-binding sites is extremely sensitive to small changes in the relative affinity of N for DNA and RNA.

N binds to dsDNA more weakly than to ssRNA (Figure 2), but elongation complexes contain more DNA sites than RNA sites, and a significant fraction of N protein binds the DNA template. At low

concentrations of salt, the affinity of N for DNA is nearly as strong as the affinity of N for RNA, and N protein binds predominantly to the DNA template contained in the elongation complexes. This causes a decrease in N–RNA binding, and a corresponding decrease in antitermination activity (Figure 4). At high concentrations of salt, the affinity of N for DNA is much weaker than the affinity of N for RNA, and N protein binds predominantly to RNA. This causes activity to increase at moderate concentrations of salt, while at high concentrations of salt the observed decreases in N binding and activity are dictated by the salt-dependence of N–RNA binding.

Figure 4 shows that antitermination activity reaches a maximum that is determined by the RNA–DNA competitive binding equilibrium and the distributive binding of multiple N protein molecules to a single RNA transcript. Thus, antitermination can be less than 100% even when multiple N protein molecules are bound, on average, to RNA transcripts. In Figure 4, the mass balance (equation (1)) predicts that the concentration of N–RNA complexes is 87 nM at 0.1 mM potassium acetate, where activity reaches a maximum. The concentration of RNA transcripts is 25 nM, an average of more than three N protein molecules per transcript. The probability that a single RNA site will be bound by N is (87 nM N)/(165 RNA binding sites of 25 nM transcripts) = 0.021, and the probability that the 165 binding site transcript is bound by one or more N protein molecules is $1-(1-0.021)^{165} = 0.970$.

The maximum activity of N can be estimated without detailed knowledge of the mass balance if we make the reasonable (see Figure 2) approximation that N binds to DNA and RNA with equal affinity. In this case, the probability that a transcript is bound by N is the number of N molecules per elongation complex multiplied by the fraction of total nucleic acid-binding sites that are RNA. Thus, using the previous example, the probability that an RNA transcript will contain one or more N molecules is (400 nM N)/(25 nM elongation complexes $\times$1127 total sites) = 0.014, and the fraction of RNA transcripts with one or more N proteins bound is $1-(1-0.014)^{165} = 0.902$.

We note that setting the change in antitermination activity (Figure 1) to unity can produce an artificial agreement between binding and activity at high transcript occupancies. The concentration of salt at which activity maxima occur, however, is independent of the scaling procedures used to compare binding and activity, and is predicted correctly by the model.

Effect of RNA transcript length on boxB-independent antitermination

We estimate the number of RNA-binding sites contained in RNA transcripts on the basis of the length of the RNA transcript when the elongation complex is located at the terminator. Evidence suggests, however, that the nascent transcript RNA may be structured, and/or partially enclosed by RNA polymerase in the elongation complex.[37,38] We therefore tested the accuracy of our binding site estimation by comparing predicted binding with antitermination activity measured on RNA transcripts that vary in length and RNA sequence.

Three DNA templates were prepared that vary in sequence and encode 175 nt, 98 nt and 62 nt of transcript RNA and either the tR′ (pWW31, pRB35) or tR2 (ptR2-4) terminator (Figure 12). Transcription reactions containing these templates were titrated with N protein, and activity was compared with calculations of N binding to the elongation complexes present in the antitermination assays as a function of increasing concentration of N. Figure 5 shows that shorter RNA transcripts produce less antitermination activity, in a manner consistent with binding calculations. The agreement of predicted binding and measured activity suggests that antitermination activity depends primarily on the length of the RNA transcripts, and not on their sequence or secondary structure.

We have tested our estimate of the number of RNA-binding sites in RNA transcripts by leaving the number of RNA sites as a floating parameter and determining the best fit of our binding predictions to antitermination activity data using a least-squares algorithm. In addition, we have reduced the number of DNA-binding sites available by 35 bp, which corresponds to the footprint of RNA polymerase on DNA.[37,39–41] Analysis of eight data sets (including 114 separate antitermination reactions) produces an average best fit for the number of binding sites in an RNA transcript that is 18($\pm$5) nt less than the number of binding sites predicted on the basis of the length of the RNA transcript. This result is in approximate agreement with RNA footprinting and crosslinking data, which indicate that $\sim$14 nt of transcript RNA are involved with RNA–protein contacts between the transcript and RNA polymerase.[37,38] We propose that these residues are concealed within the elongation complex and inaccessible to N, and we use this number to estimate the concentration of non-specific RNA-binding sites in subsequent calculations.

Effect of the concentration of N on the salt-dependence of antitermination

We next tested the effect of the concentration of N on the salt-dependence of binding and activity. Template pRB35 was titrated with potassium acetate at three different concentrations of N protein. Figure 6 shows that the primary effect of the concentration of N is to increase the magnitude of binding and antitermination, while preserving the salt-dependence of binding and the antitermination activity observed at high concentrations of salt. As in Figure 4, a maximum in binding and activity is observed at $\sim$0.1 M potassium acetate. The concentration of salt at which these maxima

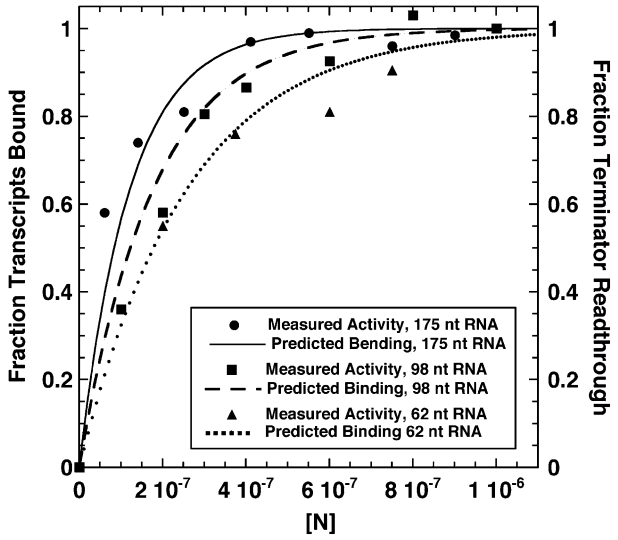

Figure 5. Effect of transcript length on predicted N binding and measured antitermination activity. Longer RNA transcripts, containing more non-specific N protein binding sites, exhibit increased antitermination activity, consistent with binding measurements. The predicted fraction of RNA transcripts bound by N is compared with the antitermination activity measured on DNA templates encoding various RNA transcript lengths. Circles, antitermination activity on template pWW31, encoding 175 nt transcript RNA. Continuous line, predicted fraction of RNA transcripts bound by N given 175 nt RNA transcript. Squares, antitermination activity on template pRB35, encoding 98 nt transcript RNA. Broken line, predicted fraction of RNA transcripts bound by N given 98 nt RNA. Triangles, antitermination activity on template ptR2-4, encoding 62 nt transcript RNA. Dotted line, predicted fraction of RNA transcripts bound by N given 62 nt RNA transcript. Predictions of the fraction of RNA transcripts bound by N were calculated using Figure 2 to calculate the K_a of N binding to RNA and DNA at the concentration of salt used in the activity assays (50 mM); these values were substituted into equation (5) and curves were generated by varying the concentration of N.

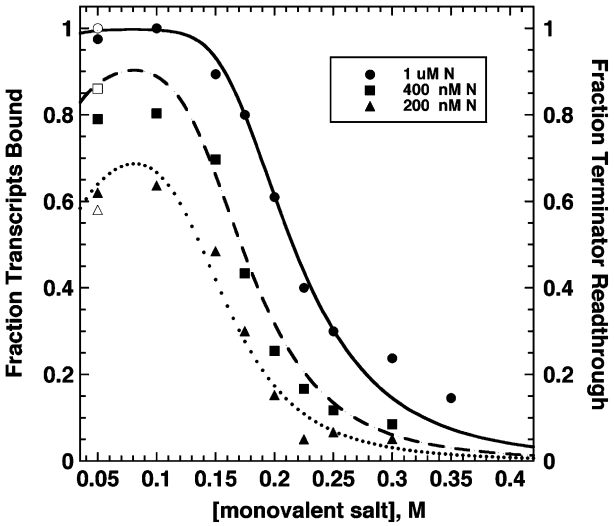

Figure 6. Effect of the concentration of N protein. Comparison of predicted and measured salt-dependence of binding and activity at three concentrations of N protein on a short RNA transcript. Antitermination activity in reactions containing the 98 nt pRB35 transcript was measured at 1000 nM (circles), 400 nM (squares), and 200 nM (triangles) N protein. Open symbols represent average activity values from Figure 1. Binding of N protein to transcripts was estimated at 1000 nM (continuous line), 400 nM (broken line), and 200 nM (dotted line) N protein with 18 nt of transcript RNA unavailable for N binding (see the text). The fraction of RNA transcripts bound by N was estimated as in Figure 4, with various concentrations of N.

occur is independent of the concentration N, as predicted by the relative salt-dependence of N-RNA and N–DNA binding (Figure 2). Thus, the primary effect of changes in the concentration of N is to increase the amplitude, but not the shape, of the N binding and N-dependent antitermination curves.

Effect of competitor dsDNA

We next tested our assumption that N–DNA complexes bound to template DNA display similar affinity (as seen through the lack of N-dependent antitermination activity) to the N–DNA complexes studied *in trans* in fluorescence experiments. To do this, we measured the activity of N protein in the presence of increasing amounts of competitor dsDNA, using template pRB35 at three concentrations of N protein. We then compared the results with the predicted fraction of transcripts bound by N in the presence of increasing concentrations of non-specific DNA sites, using the binding overlap approximation that the number of DNA sites is approximately equal to the concentration of competitor DNA, expressed in nucleotide residues.[34]

Figure 7 shows that competitor DNA inhibits antitermination in an N concentration-dependent manner. Because these reactions are performed at a low concentration of salt (50 mM), N binds predominantly to non-specific DNA sites in the elongation complex, resulting in a decrease in both predicted binding and measured activity. The agreement between predicted binding and measured activity is not exact, as the magnitude of measured activity exceeds the predicted fraction of transcripts bound. This deviation of predicted binding from activity could mean that N bound to dsDNA *in trans* still exhibits a small amount of residual antitermination activity; alternatively, the large quantities of dsDNA used in these experiments may exaggerate experimental errors in our measurements of the binding of N to dsDNA. These deviations do not alter the overall behavior of the system; we therefore conclude that the primary effect of N–DNA binding is to serve as a "sink" for free N protein, thus rendering this fraction of the N protein transcriptionally inactive.

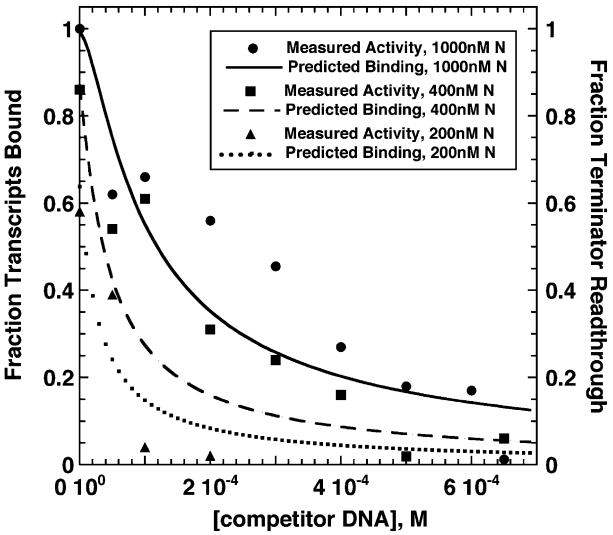

Figure 7. Effect of non-specific competitor DNA on transcript binding and boxB-independent antitermination activity. Antitermination activity on template pRB35 in the presence of increasing amounts of calf thymus DNA was measured at 1000 nM (circles), 400 nM (squares), and 200 nM (triangles) N protein. Binding of N protein to transcripts at 1000 nM (continuous line), 400 nM (broken line) and 200 nM (dotted line) N protein was estimated using equation (5), with K_a values at 50 mM salt and at the concentrations of N protein indicated, as a function of increasing concentration of DNA-binding site. The number of dsDNA-binding sites was approximated by the concentration of DNA nucleotide residues for N binding at low concentrations of protein.[34]

An equilibrium model relating the binding of N to boxB RNA, non-specific RNA, and non-specific DNA predicts the boxB-dependent antitermination activity of N protein

The relationship between binding of N protein to boxB and antitermination activity has been the subject of numerous studies.[14,16,28,33,42] However, the degree to which boxB increases transcript binding and antitermination activity has not been characterized quantitatively. N binds to the transcript-encoded boxB RNA with 1:1 stoichiometry and high affinity,[14,16] while simultaneously binding multiple non-specific RNA-binding sites. This results in antitermination from both boxB-bound and non-specifically bound N protein in the same reaction. The maximum activity of N bound to non-specific RNA and N bound to boxB are equal in the absence of NusA (see above). Thus, to find the increase in activity due to boxB binding, we calculated the fraction of transcripts that bind N protein at boxB, *versus* the fraction that binds N at non-specific RNA sites.

To do this, we repeated the procedure described above for the prediction of binding in boxB-independent systems, but with an added term in the N mass balance (equation (1)) that represents the concentration of N protein bound by boxB RNA. N–boxB and N–RNA concentrations obtained from the new mass balance were then used to calculate the ratio of N–boxB and N–non-specific RNA complexes to RNA transcripts, and equation (5) was used to calculate the distribution of multiple non-specific N–RNA binding events on RNA transcripts. Thus we set the number of transcripts that are antiterminated as equal to the fraction of transcripts bound at boxB, plus the fraction of transcripts not bound at boxB, multiplied by the fraction of transcripts bound by one or more N protein molecule at non-specific RNA sites. The fraction of RNA transcripts bound by N determined in this way was then used to estimate the fraction of the 25 nM boxB RNA present in antitermination assays that is bound by N at various concentrations of salt (Materials and Methods).

We then compared the predicted salt-dependence of N binding to specific and non-specific sites on the RNA transcript with the salt-dependence of N activity in reactions that contain a transcript-encoded boxB RNA sequence. A DNA template (pWW16; see Figure 12) encoding boxB RNA was transcribed in the presence of two different concentrations of N protein and increasing concentration of salt. Figure 8 displays the expected curves for the salt concentration-dependence of N binding, superimposed on the results of transcription antitermination activity assays at various concentrations of salt. A direct relationship between the binding of N to the boxB stem-loop and non-specific RNA-binding sites on the transcript and N-dependent antitermination activity was observed for both concentrations of N protein tested. These results demonstrate that the boxB-dependent antitermination activity of N protein can be described by a simple equilibrium-based model that describes the binding of N to RNA transcripts. In addition, the data confirm the observation (see above; and see Figure 1) that the activity of N protein as an antiterminator is not changed by its binding to boxB RNA.

Magnitude of the boxB-specific effect

Using the above procedures, we have been able to separate the contributions of specific (boxB) and non-specific RNA binding to antitermination function. This allows us to quantify the magnitude of the specific binding effect, and suggests that empirical measurements of binding specificity may produce inaccurate estimations of the contribution of N–boxB binding to activity. We illustrate this point by parsing out the antitermination activity that results from binding of N to specific (boxB) and non-specific RNA sites on the RNA transcripts of Figure 8.

Previous measurements of the effect of boxB on antitermination activity (i.e. the magnitude of the specific antitermination effect) were determined empirically by comparing measurements of antitermination activity made in separate reactions in the presence and in the absence of transcript-encoded boxB RNA. We have reproduced this

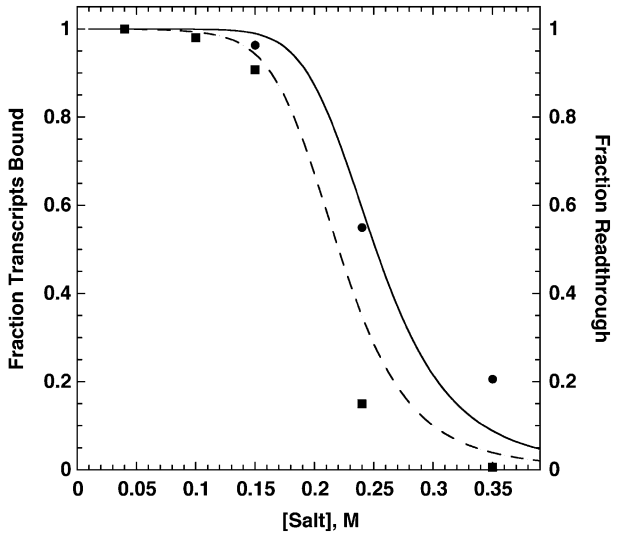

Figure 8. BoxB-dependent antitermination activity and the fraction of transcripts bound non-specifically by N as a function of salt concentration. The fraction of RNA transcripts bound by one or more N proteins as a function of salt concentration was calculated by inserting K_a values for N binding to boxB RNA, non-specific RNA and non-specific DNA as determined from the linear fit in Figure 2 into equation (5) at 350 nM (continuous line) and 150 nM (broken line) N protein. The activity of 350 nM (circles) and 150 nM (squares) N protein was measured using transcription runoff assays to determine the change in termination efficiency at various concentrations of salt using the Δ*nut* template, pWW16.

kind of measurement by calculating the fraction of transcripts that are bound by N when the boxB sequence is present and when it is absent under the conditions of the experiments shown in Figure 8. Figure 9(a) shows that the sum of the antitermination levels measured in the presence and in the absence of boxB predicts termination read-through that is greater than 100%. This overstatement of N activity is caused by the failure of such summations to account for binding of multiple N proteins to specific and non-specific sites on the same RNA transcript. Thus, while one N protein molecule is sufficient for complete antitermination, several may bind; the probability that an individual protein will interact with RNAP to produce antitermination is reduced when many protein molecules are bound. When the boxB RNA site is present, more protein molecules are bound to each transcript, and the probability that an individual N protein molecule will act is low. When boxB is absent, fewer protein molecules are bound to each transcript, and the probability that an N protein molecule will cause antitermination is high. Thus, empirical measurements of specificity activity made by deleting boxB from RNA transcripts measure activity that, in part, does not exist in the presence of boxB. Thus, when separate reactions performed in the presence and in the absence of boxB are compared, non-specific binding-related N activity in the absence of boxB is greater than non-specific binding-related activity in the presence of boxB, and the sum of antitermination activities in the presence and in the absence of boxB appears to exceed 100%, which cannot be the case.

A more accurate estimate of the contribution of specific and non-specific binding to activity is obtained when both boxB and non-specific RNA binding are quantified in a single reaction (Figure 9(b)). We assume that N–boxB complexes and N–non-specific RNA complexes have an equal chance to interact with RNAP and cause antitermination. Thus, when N protein is bound to both boxB and non-specific sites on the same RNA transcript, the probability that the N–boxB complex will cause antitermination depends on the number of other N molecules that are bound to the same RNA transcript. We calculate the number of N molecules bound to non-specific sites on a given RNA transcript by calculating a binomial distribution of non-specific RNA binding events (equation (4)). This distribution gives the fraction of RNA transcripts that contain n molecules of N protein bound to non-specific RNA. The fraction of transcripts antiterminated by N–boxB on a transcript that also contains n molecules of N bound to non-specific RNA is then equal to the fraction of complexes that contain n non-specifically bound N proteins and one boxB complex, divided by the total number of N molecules bound to the transcript ($n+1$; Materials and Methods). The contribution of specific (boxB) binding to antitermination activity is the sum of the fraction of transcripts antiterminated by N–boxB for all numbers n of N molecules multiply bound to non-specific RNA.

We have calculated the fraction of transcripts that are bound by one or more N protein molecule (Figure 9(b), continuous line), the fraction of transcripts that are antiterminated by complexes of N and boxB (Figure 9(b), broken line), and the fraction of transcripts that are antiterminated by N bound to non-specific RNA (Figure 9(b), dotted line). This method produces an estimate of the contribution of specific binding to activity that is significantly higher (∼65%; in Figure 9(b), the magnitude of the specific effect is the area underneath continuous line) than previous empirical estimates of activity measured in separate reactions using transcripts that contain or lack a boxB RNA sequence (∼30%; the observed magnitude of the specific effect is the difference between Figure 9(a), in the continuous line and the dotted line), and produces a rational sum of boxB-dependent and boxB-independent antitermination activity that is less than 100%, as expected. Thus, accurate measurement of the magnitude of specific binding-related activity must account for the distribution of multiple N binding events at both specific and non-specific sites on the same RNA transcript.

Non-specific binding may obscure the comparison of the antitermination activities of boxB mutants. Several authors have examined N binding and activity relationships by comparing the

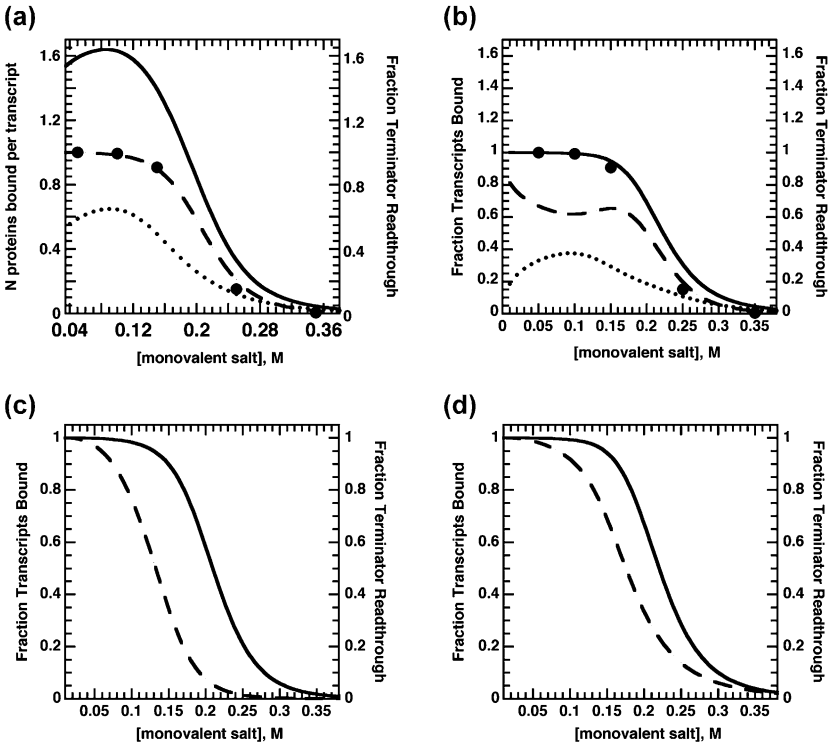

Figure 9. Contribution of N–boxB binding to antitermination activity. (a) Predicted antitermination activities resulting from the binding of N to boxB and non-specific RNA in Figure 8 (150 nM). Circles, antitermination activity from Figure 8 (150 nM N); Dotted line, predicted activity from binding of N to non-specific sites contained in the pWW16 transcript of Figure 8; broken line, predicted activity from binding of N to boxB site of pWW16 in Figure 8; continuous line, predicted activity from binding of N to both boxB and non-specific sites. Activity measured in separate reactions in the presence and in the absence of boxB gives a total fraction terminator read-through greater than 1. (b) Antitermination activity measured in a single reaction using the predictive binding model and accounting for multiple binding events on the same RNA transcript. Circles, Figure 8 activity; broken line, predicted activity from binding of N to boxB in Figure 8; dotted line, predicted activity from binding of N to non-specific RNA complexes; black line, predicted activity due to binding of N to both boxB and non-specific RNA sites. The contribution of specific (boxB) binding to activity is greater when multiple specific and non-specific binding events are considered in the context of a single reaction. (c) and (d) Non-specific binding obscures the inhibitory effect of boxB mutations on binding and activity. (c) The predicted binding and antitermination activity of N for boxB and a hypothetical boxB mutant sequence that has 100-fold lower affinity for N when the contribution of non-specific binding to activity is ignored. WT boxB (continuous line), mutant boxB (broken line). (d) The predicted binding and antitermination activity of N when the contribution of non-specific binding to activity is included. WT boxB (continuous line), mutant boxB (broken line).

behavior of boxB mutants *in vivo*[14,42–44] and *in vitro*.[14] Many of these mutant boxB sequences bind N with lower affinity than wild-type. We have depicted the binding and activity of a hypothetical mutant of boxB that possesses 100-fold weaker binding affinity for N in Figure 9(c). In the absence of non-specific binding, the difference in binding and activity between wild-type boxB and the weak-binding mutant is large (Figure 9(c)); however, when the contribution of non-specific RNA binding to antitermination is taken into account, the difference is much smaller (Figure 9(d)). Thus, analysis of binding-function relationships of N and boxB mutants must account for the activity that results from non-specific binding of N to RNA. This too requires calculation of the distribution of multiple N-transcript binding events.

Inhibition of boxB-independent antitermination by NusA

The *E. coli* host protein NusA has several effects on termination and antitermination activity. In the absence of N, NusA increases termination. In contrast, in the presence of N and boxB, NusA increases antitermination. In the simplest antitermination system, including N and transcripts deleted for boxB, NusA inhibits antitermination to an extent that depends on the concentration of N protein in the solution (Figure 11).[32] This concentration-dependence led us to ask whether an equilibrium-based model could explain the inhibition of boxB-independent antitermination activity by NusA.

The equilibrium between N and the RNA and DNA of the transcription complex is complicated by addition of NusA, which binds both N and RNAP with high affinity. In addition, while NusA does not bind RNA, complexes of NusA and N do bind to RNA. A model illustrating the possible binding states of N, NusA, transcript RNA, and RNAP, together with our prediction of whether each form will exhibit antitermination activity, is depicted in Figure 10.

The equilibrium between N, NusA, RNA, and RNAP can be approximated using assumptions that are supported by existing reports in the literature. First, complexes of N and NusA bind RNA with the same affinity as N alone;[27,28] thus, we assume that the partitioning of N among RNA and DNA sites is the same for both N and N–NusA complexes. Second, we assume that binary N–NusA complexes with RNAP are inactive for antitermination, because direct association of N with RNAP *via* NusA would be expected to increase, not inhibit,

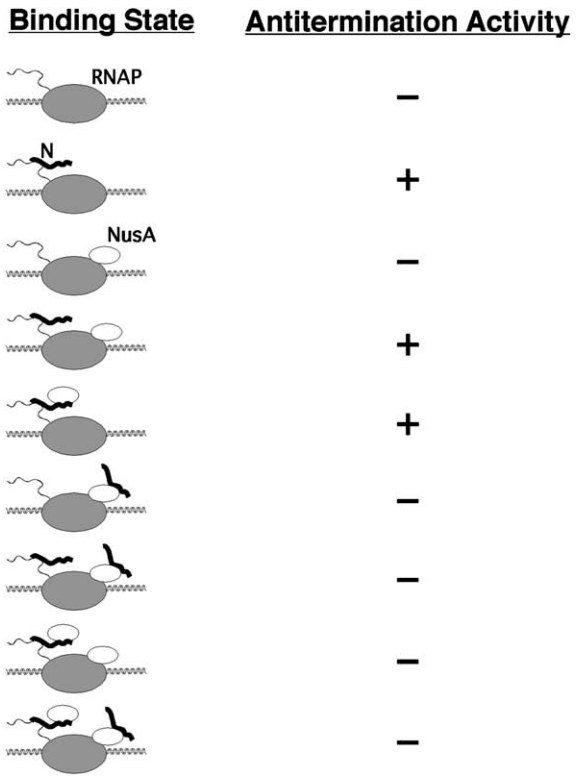

Figure 10. A model for inhibition of boxB-independent antitermination by NusA. Possible binding states of N, NusA, and RNA polymerase and antitermination outcomes predicted by the equilibrium binding model (see the text), assuming that direct binding of N–NusA complexes to RNAP does not increase antitermination.

antitermination. Third, we have used existing reports to determine the equilibrium constant for binding of NusA to RNAP.[43] Finally, we have measured the salt-dependence of N binding to NusA:

$$\log K_a = 6.74 + 0.6(-\log [K^+])$$

see Materials and Methods.

These assumptions provide the basis for the possible association states and antitermination activities shown in Figure 10, and suggest a possible mechanism by which NusA can inhibit boxB-dependent antitermination. NusA inhibits antitermination at low concentrations of N, but not at high concentrations of N. This suggests that the inhibitory effect of NusA can be overcome by excess N protein, and that inhibition must result from formation of binding states of N, NusA and RNAP that prevail at high concentrations of NusA and low concentrations of N.

One such binding state occurs when RNAP is bound by NusA and the RNA transcript is bound by a complex of N and NusA (Figure 10). This is a likely situation in our assays, as the concentration of NusA in our experiments (120 nM) is high enough to bind RNAP completely, and to bind up to ~100 nM N protein. We propose that when NusA is bound to both N and RNAP, the polymerase-bound NusA interferes with the activity of N-NusA–RNA complexes on the RNA transcript, and inhibits antitermination. This mechanism explains the inhibition of antitermination at low concentrations of N, as both N and RNAP are expected to be bound nearly completely by NusA. The mechanism also explains the lack of NusA-dependent inhibition at high concentrations of N, as excess N, unbound by NusA, is expected to bind transcript RNA at multiple sites on each RNA transcript, allowing antitermination.

We have calculated the fraction of transcripts that are bound by one or more N protein molecule on elongation complexes that are not bound also by NusA at RNAP (Materials and Methods), and compared the predictions of the model with antitermination activity. Figure 11 shows the effect of NusA on boxB-independent antitermination assays carried out with DNA template pRB35. Also shown is the predicted fraction of transcripts bound by N in the presence (Figure 11, broken line) and in the absence (Figure 11, continuous line) of NusA, allowing for the inactivity of elongation complexes that are bound by NusA at both RNAP and N–RNA complexes. The close correspondence of activity data with predicted binding in the presence and in the absence of NusA supports the hypothesis that NusA bound to RNAP renders

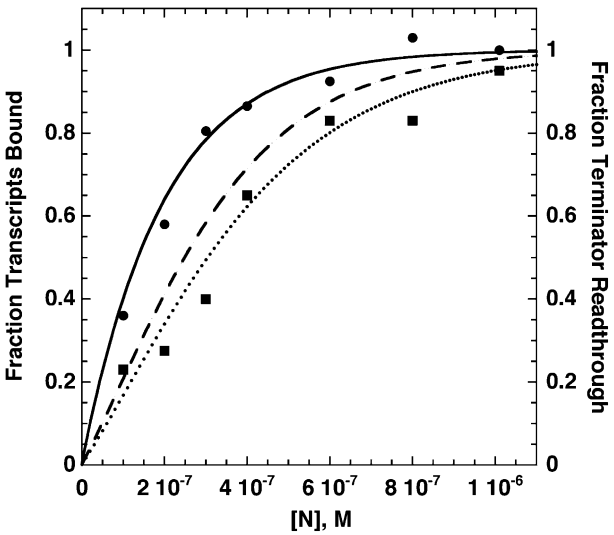

Figure 11. Inhibition of boxB-independent antitermination in the presence of NusA protein. Antitermination activity on Δ*nut* template pRB35 was measured in the presence (squares) and in the absence (circles) of 120 nM NusA protein. Continuous line, fraction of transcripts bound by N in the absence of NusA, assuming 70 nt of transcript RNA are available for binding to N. Also shown is the fraction of transcripts bound by N in the presence of 120 nM NusA, with 70 nt (broken line) and 56 nt (dotted line) of transcript RNA available for binding to N protein. Curves were generated as in Figure 5, with correction for multiple NusA binding events as described in the text.

N–NusA–RNA complexes inactive for antitermination. This model predicts that antitermination will decrease at low concentrations of N and high concentrations of NusA; such behavior has been reported.[31]

We have estimated the best fit number of non-specific binding sites in the RNA transcript available for binding to N in the presence of NusA, again by setting the concentration of RNA sites as a floating parameter and calculating the best fit of our curve to the activity data. We find that our model best fits the data when the number of binding sites on the RNA transcript is 32 nt less than the number of binding sites predicted on the basis of the length of the RNA transcript. As the number of nucleotides concealed from N in the absence of NusA is 18, this suggests that NusA prevents N from binding an additional 14 nt of transcript RNA. This result is in good agreement with RNA footprinting[45] and crosslinking[46] data, which demonstrate that RNA is contacted by NusA at RNA lengths up to 35 nt from the 3′ end of the transcript.

Discussion

The central role of the boxB hairpin in λ N-dependent antitermination is likely to be bipartite; it provides both a high-affinity binding site for N protein on the nascent transcript RNA, and it may increase the intrinsic ability of N to cause antitermination once bound to RNAP. Understanding the role of boxB is therefore likely a matter of how much, not whether, boxB contributes to binding or activity, and therefore requires quantitative and independent measurements of both binding and activity.

This, in turn, requires a model that relates equilibrium binding affinities measured in isolation to *in vitro* measurements of N antitermination. We have attempted to make such a model, beginning with three major assumptions: (i) that interactions between N and elongation complex components are at equilibrium; (ii) that binding of a single N protein molecule to RNA transcripts results in maximal antitermination; and (iii) that binding interactions are exclusive and independent. In addition, we have made several approximations that are supported by experimental data. For example, we have assumed that RNA looping of short RNA transcripts is 100% efficient, that N–DNA complexes do not produce antitermination, and that overlapping non-specific RNA-binding sites can be considered homogenous. We argue that the cumulative uncertainty implicit in these assumptions and approximations is balanced by the ability of this equilibrium model to make testable and quantitative predictions that can serve as a "straw man" for further investigations of *cis*-regulatory mechanisms in N antitermination.

We have calculated how much N protein binds to each of the nucleic acid components of elongation complexes, providing a detailed picture of the interaction of a regulatory protein with those elements common to all transcription complexes; i.e. DNA, RNA and RNA polymerase. Elongation complexes are covered with numerous overlapping non-specific nucleic acid-binding sites; so many, in fact, that even in the minimal complexes used in this study, the 1000-fold difference in binding affinity between boxB and non-specific sites for N protein is compensated by the presence of hundreds of non-specific binding sites. N therefore binds multiple non-specific sites, both on DNA and RNA, even when experimental conditions favor binding of N to boxB. Thus, while N binds a number of low-affinity sites, or a single high-affinity site, N does not bind the single low-affinity site provided by RNAP at physiological concentrations of N protein. Thus, N activity (and possibly the activity of other systems that are embedded in nucleic acids) is determined by a binding equilibrium involving all the nucleic acid components present.

An important feature of this equilibrium is that multiple proteins can bind non-specifically to RNA transcripts. When the concentration of free N protein is limiting, this effectively removes protein from other elongation complexes. Thus, even at a stoichiometry of two to three protein molecules per transcript, antitermination activity can be less than 100%. We find that the inclusion of a binomial distribution of multiple binding events on RNA transcripts containing multiple non-specific binding sites is essential to permit our model to fit the activity data. This statistical distribution of multiple binding events is likely to regulate the behavior of any system where one protein molecule suffices for activity, but several can bind at random.

The statistical distribution of multiple binding events is essential for correct calculation of specificity. Because specificity corresponds to a difference (i.e. the binding and activity in the presence of a high-affinity sequence regulatory element minus the binding and activity in its absence) a good measure of specificity must account for both specific and non-specific binding and activity. We note that simple measurement of activity in the presence and in the absence of the binding site, and in separate reactions, overestimates the contribution of non-specific binding events to activity, and underestimates the contribution of the specific sequence regulatory element to activity (Figure 9).

Another general feature of transcription illustrated by the quantitative approach to N-dependent antitermination is the competition between "active" RNA-binding sites and transcriptionally inert DNA-binding sites for free N protein. We find that increasing the length of the DNA template inhibits antitermination (Figure 3), and that dsDNA added *in trans* has qualitatively the same effect (Figure 7). We assume that complexes of N and DNA are transcriptionally inactive, and find that small changes in the relative affinity of N for RNA and DNA result in large changes in activity. This competition between RNA and DNA allows us to explain activity maxima in plots of the

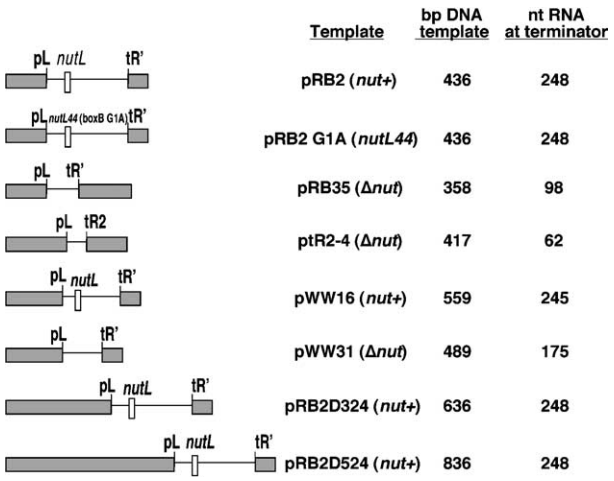

Figure 12. Diagrams of the transcription templates used in this study. DNA templates used to measure antitermination activity are shown, as are positions of the phage λ pL promoter and terminators. Lines represent transcribed regions; shaded boxes represent non-coding DNA. Amounts of non-specific DNA and RNA present in antitermination complexes are indicated.

salt-dependence of antitermination that we had previously dismissed as experimental artifacts.

Our comparisons of binding and activity raise the question: what is the "maximal activity" of N protein? Binding of N protein to RNAP does not always result in antitermination. N activity is less than 100% even when transcripts are bound, on average, by four to five N protein molecules (Figure 1). Thus, some process that follows transcript binding (e.g. RNA looping, or the antitermination event itself) must be inefficient. We have assumed that RNA looping is completely efficient on the short transcripts used in these experiments (see above); it follows that the inefficiency must lie in the intrinsic ability of N protein, once delivered to RNAP, to cause antitermination. Thus, N protein, like all enzymes, does not cause antitermination, but rather increases the probability that antitermination will occur. This probability is reflected in the terminator read-through measured *in vitro*, and is what we call maximum antitermination activity.

We find that boxB and NusA together increase the maximum antitermination activity of N protein by approximately 20%; thus, the role of boxB is both to provide a high-affinity binding platform for N protein, and to increase the intrinsic ability of N protein to cause antitermination. We have not been able to quantify the binding of N protein to RNA transcripts in the presence of boxB and NusA. However, antitermination activity at saturating concentrations of N in the presence of boxB and NusA is nearly 100%, while in the absence of either NusA or boxB, or both, we find this value to be approximately 90%. Thus, the combination of boxB and NusA increases the intrinsic activity of N protein by a small but reproducible amount. Exactly how boxB and NusA increase N activity is not known, but because both NusA and boxB binding induce specific conformations in N protein,[16,47] and because N is unstructured in their absence, it seems possible that boxB and NusA may act by placing conformational constraints on N protein.

Non-specific binding does not cause antitermination *in vivo*. Therefore, mechanisms must exist *in vivo* that suppress non-specific binding and/or activity. Our studies of NusA-dependent inhibition of antitermination in the absence of boxB suggest that one specificity-enhancing mechanism present *in vivo* is the simple blocking of N access to non-specific binding sites on the RNA transcript. We find that ~18 nt of transcript RNA are unavailable for N binding in the absence of NusA, and that ~32 nt of transcript RNA are concealed from N in the presence of NusA, consistent with previous studies of transcript accessibility in the elongation complex.[37,38,45,46] NusA thus prevents N from binding the non-specific RNA sites that are closest to the transcription complex. These sites are the most likely to provide the increases in local concentration sufficient for a productive N–RNAP interaction. However, *nutL* is 64 bp from *pL* and *nutR* is 240 bp from *pR*, so covering 32 nt of transcript RNA is not sufficient to explain the ability of N to discriminate between boxB and adjacent non-specific sites. Just how this occurs in the presence of other components of the antitermination complex, other RNA-binding proteins, and the translational apparatus is an important question for N function, and for *cis*-regulatory mechanisms, *in vivo*.

The studies presented here indicate that the magnitude of specificity depends on the strength of specific interactions, and critically on the nature of the underlying non-specific interactions present in a system. These studies may therefore prove helpful in cases where regulatory proteins have low affinity for binding sites (e.g. in the absence of co-activators, or when sequences of nucleic acid-binding sites are degenerate or contain mutations). Our studies indicate that, in the simple case of *in vitro* λ N antitermination, these effects are determined by an N-nucleic acid binding equilibrium. However, this model does not address the regulatory mechanism by which *cis*-RNA looping decreases antitermination at terminators located far from the *nut* site. Nor does an equilibrium model explain the ability of accessory factors to restore antitermination at these same terminators. We have examined these issues by determining the relationship between *nut*-to-terminator distance and antitermination activity; these results will be presented elsewhere.

Materials and Methods

Protein and oligonucleotide preparation

The procedures used to prepare and purify the N protein of phage λ,[16] the NusA protein,[43,48] and the RNA

polymerase of *E. coli*,[49] have been described. RNA oligomers (boxB sequence: AUUCCAGCCCUGAAAA AGGGC; non-specific RNA sequence: ACUGACUAACU GACUAACUGACUAACUGACUG; bacteriophage phi21 RNA hairpin control sequence: UCACCUCUAACCG GUGA) used in our binding measurements were purchased from Dharmacon, Inc. (Boulder, CO) and were treated before use by heating to 95 °C, diluting into N/BoxB buffer (defined below), and cooling slowly to room temperature. Double-stranded (ds) DNA oligomers (non-specific dsDNA 5′→3′ sequence: flourescein-GCTCGCTCAGCC) were purchased from IDT technologies, Inc. (Coralville IA), and prepared as described above by annealing to complementary strands. All oligonucleotides were analyzed (and, where necessary, purified) by PAGE. Calf thymus DNA was purchased from Sigma (St Louis, MO). Concentrations of DNA and RNA were determined spectrophotometrically using calculated molar extinction coefficients for oligonucleotides of $6.6\times10^3\ M^{-1}\ cm^{-1}$ for transcription templates and calf thymus DNA (in nt units). Concentrations of RNAP, NusA, N and N peptide were determined spectrophotometrically using calculated molar extinction coefficients.[50]

Plasmids and transcription templates

Plasmids pRB2 and pRB35,[36] pWW16 and pWW31,[17] and ptR2-4[51] were constructed as described (see Figure 12). Templates used in *in vitro* transcription reactions were prepared from parent plasmids by digestion with restriction endonucleases BstUI, ApaLI, and AseI (New England Biolabs, Beverly, MA). Templates pRB2D324 and pRB2D524, containing 200 bp and 400 bp of extra DNA upstream of the promoter in pRB2, were prepared by PCR, using primers complementary to plasmid pRB2 at positions 306 and 506 on plasmid pRB2. Template ptR2-4 was prepared by PCR, using primers positioned at positions 1450–341 of plasmid ptR2-4. BstUI and PCR fragments containing the pL-tR′ transcription unit were isolated by electrophoresis on 1% (w/v) agarose and purified using Quiagen (Valencia, CA) gel-purification spin columns. Transcription templates were used as described in the Figure legends.

Measurement of N–boxB, N–RNA, N–DNA, and N–NusA binding constants

Values of the equilibrium association constants, K_a, for the binding of N to RNA oligonucleotides were measured by monitoring the quenching of the intrinsic (tryptophan) fluorescence of the λ N protein as a function of RNA concentration, as described.[16] Fluorescence measurements using oligonucleotides were performed at 30 °C in N/BoxB buffer (20 mM KPO_4 (pH 7.6), 0.1 mM EDTA, 1 mM β-mercaptoethanol) plus various concentrations of KCl. The K_d values for N binding to boxB RNA were calculated by plotting the change in fluorescence intensity as a function of total RNA added, and then fitting the data to equation (1), as described.[16]

Binding of N to non-specific RNA and the bacteriophage phi21 boxB RNA hairpin control was measured as above, and analyzed using the McGhee–von Hippel model for ligands binding to a linear homogenous lattice,[34] as follows: the binding site size n of N protein on non-specific RNA and the maximum quenching of fluorescence signal Q_{max} was determined by titration of N with RNA at low salt (45 mM total K^+) and several concentrations of N protein. Under these conditions, n and Q_{max} were invariant with respect to the concentration of N, indicating non-cooperative binding of N to the RNA lattice ($n=11.5(\pm1.6)$ nt, $Q_{max}=0.58\pm0.04$). The standard error in the measurement of n and Q_{max} was then used to define maximum and minimum values for n and Q_{max}, and curves were fit to obtain the cooperativity parameter, ω, and the equilibrium constant, K_a. Curve fits were obtained using Profit (Quansoft, Uetikon am See, Switzerland) and a robust curve-fitting algorithm based on minimization of χ^2 values for sets of incremented parameters. Curves thus fitted had ω values over a range of 0.1–10, consistent with non-cooperative binding.

Values of the equilibrium association constant (K_a) for the binding of N to dsDNA oligonucleotides were measured by monitoring the change in fluorescence anisotropy of fluorescein-labeled dsDNA oligonucleotides as a function of added N protein. The binding site size n (17.3(±1.9) nt), and the maximum change in anisotropy r_{max} (0.12 ± 0.02), were again determined at low salt (45 mM total K^+) and various concentrations of DNA (1–5 μM total nucleotides). Data were analyzed as described above, with rearrangement of the McGhee–von Hippel binding isotherm to depict binding as a function of total lattice (DNA) concentration. Competition experiments with unlabeled dsDNA gave no change (within error) in equilibrium constant values.

Values for association constants as a function of salt concentration were plotted as $\log K_a$ *versus* $-\log$ [total K^+] using Kaleidagraph (Synergy Software, Reading, PA). Linear curve fits and errors were estimated using least-squares analysis. The salt-dependence of the N–NusA binding equilibrium was obtained as described;[27] measurements of the binding of N to NusA were performed at 150 mM, 250 mM, 350 mM, and 550 mM KCl.

In vitro transcription reactions

Transcription reactions (10 μl) containing 25 nM template dsDNA, and 25 nM *E. coli* RNAP in transcription buffer (20 mM Tris–HCl (7.6), 5 mM magnesium acetate, 0.1 mM EDTA, 5 mM DTT, 5% (v/v) glycerol, and 50 mM (unless otherwise indicated) potassium acetate) were incubated with 150 μM ApU, 10 μM ATP, CTP, and GTP, and 250 nM α-ATP to form elongation complexes stalled at the first U position on transcription templates. Complexes were chased with 1 mM NTPs and 10 μg/ml of rifampicin, and the indicated amounts of N and NusA protein and/or competitor dsDNA for ~11 ms/bp of transcript RNA at the terminator position. Reactions were quenched with 0.25% (w/v) SDS/10mM EDTA and digested with 20 μg/ml of proteinase K for 45 minutes at 45 °C, diluted with an equal volume of formamide loading buffer, heated to 94 °C for three minutes, and loaded onto a 7 M urea/7% (w/v) polyacrylamide gel. Termination efficiency (*TE*) is are defined as:

$$TE = \mathrm{RNA_{term}}/(\mathrm{RNA_{term}} + \mathrm{RNA_{runoff}})$$

where RNA_{term} and RNA_{runoff} are the amounts of radioactivity in terminated and full-length bands (after subtraction of background determined by quantification of the area between bands) as measured with a model 60 Phosphoimager and accompanying software (Molecular Dynamics, Sunnyvale, CA). To account for constitutive read-through of terminators in the absence of N (~15% on the tR′ terminator used in these studies) and incomplete maximum antitermination in the presence of N (the maximum published N effect is 95%),[32] termination efficiencies were normalized to total change in termination efficiency (ΔTE_M). The change in termination

efficiency ΔTE_M due to addition of N protein is defined as $(TE_0 - TE)\ /(TE_0 - TE_N)$, where TE_0 is the maximum termination efficiency observed in the absence of N, and TE_N is the minimum termination efficiency observed in the presence of N. TE_N was determined separately for each transcription template by titration with 1 μM N protein at 50 mM potassium acetate; values for TE_0 and TE_N are given in the Figure legends. TE_0 was determined separately for each template, and the concentration of each salt and competitor DNA (TE_0 on pWW31, containing tR′ terminator, and [potassium acetate]: 0.165 at 50 mM potassium acetate; 0.16 at 100 mM potassium acetate; 0.16 at 150 mM potassium acetate; 0.22 at 175 mM potassium acetate; 0.23 at 200 mM potassium acetate; 0.22 at 225 mM potassium acetate; .022 at 250 mM potassium acetate; 0.23 at 300 mM potassium acetate; 0.27 at 350 mM potassium acetate; and 0.29 at 400 mM potassium acetate; results for pRB35 were the same (±0.01). Where indicated by error bars, activity measurements are the average of three or more independent measurements, otherwise they represent a single measurement. Experimental artifacts necessitated the use of Chauvenet exclusion criteria with a rejection threshold of 0.5.

Prediction of N binding

Curve fits predicting the fraction of RNA transcripts bound by one or more N protein molecule in the absence of boxB and NusA were generated using equations and values given in the text. Calculation of fractions of transcripts bound by N were obtained using Kaleidagraph (Synergy Software, Reading, PA) and Levenburg-Marquart analysis for linear binding fits, and Mathematica (Wolfram Research, Inc., Champaign, IL) for fits involving determination of the N mass balance (equation (1)).

Curve fits describing the binding of N to RNA transcripts that encode a boxB RNA sequence were generated as above, with the following modifications. The mass balance for N binding to RNA and DNA sites in the elongation complex was modified to include a term for the concentration of N-boxB RNA complexes:

$$N_{tot} = N_F + N_{RNA} + N_{DNA} + N_{boxB}$$

The concentration of free N was determined numerically using equation (3), including a term for boxB binding and using measured values for the association constant of N for boxB RNA (Figure 2) and a total concentration of boxB of 25 nM. The concentration of free N protein was then used to calculate N_{RNA} and N_{boxB}. The fraction of transcripts that are bound by N at the boxB site was calculated by dividing N_{boxB} by the total concentration of elongation complexes. The fraction of transcripts that are bound by a number n of N molecules at non-specific sites in the presence of boxB was calculated using equation (4), where the probability p of a single non-specific binding event is given by the fraction of non-specific RNA sites bound by N at the concentration of RNA sites used in activity assays using equilibrium constants shown in Table 1, and the number of trials k is equal to the number of RNA sites present in the RNA transcript. The fraction of transcripts that are bound by one or more N protein molecules at boxB and non-specific sites (Figure 8) was calculated by adding the fraction of transcripts bound at boxB (NB/B_{tot}) and the fraction of transcripts that are not bound at boxB but that are bound by one or more N molecules at nonspecific sites (equation (5)):

$$\text{Transcripts bound by N} = (NB/B_{tot}) + [(1 - (NB/B_{tot}))\,(1 - P(0|k))].$$

The fraction of transcripts antiterminated by N–boxB complexes in Figure 9(b) was obtained by summing the products of the fraction of transcripts bound at boxB and the fraction of transcripts that contain n complexes of N and non-specific RNA divided by the total number of N protein molecules bound to the transcript using equation (4):

$$AT_{boxB} = (NB/B_{tot})\,[P_p(0|k) + P_p(1|k)/2 + P_p(2|k))/3 + \ldots + P_p(n|k)/(n+1)]$$

Under experimental conditions, the fraction of transcripts bound by n molecules of N protein at non-specific sites was determined to be negligible for $n > 5$; calculations were therefore summed from $n = 0$ to 8.

Curve fits describing the binding of N to RNA transcripts in the absence of a boxB site and in the presence of NusA were generated by including a term for N–NusA binding in the mass balance for N binding to RNA and DNA sites in the elongation complex:

$$N_{tot} = N_F + N_{RNA} + N_{DNA} + N_{NusA}$$

The concentration of each component of the mass balance was then determined as above, using measured affinities of N for NusA (Results) and a value for the concentration of NusA obtained by subtracting the concentration of NusA bound to RNA polymerase from the total concentration of NusA. The concentration of NusA–RNAP complexes was determined using previously reported dissociation constants[43] (30 nM at 50 mM KCl) and total NusA and RNAP concentrations:

$$F_{b,NusA} = \frac{K_d + [NusA_0] + [RNAP_0] - \sqrt{(-K_d - [NusA_0] - [RNAP_0])^2 - 4[NusA_0][RNAP_0]}}{2[NusA_0]}$$

Values for the concentration of N–NusA complexes thus calculated were used to iteratively recalculate the total concentration of NusA available for binding to NusA in the presence of increasing concentrations of total N. Under these conditions, the concentrations of NusA–RNAP complexes and N–NusA complexes were essentially invariant.

The concentration of N–NusA complexes derived from the mass balance was then partitioned among RNA and DNA sites using the assumption that binding of N to RNA and DNA is not affected by the presence of NusA. The resulting concentrations of N–RNA, N–NusA–RNA, N–DNA, and N–NusA–DNA sites were converted into fractions of RNA and DNA-binding sites bound by N, NusA, and N–NusA by dividing the concentrations of N complexes by the total concentration of RNA and DNA sites.

The probability that an N molecule (or an N–NusA complex) that is bound to an RNA site is bound also to an elongation complex that harbors RNAP bound to NusA was calculated by multiplying the fraction of RNA-binding sites bound by N and N–NusA and the fraction of RNAP bound by NusA. These probabilities were then used to calculate the binomial distribution of multiple binding events on the same transcript, as described above (equation (5)), yielding the fraction of transcripts bound

by one or more N and N–NusA complex on elongation complexes bound and unbound by NusA at RNAP.

Estimation of the number of binding sites exposed for N binding was performed using a least-squares analysis of the predicted fraction of transcripts bound by one or more N protein molecule (or, in the case of NusA experiments, the fraction of transcripts bound by one or more N and N–NusA complex on elongation complexes bound and unbound by NusA at RNAP) to experimental data with variable RNA non-specific binding site concentration. The best fit number of binding sites thus determined was related to nucleotides of transcript available for binding to N using the binding overlap assumption that for N binding sites to be available and a binding site size n, the number of exposed nucleotide residues will be $(N + n) - 1$.[34]

Acknowledgements

This work was supported, in part, by NIH Research grants GM15792 and GM29158 to P.H.vH., and a Lucille P. Markey Charitable Trust grant to the Institute of Molecular Biology at the University of Oregon. M.R.V.G., C.R.C. and W.A.R. were trainees on NIH Institutional Predoctoral Training grant GM-07759. P.H.vH. is an American Cancer Society Research Professor of Chemistry. We are grateful to our laboratory colleagues for many useful discussions of this work.

References

1. Friedman, D. & Gottesman, M. (1983). Lytic mode of lambda development. In *Lambda II* (Hendrix, R. W., ed.), pp. 21–52, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. Greenblatt, J., Nodwell, J. R. & Mason, S. W. (1993). Transcriptional antitermination. *Nature*, **364**, 401–406.
3. Das, A., Pal, M., Mena, J. G., Whalen, W., Wolska, K., Crossley, R. *et al*. (1996). Components of multiprotein–RNA complex that controls transcriptional elongation in *Escherichia coli* phage lambda. *Methods Enzymol.* **274**, 374–402.
4. Van Gilst, M. R. & von Hippel, P. H. (2000). Quantitative dissection of a transcriptional control system: the N-dependent antitermination complex of phage λ as a regulatory paradigm. *Methods Enzymol.* **323**, 1–31.
5. Salstrom, J. S. & Szybalski, W. (1978). Coliphage λ *nutL*$^-$: a unique class of mutants defective in the site of gene N product utilization for antitermination of leftward transcription. *J. Mol. Biol.* **124**, 195–221.
6. Rosenberg, M., Court, D., Shimatake, H., Brady, C. & Wulff, D. L. (1978). The relationship between function and DNA sequence in an intercistronic regulatory region in phage λ. *Nature*, **272**, 414–423.
7. Warren, F. & Das, A. (1984). Formation of termination-resistant transcription complex at phage lambda nut locus: effects of altered translation and a ribosomal mutation. *Proc. Natl Acad. Sci. USA*, **81**, 3612.
8. Olson, E. R., Tomich, C. C. & Friedman, D. I. (1984). The nusA recognition site. Alteration in its sequence or position relative to upstream translation interferes with the action of the N antitermination function of phage lambda. *J. Mol. Biol.* **180**, 1053.
9. Horwitz, R. J., Li, J. & Greenblatt, J. (1987). An elongation control particle containing the N gene transcriptional antitermination protein of bacteriophage lambda. *Cell*, **51**, 631.
10. Nodwell, J. R. & Greenblatt, J. (1991). The *nut* site of bacteriophage λ is made of RNA and is bound by transcription antitermination factors on the surface of RNA polymerase. *Genes Dev.* **5**, 2141–2151.
11. Barik, S., Ghosh, B., Whalen, W. & Lazinski, D. (1987). An antitermination protein engages the elongating transcription apparatus at a promoter-proximal recognition site. *Cell*, **550**, 885.
12. Das, A. (1992). How the phage lambda N gene product suppresses transcription termination: communication of RNA polymerase with regulatory proteins mediated by signals in nascent RNA. *J. Bacteriol.* **174**, 6711–6716.
13. Nodwell, J. R. & Greenblatt, J. (1993). Recognition of boxA antiterminator RNA by the *E. coli* antitermination factors nusB and ribosomal protein S10. *Cell*, **72**, 261–268.
14. Chattopadhay, S., Garcia-Mena, J., DeVito, J., Wolska, K. & Das, A. (1995). Bipartite function of a small RNA hairpin in transcription antitermination in bacteriophage λ. *Proc. Natl Acad. Sci. USA*, **92**, 4061–4065.
15. Tan, R. & Frankel, A. D. (1992). Structural variety of arginine-rich RNA-binding peptides. *Biochemistry*, **31**, 10288–10294.
16. Van Gilst, M. R., Rees, W. A., Das, A. & von Hippel, P. H. (1997). Complexes of N antitermination protein of phage λ with specific and non-specific RNA target sites on the nascent transcript. *Biochemistry*, **36**, 1514–1524.
17. Whalen, W. A. & Das, A. (1990). Action of an RNA site at a distance: role of the *nut* genetic signal in transcription antitermination by phage λ *N* gene product. *New Biol.* **2**, 975–991.
18. Friedman, D. I. (1971). A bacterial mutant affecting lambda development. In *The Bacteriophage Lambda* (Hershey, A. D., ed.), pp. 733–738, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
19. Keppel, F., Georgopoulos, C. P. & Eisen, H. (1974). Host interference with expression of the lambda N gene product. *Biochimie*, **56**, 1505–1509.
20. Greenblatt, J. & Li, J. (1981). The nusA gene protein of *Escherichia coli*. Its identification and a demonstration that it interacts with the gene N transcription antitermination protein of bacteriophage lambda. *J. Mol. Biol.* **147**, 11–23.
21. Friedman, D. I., Schauer, A. T., Baumann, M. R., Baron, L. S. & Adhya, S. L. (1981). Evidence that ribosomal protein S10 participates in the control of transcription termination. *Proc. Natl Acad. Sci. USA*, **78**, 1115–1118.
22. Li, J., Horwitz, R., McCracken, S. & Greenblatt, J. (1992). NusG, a new *Escherichia coli* elongation factor involved in transcriptional antitermination by the N protein of phage lambda. *J. Biol. Chem.* **267**, 6012–6019.
23. Mason, S. W. & Greenblatt, J. (1991). Assembly of transcription elongation complexes containing the N protein of phage λ and the *Escherichia coli* elongation factors NusA, NusB, NusG, and S10. *Genes Dev.* **5**, 1504–1512.
24. Mogridge, J., Mah, T. F. & Greenblatt, J. (1995). A protein–RNA interaction network facilitates the template-independent cooperative assembly on RNA

polymerase of a stable antitermination complex containing the lambda N protein. *Genes Dev.* **9**, 2831–2845.
25. Mason, S. W., Li, J. & Greenblatt, J. (1992). Host factor requirements for processive antitermination of transcription and suppression of pausing by the N protein of bacteriophage λ. *J. Biol. Chem.* **5**, 19418–19426.
26. DeVito, J. & Das, A. (1994). Control of transcription processivity in phage λ: Nus factors strengthen the termination-resistant state of RNA polymerase induced by N antiterminator. *Proc. Natl Acad. Sci. USA*, **91**, 8660–8664.
27. Van Gilst, M. R. & von Hippel, P. H. (1997). Assembly of the N-dependent antitermination complex of phage λ: NusA and RNA bind independently to different unfolded domains of the N protein. *J. Mol. Biol.* **274**, 160–173.
28. Xia, T., Frankel, A., Takahashi, T. T., Ren, J. & Roberts, R. W. (2003). Context and conformation dictate function of a transcription antitermination switch. *Nature Struct. Biol.* **10**, 812–819.
29. Lazinski, D., Grzadzielska, E. & Das, A. (1989). Sequence-specific recognition of RNA hairpins by bacteriophage antiterminators requires a consensus arginine-rich motif. *Cell*, **59**, 207–218.
30. Harada, K., Martin, S. S., Tan, R. & Frankel, A. D. (1997). Molding a peptide into an RNA site by *in vivo* peptide evolution. *Proc. Natl Acad. Sci. USA*, **94**, 11887–11892.
31. Whalen, W. A., Ghosh, B. & Das, A. (1988). NusA protein is necessary and sufficient *in vitro* for phage λ N gene product to suppress a rho-independent terminator placed downstream of *nutL*. *Proc. Natl Acad. Sci. USA*, **85**, 2494–2498.
32. Rees, W. A., Weitzel, S. E., Yager, T. D., Das, A. & von Hippel, P. H. (1996). Bacteriophage λ N protein alone can induce transcription antitermination *in vitro*. *Proc. Natl Acad. Sci. USA*, **93**, 342–346.
33. Franklin, N. C. & Doelling, J. H. (1988). Overexpression of N antitermination proteins of bacteriophages λ, 21, and 22: loss of N protein specificity. *J. Bacteriol.* **171**, 2513–2522.
34. McGhee, J. D. & von Hippel, P. H. (1974). Theoretical aspects of DNA–protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *J. Mol. Biol.* **86**, 469–489.
35. Record, M. T., Lohman, T. M. & de Haseth, P. (1976). Ion effects on ligand-nucleic acid interactions. *J. Mol. Biol.* **107**, 145–158.
36. Rees, W. A., Weitzel, S., Das, A. & von Hippel, P. H. (1997). Regulation of the elongation-termination decision at intrinsic terminators by antitermination protein N of phage λ. *J. Mol. Biol.* **273**, 797–813.
37. Komissarova, N. & Kashlev, M. (1997). Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3′ end of the RNA intact and extruded. *Proc. Natl Acad. Sci. USA*, **94**, 1755–1760.
38. Nudler, E., Mustaev, A., Lukhtanov, E. & Goldfarb, A. (1998). Spatial organization of transcription elongation complex in *E. coli*. *Science*, **281**, 424–428.
39. Pavco, P. & Steege, D. A. (1989). Elongation by *Escherichia coli* RNA Polymerase is blocked in vitro by a site-specific DNA binding protein. *J. Biol. Chem.* **265**, 9960–9969.
40. Nudler, E., Mustaev, A., Lukhtanov, E. & Goldfarb, A. (1997). The RNA:DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell*, **89**, 33–41.
41. Gu, W., Wind, M. & Reines, D. (1993). Nascent RNA cleavage by arrested RNA polymerase II does not require upstream translocation of the elongation complex. *J. Biol. Chem.* **268**, 25604–25616.
42. Su, L., Radek, J. T., Labeots, L. A., Hallenga, K., Hermanto, P., Chen, H. *et al*. (1997). An RNA enhancer in a phage transcriptional antitermination complex functions as a structural switch. *Genes Dev.* **11**, 2214–2224.
43. Gill, S. C., Weitzel, S. E. & von Hippel, P. H. (1991). *Escherichia coli* σ^{70} and NusA proteins. I. Binding interactions with core RNA polymerase in solution and within the transcription complex. *J. Mol. Biol.* **220**, 307–324.
44. Franklin, N. C. (1993). Clustered arginine residues of bacteriophage λ N protein are essential to antitermination of transcription, but their locale cannot compensate for *boxB* loop defects. *J. Mol. Biol.* **231**, 343–360.
45. Landick, R. & Yanofsky, C. (1987). Isolation and structural analysis of the *Escherichia coli trp* leader paused transcription complex. *J. Mol. Biol.* **196**, 363–377.
46. Liu, K. & Hanna, M. M. (1995). NusA contacts nascent RNA in transcription complexes. *J. Mol. Biol.* **247**, 547–558.
47. Bonin, I., Muhlberger, R., Bourenkov, G. P., Huber, R., Bacher, A., Richter, G. & Wahl, M. C. (2004). Structural basis for the interaction of *Escherichia coli* NusA with protein N of phage lambda. *Proc. Natl Acad. Sci. USA*, **101**, 13762–13767.
48. Schmidt, M. C. & Chamberlin, M. J. (1984). Amplification and isolation of *Escherichia coli nusA* protein and studies of its effects on *in vitro* RNA chain elongation. *Biochemistry*, **23**, 197–203.
49. Burgess, R. R. & Jendrisak, J. J. (1975). A procedure for the rapid, large-scale purification of *E. coli* DNA-dependent RNA polymerase involving Polyamin P precipitation and DNA-cellulose chromatography. *Biochemistry*, **14**, 4634–4638.
50. Gill, S. C. & von Hippel, P. H. (1991). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326.
51. Wilson, K. S. & von Hippel, P. H. (1994). Stability of *E. coli* transcription complexes near an intrinsic terminator. *J. Mol. Biol.* **244**, 36–51.
52. Rippe, K. (2001). Making contacts on a nucleic acid polymer. *Trends Biochem. Sci.* **26**, 733–740.

Edited by D. E. Draper

(Received 14 October 2004; received in revised form 11 March 2005; accepted 15 March 2005)