Chapter 5

Context and Conformation Dictate Function of a Transcription Antitermination Switch

This work has been adapted from the following publication:

Xia, T., Frankel, A., Takahashi, T.T., Ren, J. and Roberts, R.W. Context and conformation dictate function of a transcription antitermination switch. (2003) *Nat Struct Biol* **10**, 812-819.

Abstract

In bacteriophage λ, transcription elongation is regulated by the N protein, which binds a nascent mRNA hairpin (termed *boxB*) and enables RNA polymerase to read through distal terminators. We have examined the structure, energetics and *in vivo* function of a number of N–*boxB* complexes derived from *in vitro* protein selection. Trp18 fully stacks on the RNA loop in the wild-type structure, and can become partially or completely unstacked when the sequence context is changed three or four residues away, resulting in a recognition interface in which the best binding residues depend on the sequence context. Notably, *in vivo* antitermination activity correlates with the presence of a stacked aromatic residue at position 18, but not with N–*boxB* binding affinity. Our work demonstrates that RNA polymerase responds to subtle conformational changes in *cis*-acting regulatory complexes and that approximation of components is not sufficient to generate a fully functional transcription switch.

Introduction

The regulation of transcription elongation is important in both prokaryotic and eukaryotic gene regulation (1-5). However, much less is known about the regulation of elongation than of initiation, even in classical systems such as bacteriophage λ, in which termination (6) and elongation regulation were first described (7, 8). The N protein is a key regulator of transcription elongation in phage λ (9-11), interacting with (i) the nascent mRNA (ii) the transcription factor NusA and (iii) RNA polymerase, despite being only 107 amino acids long and totally unfolded in solution (Figure 5.1a,b) (12). N increases the processivity of RNA polymerase, allowing it to read through both intrinsic and Rho-dependent termination signals (9-11), and initiates this antitermination activity by specifically binding the nascent mRNA at the *boxB* stem loop within the sequence known as *nut* (N-utilization) (13, 14) (Figure 5.1b,c). Despite recent crystallographic work describing prokaryotic transcription machinery (15), various features of this antitermination switch remain mysterious. For example, controversy exists over whether approximation of components in the antitermination machinery is sufficient for function (16) or whether the detailed conformations of *cis*-acting elements (17) have any role.

The RNA-binding domain of the λ N(1–22) peptide folds into a bent α-helix upon binding the *boxB* RNA and enforces the formation of a GNRA tetraloop fold with one base extruded (Figure 5.1c) (17-21). In the complex, Trp18 stacks on top of the tetraloop fold, effectively extending the RNA stack by one residue (Figure 5.1c). Previous work *in vitro* and *in vivo* indicates that Trp18 is important for both RNA binding and processive antitermination (19, 22). We have previously used the RNA-binding domain of the λ

N(1–22)–*boxB* complex as a model system to isolate high-affinity RNA-binding peptides by mRNA display (23, 24). We were surprised when selections randomizing residues 13–22 of N showed no conservation of tryptophan at position 18 in the selected peptides (23). Instead, a loose consensus containing Glu14 and Arg15 emerged, which also differed from the energetically important wild-type residues Lys14 and Gln15 (19).

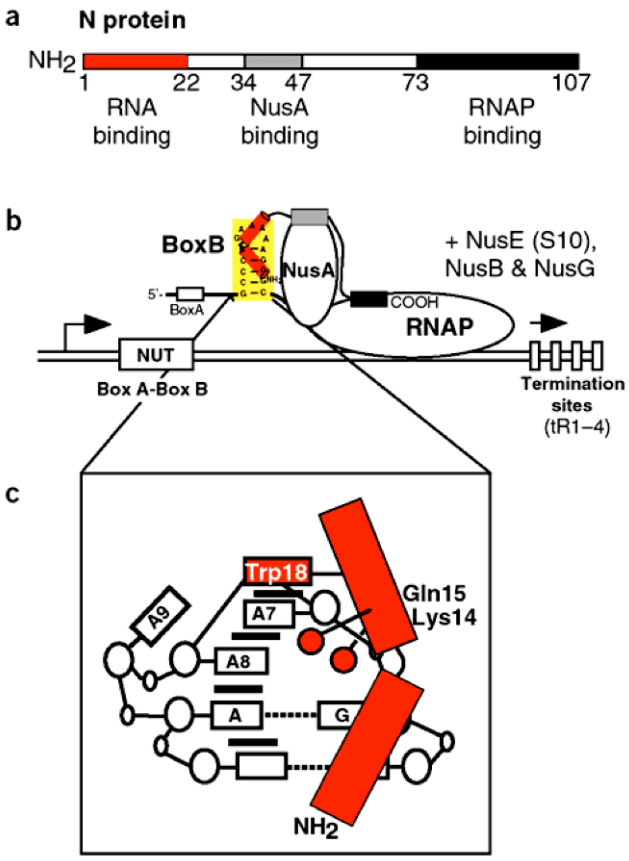

Figure 5.1. Structure and function of the N protein. **(a)** Domain structure of N. Regions for *boxB* RNA binding (residues 1–22), NusA binding (residues 34–47) and RNA polymerase binding (residues 73–107) are indicated (39). **(b)** Formation of the processive antitermination complex. The minimal antitermination complex (50) forms when N binds the *boxB* in the nascent mRNA and is stabilized by NusA binding. Association of the other Nus proteins results in a processive antitermination complex. **(c)** Schematic of N peptide–*boxB* RNA complex indicating the peptide and loop stacking geometry.

Notably, selections in which residues 7, 14 and 15 were also randomized resulted in a sequence containing the Glu14Arg15 pair in the presence of Trp18. Here we address the structural, energetic and functional consequences of these mutations.

Results

Context-Dependent Energetics

Our previous selections hinted that certain residues in the RNA-binding domain of N might be energetically coupled (23). Subsequent sequence-covariation analysis of more than 60 different RNA-binding peptides isolated through *in vitro* selection experiments supported this idea, indicating a statistically significant linkage between (i) positions 14 and 15 and (ii) positions 15 and 18, and general linkage between (iii) adjacent residues and (iv) those positioned one helical turn away (25).

We were curious to examine the structural and functional consequences of these changes and assess why these selected sequences are not found in the wild-type phage. We constructed 20 peptides with systematic variations at positions 14, 15 and 18 and measured the binding of each one to *boxB* RNA to investigate any energetic coupling experimentally. In wild-type λ N(1–22; K14Q15), the residues at position 18 with highest binding affinities were tryptophan and tyrosine, whereas arginine, alanine and glutamate variants showed similar stability (Figure 5.2). In other sequence contexts, both the best residues at position 18 and the rank order shifted substantially. For example, peptides containing Trp18, Ala18 and Arg18 all showed similar affinity when residues 14 and 15 were glutamate and arginine (that is, in the E14R15 variant), respectively. However, the peptide with Arg18 was the best binder in a designed sequence containing

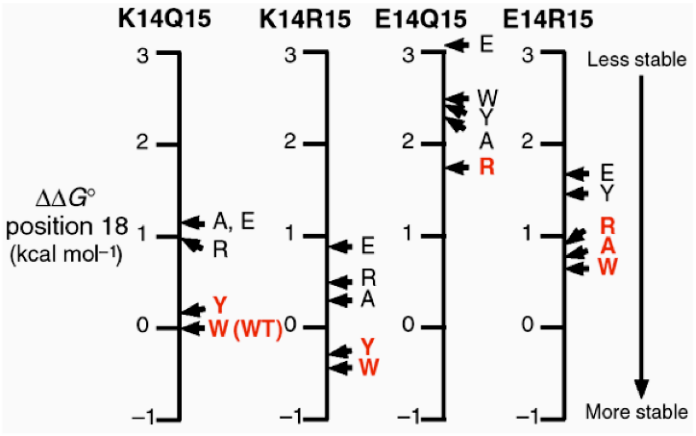

Figure 5.2. Relative stability of N peptide–*boxB* RNA complexes that differ at positions 14, 15 and 18. The dissociation constant (K_d) for the wild-type (WT) complex was determined to be 1.2 nM for 2AP-7–labeled *boxB* RNA. Uncertainties in K_d values were <10%. For complexes with small fluorescence change at 2AP-7 (such as E14Q15), 2AP-8 and 2AP-9 *boxB* RNA were used. Complexes are ranked by their free energy change relative to the wild-type N–*boxB* complex (G°; kcal mol^{-1}), with the most stable shown in red. Complexes are indicated by the peptide residue identity at positions 14 and 15 (top) and residue 18 (adjacent to arrows).

E14Q15. Thus, the energetic landscape at position 18 depends on the sequence context one helical turn away at positions 14 and 15.

Probing Sequence Preferences with mRNA Display

We next investigated the context-dependent energetics of RNA binding using *in vitro* selection experiments to rank the amino acids at position 18 in relation to different sequence contexts at positions 14 and 15. We inserted a saturation cassette (NNG/C codon) into the position corresponding to residue 18 of the protein. We then carried out mRNA display–based peptide selections (23) for five libraries with variants differing at positions 14 and 15: (i) K14Q15 (ii) L14Q15 (iii) K14H15 (iv) K14R15 and (v) E14R15. After two rounds of selection against *boxB* RNA (23, 24), 20 sequences were cloned from each library and assessed (Table 5.1). We observed good correlation between the number of times a peptide sequence occurred and its binding affinity for the *boxB* RNA,

Table 5.1. Amino acid composition at position 18 after two rounds of *in vitro* selection for binding *boxB* RNA

Library	Trp	Tyr	Phe	His	Total aromatic	Total other residues (%)
K14Q15-X18	19.2	19.2	7.7	3.8	50	50
L14Q15-X18	30.0	35.0	0	5.0	70	30
K14H15-X18	0	17.9	17.9	3.6	39.4	61.6
K14R15-X18	4.2	4.2	4.2	12.4	25	75
E14R15-X18	0	0	5.3	15.8	21.1	78.9

For each library, fractions (%) of sequences with tryptophan, tyrosine, phenylalanine and histidine at position 18 are listed separately, and their sum is listed as Total aromatic; the rest of the residues are listed as Total others. The numbers of sequences for each library are as follows: 26 for the K14Q15-X18 library, 20 for L14Q15-X18, 28 for K14H15-X18, 24 for K14R15-X18 and 19 for E14R15-X18.

confirming our approach. The selected sequences showed strong context effects for *boxB* RNA binding and peptide positions 14, 15 and 18 that were unique to each library. For the K14Q15-X18 and L14Q15-X18 libraries, we observed a strong preference for tryptophan and tyrosine at position 18. Fluorescence measurements of these variants showed quenching of 2-aminopurine (2AP)-7-substituted *boxB* similar to that of the wild-type peptide, consistent with -stacking of the aromatic side chain on the RNA loop (19, 22) (Figure 5.3b). Our selection results correlated well with previous genetic work demonstrating that the L14Q15 variant is highly functional and that tryptophan and tyrosine are the only functional substitutions at position 18 of N protein (22). In contrast, our selections showed that aromatic and planar residues rarely occur at position 18 in the K14R15 and E14R15 variants. The selections therefore provided an experimental demonstration that the fitness landscape for *boxB* binding is rugged for the interfacial residues, and were suggestive of structural differences that would explain the observed sequence preferences.

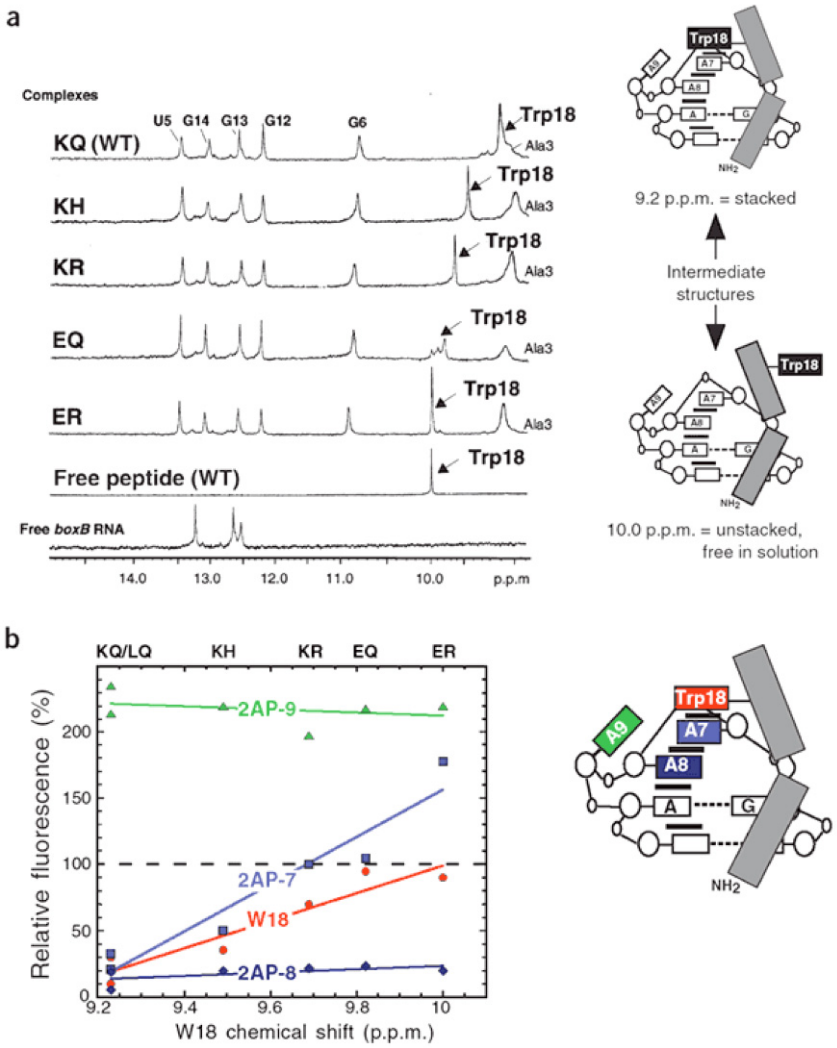

Figure 5.3. Structure and folding characteristics of N–*boxB* complexes as detected by NMR and fluorescence spectroscopy. **(a)** One-dimensional NMR spectra of peptide–RNA complexes containing Trp18 and residues K14Q14 (KQ; wild-type, WT), K14H15 (KH), K14R15 (KR), E14Q15 (EQ) or E14R15 (ER) at positions 14 and 15. The positions of the imino protons and the tryptophan indole proton are indicated. One-dimensional spectra are also shown for the free peptide (wild-type) and free *boxB* RNA. **(b)** Correlation of N–*boxB* fluorescence with Trp18 chemical shift. The relative fluorescence of Trp18 (unlabeled complex) or 2-aminopurine at RNA position 7, 8 or 9 (2AP-7, 2AP-8 and 2AP-9) is plotted versus the 1H chemical shift of Trp18 (p.p.m.). The identities of peptide residues 14 and 15 are indicated at the top of the figure. Linear fits to the data and the stacking schematic of the wild-type complex are also shown.

An Adaptive RNA-Peptide Interface

NMR provided a means to examine the structural origins of the context dependences we had observed. The indole NH proton of Trp18 and the RNA imino protons provide well-resolved spectroscopic handles in our RNA–peptide complexes. We reasoned that structural variations in the complexes should be revealed by intensity or chemical-shift changes of some of these resonances. All of the free peptides examined were unfolded in solution (23) and showed induced α-helix formation upon binding the *boxB* RNA (Figure 5.S1). The free *boxB* RNA showed three imino resonances; the rest were lost to exchange with solvent (Figure 5.3a). In the wild-type N(1–22)–*boxB* complex (K14Q15, Figure 5.3a), the RNA showed two more resonances from imino protons (five total), four corresponding to stem base pairs (U5, G14, G13 and G12) and one upfield at 11 p.p.m. for the G6–A10 sheared pair. The most notable change in the peptide spectra occurred at Trp18, which shifted upfield by 0.8 p.p.m. as a result of the ring-current effect of the adjacent, π-stacked RNA base (A7), consistent with previous structural work (17-19).

All complexes gave the same number and pattern of imino proton peaks at 10–14 p.p.m. (including the G6–A10 sheared pair imino), indicating that the mutant peptides induce similar RNA stem structure in all cases. Consistent with this idea, the N-terminal peptide (N(1–11)) present in all our mutants also induced the same RNA spectra with *boxB* (data not shown), indicating that stem recognition occurs predominantly with the N-terminal 11 residues. However, the Trp18 indole NH resonance in five of the RNA–peptide complexes showed considerable variability (Figure 5.3a). Whereas both the wild-type complex and L14Q15 complex (data not shown) showed almost complete

stacking on the loop as indicated by the large upfield Trp18 chemical shift, in the E14R15 complex, Trp18 appeared almost completely unstacked as indicated by the same chemical shift as in the free peptide. The other complexes (K14H15, K14R15, and E14Q15) showed intermediate levels of stacking, as gauged by their Trp18 chemical shifts between the wild-type and E14R15.

The NMR observations correlated well with fluorescence experiments used to determine the K_d values for RNA–peptide complex formation. In these experiments, we measured the fluorescence ratio of the complex to free peptide or free 2AP-labeled RNA. Quenching indicates increased stacking of the fluorophore, whereas enhancement indicates increased exposure to aqueous solvent (26). The stacked structure seen in the wild-type complex showed quenching of both position 7 (2AP-7) of the RNA and Trp18 of the peptide, whereas fluorescence enhancement for 2AP-7 was seen for the unstacked E14R15 variant. The relative fluorescence of 2AP-7 and Trp18 showed a linear correlation, as did 2AP-7 fluorescence and the Trp18 chemical shift we observed by NMR (Figure 5.3b). RNA positions adjacent to this interface (2AP-8 and 2AP-9) showed a consistent amount of fluorescence quenching (22–28%) or enhancement (200–235%) irrespective of the Trp18 chemical shift for different complexes (Figure 5.3b).

Taken together, the NMR, CD and fluorescence data indicate that the complexes have similarity in the RNA stem, the sheared GA pair, the flipped base (position A9) in the loop and the induced helical structure of the peptides. However, the critical Trp18–A7 interaction showed structural and energetic variability depending on the sequence context at positions 14 and 15, one helical turn away. Structurally, variable amounts of stacking between Trp18 and A7 were observed, with the wild-type and L14Q15

complexes being fully stacked and the E14R15 complex completely unstacked. It is likely that different complexes have different population distributions of stacked and unstacked conformations in dynamic exchange (27). In terms of the energetics and selection, the wild-type and L14Q15 sequence contexts showed that Trp18 and Tyr18 are the clear winners, whereas the best residues in other sequence backgrounds varied somewhat, and in the E14R15 context, Arg18, Ala18 and Trp18 appeared energetically similar.

Stacking, Not Energetics, Dictates *In Vivo* Function

Previous work from a number of groups has supported the idea that binding alone between N and *boxB* is sufficient to construct a functional antitermination switch. *In vitro* and *in vivo* studies have indicated that arginines 6, 7, 8, 10 and 11, as well as Trp18 or Tyr18, are critical for *in vivo* function as a result of their essential role in binding *boxB* RNA, whereas other positions such as Thr5, Glu9 and Glu13 can be readily substituted *in vitro* and *in vivo (19, 22)*. Previous *in vitro* selection work from our laboratory supported the idea that any substitutions for Arg6 and Arg7 disrupted stringent *boxB* RNA binding (23). Additionally, NMR and alanine scanning mutagenesis has also indicated that positions 3, 14 and 19 are important in recognition by phage N protein (17-19, 28). Finally, one group of researchers had reported an *in vivo* screen based solely on approximation between a nascent transcript and N mutants containing arginine-rich RNA-binding domains at their amino terminus (29). However, this system alone generated only 2% as much antitermination as the wild-type N–*boxB* system, suggesting that the distinct conformation of the components has an important role in generating a fully functional antitermination switch. Furthermore, some RNA mutations that support

N binding do not support the formation of higher-order complexes with the transcription elongation-termination factor NusA (13, 17). The identities of RNA bases in the pentaloop of *boxB* affect N binding (19) and antitermination (30) in somewhat different ways. G6 and A10 are absolutely required to maintain the GA sheared base pair, and purine bases are highly preferred for positions 7–9, with most pyrimidine substitutions resulting in <10% of wild-type antitermination activities except the A9U mutation (the flipped-out base), which has 35% of wild-type activity (30). Work in our laboratory had previously shown that wild-type N peptide discriminated between *boxB* RNA and a GAAA tetraloop with approximately ten-fold specificity (23).

These previous studies therefore motivated us to examine the *in vivo* activity of the N mutants we had studied. Our selected proteins are one-, two- and three-residue N mutants that preserve the known essential sequence elements that had been studied previously. Our measurements here of these mutants showed a wide range of stabilities (3.5 kcal mol^{-1}) and marked structural differences at the Trp18–A7 peptide-RNA interface. We were interested to see if differences in the conformation or stability of these complexes would affect *in vivo* antitermination activity. We inserted these sequences into a β-galactosidase reporter construct and measured transcription antitermination efficiency (22) in solution (Figure 5.4a) and on agar plates (Figure 5.4b,c). The activities of the complexes showed a full dynamic range, with some having activity equal to or better than that of the wild-type complex, whereas others show no activity, comparable to strains lacking N altogether (Figure 5.4a). In the wild-type sequence context (K14Q15), tryptophan and other aromatic residues were necessary at position 18 for efficient antitermination, consistent with previous *in vivo* work (22).

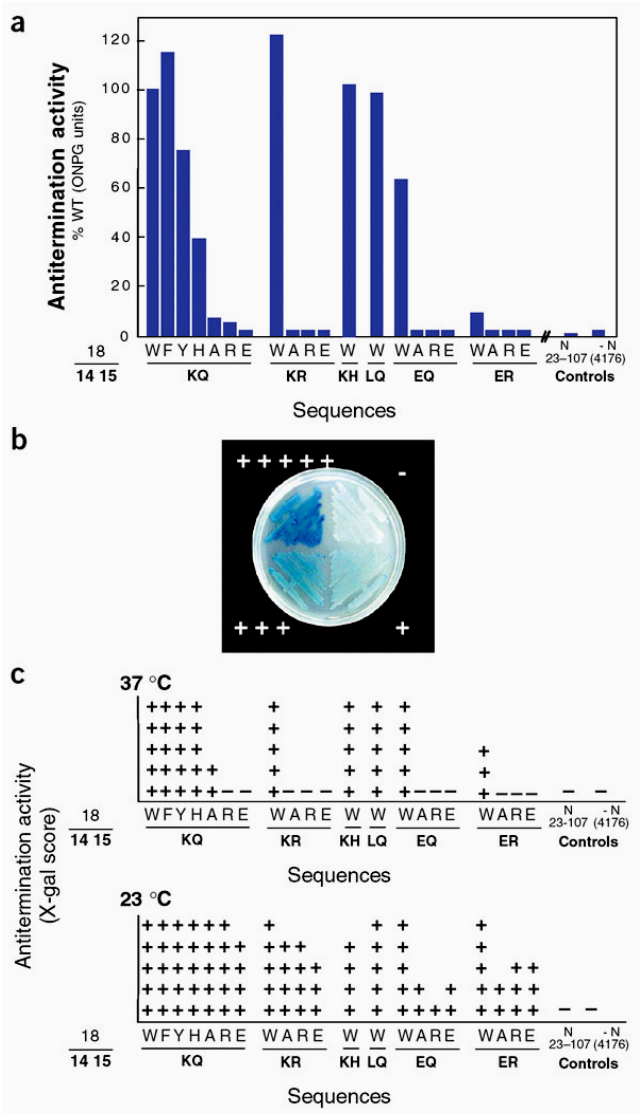

Figure 5.4. Antitermination activities of N mutants, as determined using *E. coli* strains carrying a two-plasmid reporter system (22). Functional antitermination through four terminators (tR1–4) results in expression of β-galactosidase. **(a)** Antitermination activity in solution, determined by quantifying β-galactosidase using a colorimetric assay based on *o*-nitrophenyl-β-*D*-galactopyranoside. β-galactosidase activity is expressed as a percentage of that of the wild-type (WT) N reporter construct (51). The residue identities at positions 14, 15 and 18 are indicated. Control constructs in which N lacks an RNA-binding domain (N(23–107)) or the N expressor plasmid is absent (N 4176) (22) are indicated. **(b)** Plate-scoring assay of mutants quantified on the basis of X-gal signal. **(c)** Plate-scoring assay of temperature sensitivity in the antitermination activity of N mutants grown at 37 °C or 23 °C.

However, assessment of the unstacked E14R15 mutant indicated that Trp18 alone is not sufficient for efficient antitermination, resulting in only about 9% the activity of the wild-type. This loss of activity did not seem to be caused by lack of expression, as both the E14R15 mutant and wild-type N proteins were expressed *in vivo* as judged by northwestern analysis (Figure 5.S2). Thus, variations at positions 14 and 15 that tend to unstack Trp18 resulted in little or no antitermination activity. Finally, constructs lacking the RNA-binding domain of N(23–107) showed no activity in this assay, indicating that the signal does not result from *boxB*-independent antitermination by N (31).

Antitermination activity tested on agar plates expanded the lower dynamic range of the assay and revealed that a number of our variants have temperature-sensitive phenotypes (Figure 5.4c). All variants that were positive at 37 °C also showed full activity at 23 °C. However, many variants that were inactive at 37 °C showed some level of antitermination activity at 23 °C. In particular, the fact that the E14R15 mutant has activity at lower temperature suggests that it binds *boxB* specifically *in vivo* and that its lack of functionality at higher temperature is not due to nonspecific RNA binding. Previous work has shown that some Nus proteins are dispensable for growth at lower temperatures (30 °C) (32), consistent with our observations that many more of the complexes are functional at the permissive temperature.

Our data demonstrated that the structure of the N–*boxB* complex interface, rather than its stability, is the principal determinant of functional antitermination (Figure 5.5). Our structural data suggested that the wild-type and mutant peptides have structural variation at the Trp18–A7 interface, while having essentially the same stem-helix interaction and overall loop folding (A9 extruded). The relative stability of the

complexes showed little correlation with antitermination activity (Figure 5.5a). Six of the complexes we examined showed stability within 1 kcal/mol of the wild-type complex but <10% antitermination activity, whereas one (E14Q15-W18), which binds more than 2.5 kcal mol^{-1} more weakly than the wild-type complex, had 63% of wild-type activity *in vivo* (Figure 5.5a). By contrast, complexes containing a stacked Trp18–A7 interface were fully active for antitermination (Figure 5.5b). Complexes in which Trp18 is unstacked (E14R15, Figure 5.5b). seemed to hinder or block processive antitermination. These data led us to conclude that there are different levels of control on the full dynamic range of transcription antitermination by N. The structure of the Trp18–A7 interface is critical in forming a fully functional antitermination switch, in addition to the effect of simple approximation of components (29).

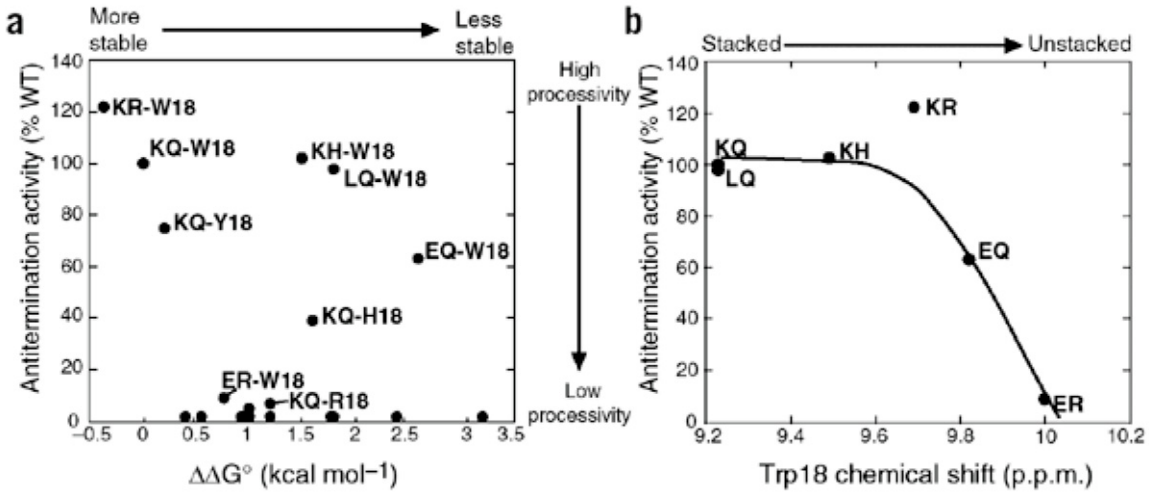

Figure 5.5. Energetic and structural correlation of antitermination activity. **(a)** No correlation is observed between the stability of the N–*boxB* complexes and their antitermination activity in solution. **(b)** Antitermination activity does correlate with Trp18-A7 stacking as detected by the chemical shift of Trp18 or the fluorescence of 2AP-7 (not shown).

Role of NusA

Our conclusion raised the questions of how the N-*boxB* conformation is sensed by the elongating polymerase, and how a single stacking interaction can dictate the processivity of a 400-kDa molecular machine. NusA is the most likely suspect as a conformational sensor. NusA has an important role in regulating the processivity of RNA polymerase (33-36). Previous work has indicated that NusA can discriminate between N–*boxB* complexes differing in a single nucleotide (13, 17) and can bind to N with high affinity at a site distinct from the RNA-binding domain (12). We therefore examined the binding and fluorescence effects of NusA on various 2AP-labeled N–*boxB* complexes (Figure 5.6).

NusA showed no detectable interaction with complexes constructed from an N peptide (either wild-type or E14R15 N(1–22)) and a 2AP-labeled *boxB* (data not shown), indicating that the N-*boxB* interface alone may not be sufficient to recruit NusA binding with high affinity. This observation is consistent with previous data showing that N protein is needed to bind NusA in order for NusA to bind RNA (37). NusA bound to complexes containing full-length N protein (wild-type or E14R15 N(1–107)) (Figure 5.6a and Figure 5.S3). In the wild-type N(1–107)–*boxB* complex, NusA quenched the fluorescence of the extruded base (2AP-9), but modestly enhanced the emission from 2AP at the crucial Trp18-A7 interface. In the E14R15 N(1–107)–*boxB* complex, NusA quenched both the extruded base (2AP-9) and the base located at the top of the loop (2AP-7). Thus, NusA seems to drive the binary N protein–*boxB* complexes toward similar, but distinct, final conformational states. Furthermore, NusA binds with similar, high-affinity K_d values to both wild-type and E14R15 complexes (14 ± 6 versus 9 ± 2

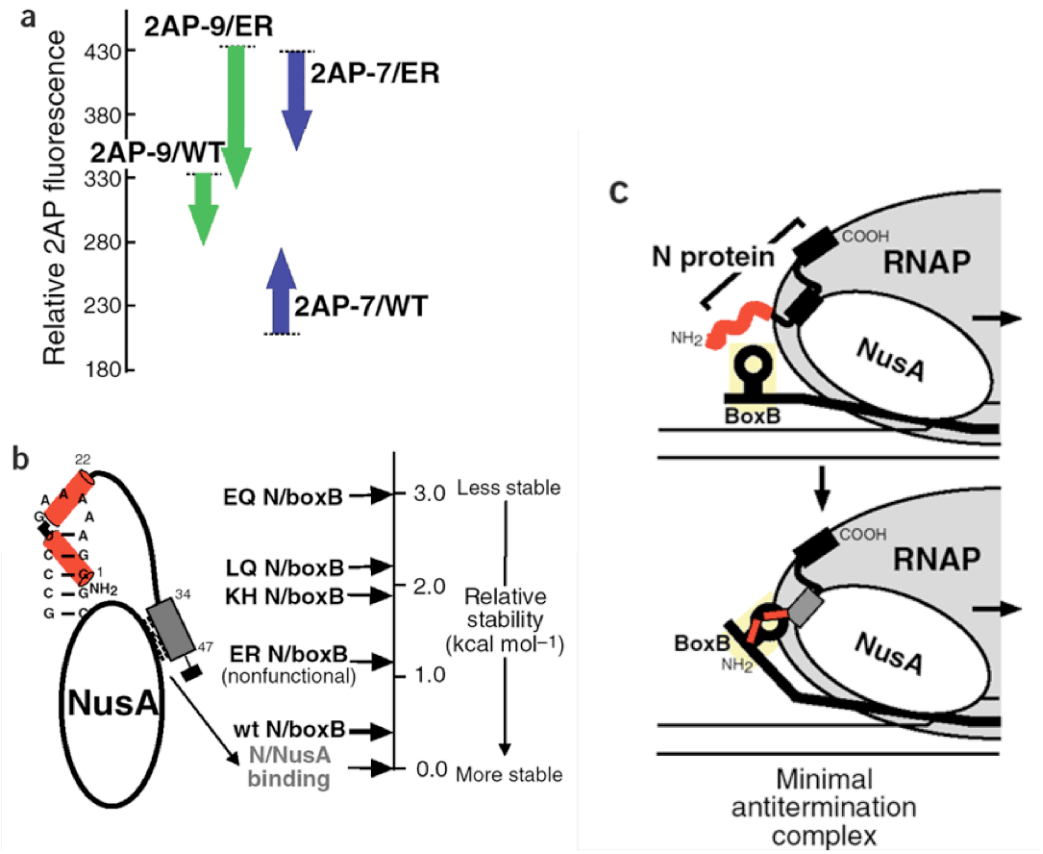

Figure 5.6. Characteristics of N-*boxB*-NusA interactions. **(a)** Fluorescence changes induced by NusA binding to N–*boxB* complexes. Binary complexes were constructed using wild-type N(1–107) or E14R15 N(1–107) and 2AP-7 (blue) or 2AP-9 (green) labeled *boxB* RNA. The arrows indicate the 2AP fluorescence change (starting to final) induced when NusA is added. **(b)** Relative energetics of NusA, N protein and *boxB* interactions. The N-NusA interaction is stronger than N binding to the *boxB* RNA. The wild-type (WT), K14H15, L14Q15 and E14Q15 complexes show >50% wild-type antitermination function in vivo, whereas ER shows <10% activity. **(c)** Model for formation of the minimal antitermination complex in vivo. The tight binding of N to NusA implies that N is prebound to polymerase during elongation and scans the message as it emerges from the transcription elongation complex. As the *boxB* hairpin emerges from the transcription bubble, the N-terminal domain of N engages the message to generate the minimal antitermination complex.

nM, respectively; Figure 5.S4), suggesting that NusA affinity is dictated by protein-protein contacts between N protein and NusA, rather than the N-*boxB* interface, in line with previous observations (12).

Notably, the K_d for NusA and N protein determined by van Gilst and von Hippel under more stringent temperature and salt conditions (12) actually indicates stronger binding between NusA and N protein (70 nM) than that of N protein to *boxB* RNA (K_d 127 nM under the same conditions; Figure 5.S5) (Figure 5.6b). An energetic ranking of a number of functional and nonfunctional N-*boxB* interactions indicates that the N-NusA interaction is as much as 3 kcal mol^{-1} more favorable than the RNA-protein interactions needed to construct a functional antitermination switch *in vivo* (Figure 5.6b).

Discussion

A Novel Model for Antitermination

In vivo, the N–*boxB* RNA complex is believed to serve the bipartite function of tethering NusA to the *nut* site RNA and sequestering N to interact directly with RNA polymerase (38, 39). The relative energetics of N-NusA and N-*boxB* interactions lead us to propose a new model for assembly of the minimal antitermination complex (Figure 5.6c). The very high affinity between N and NusA implies that N can be bound to the elongating polymerase, via interactions with NusA, where NusA associates with elongating polymerase with similar affinity (30–100 nM) (33, 34). This complex can be assembled stably before the exit of the *nut* site from polymerase. This view differs from the model generally described for antitermination, whereby N is recruited to the

transcription elongation complex (TEC) via interactions with *boxB* (Figure 5.1b) (13, 38, 40, 41). Previous structural modeling of TECs supports the idea that polymerase constantly scans the nascent transcript for termination signals via the flap-tip helix and the action of bound NusA (42). Our work implies that N protein modifies NusA, altering the function of RNA polymerase through NusA and RNAP interactions when it properly engages an appropriate RNA structure. This view of N action is consistent with experiments showing that N supports efficient transcription antitermination even in the absence of *boxB (16, 36) and with modeling efforts focused on understanding transcription termination (42). We conclude that conformational sensing of the Trp18-A7 interface occurs in the context of a TEC, containing N, NusA and RNA polymerase itself, after N has engaged the nascent boxB* hairpin. Essentially, the unfolded N acts as a template onto which the components of the antitermination complex assemble. This action of N is apparently dependent on the exact conformation of the interface of Trp18-A7, where the presence or absence of the stacking directs NusA and polymerase into either different orientations and spatial arrangements or different physical contacts. Supporting this view is previous work indicating that the quaternary NusA–N–*nut* polymerase complex can be formed and is stable with respect to competition from exogenous components (36). Although these components represent a minimal antitermination complex, assembly of the complete complex (containing N, NusA, NusB, NusE, NusG, *boxA*, *boxB* and RNAP) is essential for processive antitermination under stringent conditions *in vivo (14).*

Implications

Our results obtained under stringent conditions suggest that proper assembly of the RNA-protein interface is indispensable for full antitermination activity *in vivo*. One consequence of our observations is the conclusion that genetic screens using antitermination to isolate RNA-binding peptides (29) will likely be influenced by constraints of functional antitermination.

The N-*boxB* interface provides a clear example of adaptive recognition in which the sequence context dictates the structural meaning of interfacial residues. *In vivo*, the various conformations we observed had profound effects on polymerase function and demonstrated that approximation of components is not sufficient to generate a functional transcriptional switch. Rather, the transcription elongation complex has the means to sense the conformation of *cis*-acting regulatory elements, much like a stockbroker reading a ticker tape, profoundly altering its processivity accordingly. Similarly, it seems likely that a more complicated eukaryotic elongation complex could distinguish between various *cis*-acting regulatory signals such as those provided by HIV Tat (43, 44) and Nrd1p (45-47). In addition to revealing the structural origin of this control process, our work indicates that ligands targeting elements of elongation complexes (48) offer potential tools to better understand and control the mechanisms of gene expression in both prokaryotic and eukaryotic systems.

Materials and Methods

In vitro Peptide Selection

Generation and purification of mRNA-peptide fusions and selection cycles were done as previously described (23). Approximately 5 pmol of purified ^{35}S-labeled fusions were incubated with 200 pmol of biotinylated *boxB* RNA (5'-GGCCCUGAAAAGGG CCAAA-biotin-3') in the selection step in the presence of yeast tRNA competitor (Roche).

Synthesis of RNAs and Peptides

The 2-aminopurine (2AP)-labeled *boxB* RNAs, 5'-GGCCCUGAAAAAGGGCC-3' (with underlined nucleotides individually labeled 2AP-7, 2AP-8, 2AP-9), were generated by solid-phase synthesis using standard phosphoramidite chemistry, and then deprotected and purified by 20% urea-PAGE. RNA without the 2AP label for NMR experiments (5'-GCCCUGAAAAAGGGC-3') was synthesized using T7 RNA polymerase and a synthetic DNA template (49). Peptides MDAQTRRRERRAEXXAQXKAAN(gy) (where X represents variable residues and gy represents extra residues for quantification purposes for peptides lacking Trp18) were made by automated Fmoc synthesis (using an ABI 432 peptide synthesizer), deprotected, and purified by reversed-phase HPLC.

Binding Studies

K_d values for N-peptide binding to *boxB* RNA were determined by monitoring the fluorescence change of RNAs labeled with 2AP at the seventh, eighth or ninth position

upon addition of concentrated peptide and were fitted as previously described (23). NusA binding was performed by titrating NusA solution into preformed N–*boxB* complex. Free energies were calculated on the basis of complex stability at 20 °C in standard binding buffer (50 mM potassium acetate, 20 mM Tris, pH 7.5). Individual K_d values were reproducible with an error of 0.2 kcal mol^{-1}.

NMR Spectroscopy

Spectra were collected on a Varian Inova 600 at 25 °C in a buffer of 50 mM NaCl, 10 mM sodium phosphate, 0.5 mM EDTA, pH 6. Complex formation was monitored by inspecting imino protons of the RNA. The concentrations of the complexes were 150–300 μM.

Circular Dichroism

CD spectra were collected on an Aviv 62 DS CD spectrofluorimeter in 10 mM potassium phosphate buffer at 20 °C. Spectra for the peptide portion in the complex were calculated by subtracting spectra for free RNA and excess peptide from that of the complex. Helical content was calculated as described previously (48).

Antitermination Assay

Strains expressing wild-type and mutant N protein were constructed using the two plasmid reporter system previously described (22). All sequences were verified by sequencing. N-mutant strains were plated on tryptone agar supplemented with 0.05 mM IPTG, 0.08 mg ml^{-1} X-gal and the appropriate antibiotics. Plates were incubated at either

23 °C (permissive temperature) or 37 °C (nonpermissive temperature) and scored for blue color. Temperature sensitivity was defined here as the loss of relative β-galactosidase units in comparing the blue color of colonies grown at permissive and nonpermissive temperatures. All assays were performed in triplicate. Intensities of blue color are shown for sample N–mutant strains K14Q15E18 (-), K14Q15A18 (+), Nun-N fusion (+++) and K14Q15W18 (+++++).

Protein Expression and Purification

Wild-type and E14R15 N protein were expressed from plasmid pET-N1 in BL21 (DE3) gold cells as described (12), except that a SP-cation exchange column (Amersham) was used for the final purification step and the protein was dialyzed into fluorescence binding buffer (50 mM KCl, 20 mM Tris-HCl, pH 7.5, 1 mM β-mercaptoethanol). The NusA gene was amplified from plasmid pMS7, subcloned into vector pET19b (Novagen) and verified by sequencing. BL21(DE3) cells harboring the resulting plasmid were induced at an OD_{600} of 0.4 with 1 mM IPTG, grown for 3 h, harvested by centrifugation and resuspended 20 ml buffer A (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM β-mercaptoethanol, 1 mM imidazole) per liter of culture. The cells were lysed by three passes through a French press and clarified by two centrifugations at 18,000x*g* for 30 min at 4 °C. The supernatant was loaded on a Superflow Ni-NTA column (Qiagen), washed with 5 column volumes of buffer A, with 5 column volumes of buffer A containing 500 mM NaCl, and eluted with 400 mM imidazole in buffer A. The eluted protein was dialyzed against fluorescence binding buffer.

N Protein Expression Level

Expression of wild-type N protein, E14R15 and N(23–107) variants was accomplished as above except that the cells were induced with 0.5 mM IPTG at an OD_{600} of 0.3 and grown for 3 h, and 0.1 OD units of culture was concentrated and resuspended in 100 μl of 6x SDS loading buffer (0.28 M Tris-Cl/SDS, pH 6.8, 30% (v/v) glycerol, 1% (w/v) SDS, 0.5 M DTT). Five microliters of this solution was separated by 15% SDS-PAGE and then transferred to nitrocellulose (Bio-Rad) in transfer buffer (10 mM NaHCO3-3 mM Na2CO3, pH 9.9, 20% (v/v) methanol, 0.05% (w/v) SDS) at 4 °C. The nitrocellulose was blocked at room temperature in wash buffer (10 mM Tris-HCl, pH 7.5,100 mM NaCl, 1 mM EDTA, pH 8.0, 0.01% (v/v) NP-40) containing 5% (w/v) nonfat dry milk and 50 μg per milliliter of yeast tRNA, and then 100 nM biotinylated *boxB* RNA (5'-GGCCCUGAAAAGGGCCAAA-biotin-3') was added. The nitrocellulose was washed three times with wash buffer, probed with 0.5 g ml^{-1} streptavidin–horseradish peroxidase (Amersham) in wash buffer containing 0.5% (w/v) BSA, washed four times and visualized using Supersignal West Pico substrate (Pierce).

Acknowledgments

We thank N. Franklin for the gifts of the two-plasmid N expressor–-galactosidase reporter constructs, plasmid pMS7 and N–tester strain, P. von Hippel for the gift of pET-N1 plasmid and N protein, P. Legault for sharing the coordinates of the phage N peptide–*boxB* RNA complex and J.H. Richards, C.S. Parker and members of the Roberts laboratory for helpful comments on the manuscript. This work was supported by

grants from US National Science Foundation (NSF) and National Institutes of Health (NIH) to R.W.R. R.W.R. is an Alfred P. Sloan research fellow, A.F. is an ACS postdoctoral fellow and T.T.T. was supported by an NIH training grant.

Competing Interests Statement:

The authors declare that they have no competing financial interests.

Supplemental Information

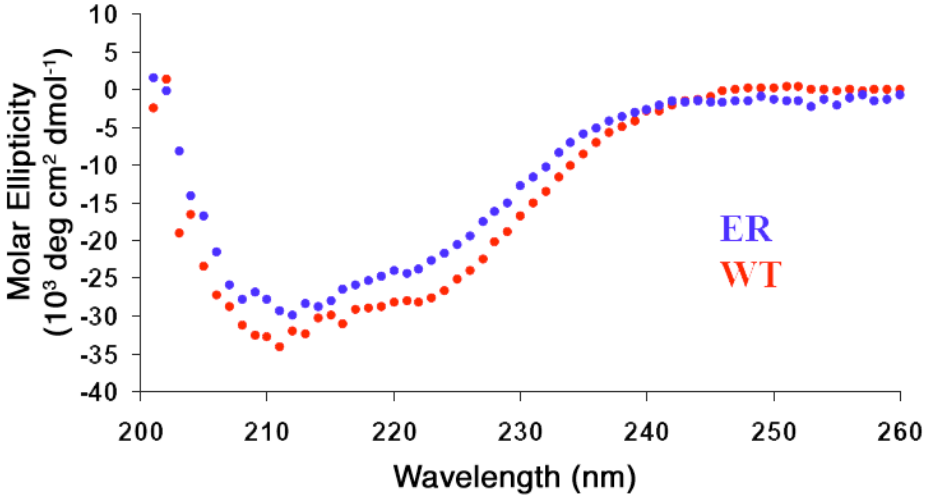

Figure 5.S1. CD spectra of the wild-type and ER peptides in complex with *boxB* RNA. Spectra were collected in 10 mM phosphate buffer at 20 °C. The contributions of the peptide portion in the complex were calculated by subtracting the RNA signal from the complex signal as described previously (48). The average number of amino acid residues in helical form is 16.4 and 13.8 for wild-tye and ER, respectively.

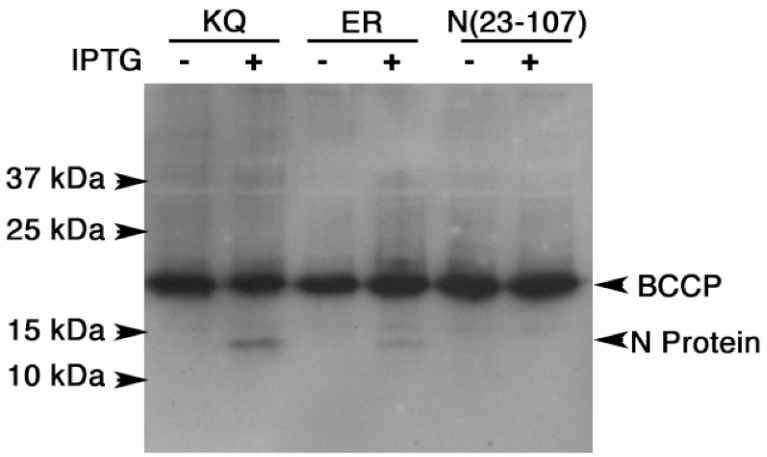

Figure 5.S2. Expression levels of full-length wild-type N protein and ER protein. Cell lysates containing wild-type N protein (KQ) or ER protein (ER) (+ IPTG lanes) were run on an SDS-PAGE gel and transferred to nitrocellulose. Wild-type and ER proteins were probed with biotinylated *boxB* RNA and visualized with streptavidin-HRP. The N(23-107) protein lacking the N-terminal RNA binding domain and lanes where cells were uninduced (- IPTG lanes) show no RNA binding. An endogenous biotinylated *E. Coli* protein (possibly BCCP) also reacts with streptavidin-HRP.

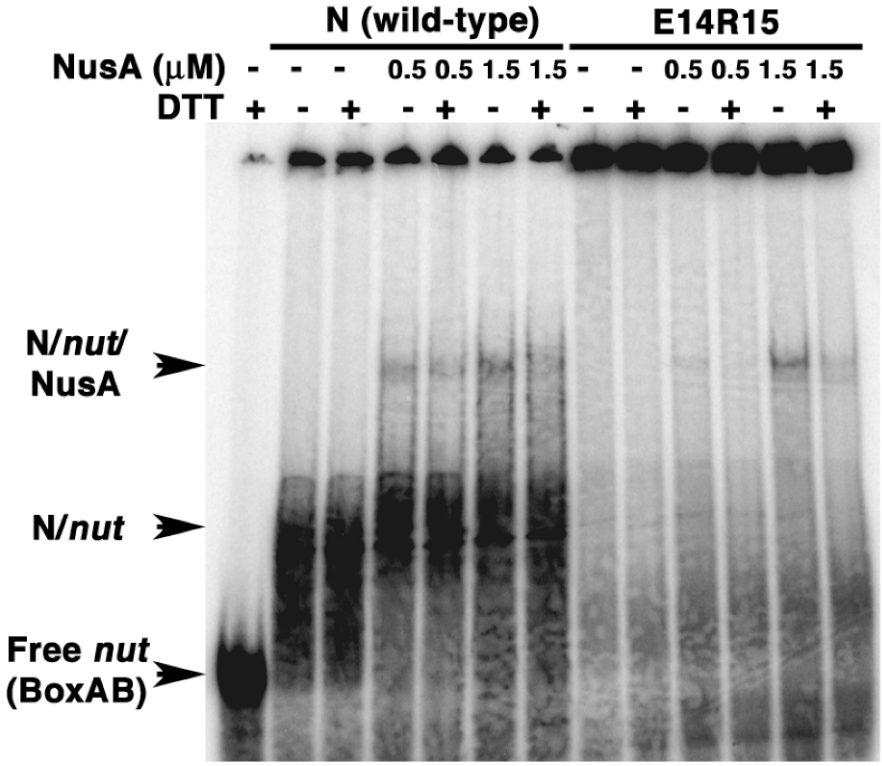

Figure 5.S3. Interaction of NusA with full-length wild-type N protein or ER protein. ^{32}P-labeled nut RNA (*boxA/boxB*) was incubated with either wild-type N protein (N (wt)) or E14R15 protein (ER) with increasing concentrations of NusA. Samples were run in the presence (+) or absence (-) of DTT. N and ER bind the *nut* RNA giving a band of slower mobility which can be supershifted with the addition of NusA. NusA supershifts both wild-type/*nut* and ER/*nut* complexes.

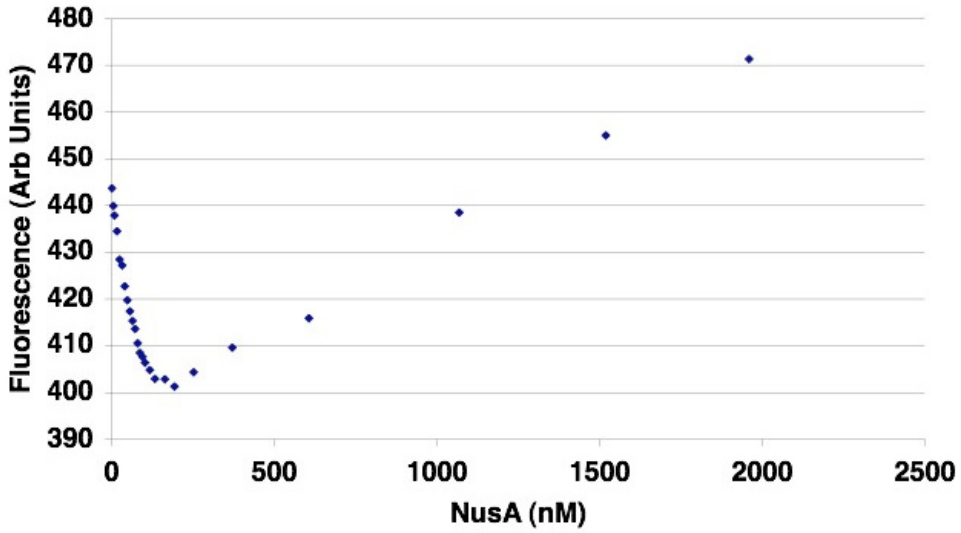

Figure 5.S4. A representative fluorescence titration isotherm of NusA binding to wild-type N in complex with *boxB* labeled at position 9. N protein and *boxB* complexes were preformed at 100 nM (1:1) in 50 mM KOAc, 20 mM Tris-OAc, pH 7.5 buffer at 20 ^{o}C. Measured dissociation constants are: 14 ± 6 nM (2AP-7), and 3 ± 1 nM (2AP-9) for NusA binding to the wildtype N-*boxB* complex; 9 ± 2 nM (2AP-7), and 35 ± 5 nM (2AP-9) for NusA binding to the ER-*boxB* complex.

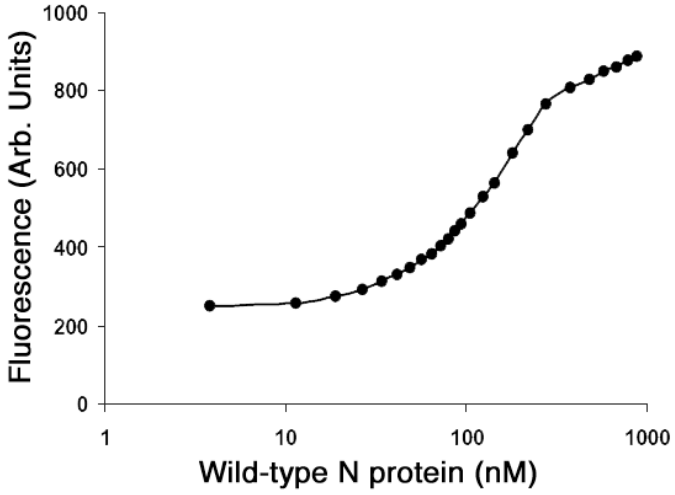

Figure 5.S5. Fluorescence titration isotherms of wild-type N protein to *boxB* RNA labeled at position 9 with 2-aminopurine (2AP-9). *BoxB* concentration is 100 nM. The titration is performed in 150 mM KOAc, 40 mM Tris-OAc, 1 mM DTT, 0.1 mM EDTA, pH 7.5 buffer at 30 oC. Measured dissociation constant is 127 nM.

References

1. Von Hippel, P.H. An integrated model of the transcription complex in elongation, termination, and editing. (1998) *Science* **281**, 660-665.
2. Yarnell, W.S. and Roberts, J.W. Mechanism of intrinsic transcription termination and antitermination. (1999) *Science* **284**, 611-615.
3. Uptain, S.M., Kane, C.M. and Chamberlin, M.J. Basic mechanisms of transcript elongation and its regulation. (1997) *Annu Rev Biochem* **66**, 117-172.
4. Zorio, D.A. and Bentley, D.L. Transcription elongation: the 'Foggy' is lifting... (2001) *Curr Biol* **11**, R144-146.
5. Maniatis, T. and Reed, R. An extensive network of coupling among gene expression machines. (2002) *Nature* **416**, 499-506.
6. Roberts, J.W. Termination factor for RNA synthesis. (1969) *Nature* **224**, 1168-1174.
7. Greenblatt, J. Positive control of endolysin synthesis *in vitro* by the gene N protein of phage lambda. (1972) *Proc Natl Acad Sci U S A* **69**, 3606-3610.
8. Salstrom, J.S. and Szybalski, W. Coliphage lambdanutL-: a unique class of mutants defective in the site of gene N product utilization for antitermination of leftward transcription. (1978) *J Mol Biol* **124**, 195-221.
9. Greenblatt, J., Nodwell, J.R. and Mason, S.W. Transcriptional antitermination. (1993) *Nature* **364**, 401-406.
10. Friedman, D.I. and Court, D.L. Transcription antitermination: the lambda paradigm updated. (1995) *Mol Microbiol* **18**, 191-200.

11. Weisberg, R.A. and Gottesman, M.E. Processive antitermination. (1999) *J Bacteriol* **181**, 359-367.

12. Van Gilst, M.R. and Von Hippel, P.H. Assembly of the N-dependent antitermination complex of phage lambda: NusA and RNA bind independently to different unfolded domains of the N protein. (1997) *J Mol Biol* **274**, 160-173.

13. Mogridge, J., Mah, T.F. and Greenblatt, J. A protein-RNA interaction network facilitates the template-independent cooperative assembly on RNA polymerase of a stable antitermination complex containing the lambda N protein. (1995) *Genes Dev* **9**, 2831-2845.

14. Zhou, Y., Mah, T.F., Yu, Y.T., Mogridge, J., Olson, E.R., Greenblatt, J. and Friedman, D.I. Interactions of an Arg-rich region of transcription elongation protein NusA with NUT RNA: implications for the order of assembly of the lambda N antitermination complex in vivo. (2001) *J Mol Biol* **310**, 33-49.

15. Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O. and Darst, S.A. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. (2002) *Science* **296**, 1285-1290.

16. Gusarov, I. and Nudler, E. Control of intrinsic transcription termination by N and NusA: the basic mechanisms. (2001) *Cell* **107**, 437-449.

17. Legault, P., Li, J., Mogridge, J., Kay, L.E. and Greenblatt, J. NMR structure of the bacteriophage lambda N peptide/*boxB* RNA complex: recognition of a GNRA fold by an arginine-rich motif. (1998) *Cell* **93**, 289-299.

18. Scharpf, M., Sticht, H., Schweimer, K., Boehm, M., Hoffmann, S. and Rosch, P. Antitermination in bacteriophage lambda. The structure of the N36 peptide-*boxB* RNA complex. (2000) *Eur J Biochem* **267**, 2397-2408.

19. Su, L., Radek, J.T., Hallenga, K., Hermanto, P., Chan, G., Labeots, L.A. and Weiss, M.A. RNA recognition by a bent alpha-helix regulates transcriptional antitermination in phage lambda. (1997) *Biochemistry* **36**, 12722-12732.

20. Jucker, F.M., Heus, H.A., Yip, P.F., Moors, E.H. and Pardi, A. A network of heterogeneous hydrogen bonds in GNRA tetraloops. (1996) *J Mol Biol* **264**, 968-980.

21. Heus, H.A. and Pardi, A. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. (1991) *Science* **253**, 191-194.

22. Franklin, N.C. Clustered arginine residues of bacteriophage lambda N protein are essential to antitermination of transcription, but their locale cannot compensate for boxB loop defects. (1993) *J Mol Biol* **231**, 343-360.

23. Barrick, J.E., Takahashi, T.T., Ren, J., Xia, T. and Roberts, R.W. Large libraries reveal diverse solutions to an RNA recognition problem. (2001) *Proc Natl Acad Sci U S A* **98**, 12374-12378.

24. Barrick, J.E., Takahashi, T.T., Balakin, A. and Roberts, R.W. Selection of RNA-binding peptides using mRNA-peptide fusions. (2001) *Methods* **23**, 287-293.

25. Barrick, J.E. and Roberts, R.W. Sequence analysis of an artificial family of RNA-binding peptides. (2002) *Protein Sci* **11**, 2688-2696.

26. Lacourciere, K.A., Stivers, J.T. and Marino, J.P. Mechanism of neomycin and Rev peptide binding to the Rev responsive element of HIV-1 as determined by fluorescence and NMR spectroscopy. (2000) *Biochemistry* **39**, 5630-5641.

27. Xia, T., Becker, H.C., Wan, C., Frankel, A., Roberts, R.W. and Zewail, A.H. The RNA-protein complex: direct probing of the interfacial recognition dynamics and its correlation with biological functions. (2003) *Proc Natl Acad Sci U S A* **100**, 8119-8123.

28. Cai, Z., Gorin, A., Frederick, R., Ye, X., Hu, W., Majumdar, A., Kettani, A. and Patel, D.J. Solution structure of P22 transcriptional antitermination N peptide-*boxB* RNA complex. (1998) *Nat Struct Biol* **5**, 203-212.

29. Harada, K., Martin, S.S. and Frankel, A.D. Selection of RNA-binding peptides *in vivo*. (1996) *Nature* **380**, 175-179.

30. Doelling, J.H. and Franklin, N.C. Effects of all single base substitutions in the loop of *boxB* on antitermination of transcription by bacteriophage lambda's N protein. (1989) *Nucleic Acids Res* **17**, 5565-5577.

31. Das, A., Pal, M., Mena, J.G., Whalen, W., Wolska, K., Crossley, R., Rees, W., Von Hippel, P.H., Costantino, N., Court, D., Mazzulla, M., Altieri, A.S., Byrd, R.A., Chattopadhyay, S., Devito, J. and Ghosh, B. Components of multiprotein-RNA complex that controls transcription elongation in *Escherichia coli* phage lambda. (1996) *Methods Enzymol* **274**, 374-402.

32. Friedman, D.I. (1998) The Bacteriophages (Plenum, New York.

33. Whalen, W., Ghosh, B. and Das, A. NusA protein is necessary and sufficient *in vitro* for phage lambda N gene product to suppress a rho-independent terminator placed downstream of nutL. (1988) *Proc Natl Acad Sci U S A* **85**, 2494-2498.

34. Gill, S.C., Weitzel, S.E. and Von Hippel, P.H. *Escherichia coli* sigma 70 and NusA proteins. I. Binding interactions with core RNA polymerase in solution and within the transcription complex. (1991) *J Mol Biol* **220**, 307-324.

35. Mason, S.W., Li, J. and Greenblatt, J. Host factor requirements for processive antitermination of transcription and suppression of pausing by the N protein of bacteriophage lambda. (1992) *J Biol Chem* **267**, 19418-19426.

36. Rees, W.A., Weitzel, S.E., Yager, T.D., Das, A. and Von Hippel, P.H. Bacteriophage lambda N protein alone can induce transcription antitermination in vitro. (1996) *Proc Natl Acad Sci U S A* **93**, 342-346.

37. Mah, T.F., Kuznedelov, K., Mushegian, A., Severinov, K. and Greenblatt, J. The alpha subunit of *E. coli* RNA polymerase activates RNA binding by NusA. (2000) *Genes Dev* **14**, 2664-2675.

38. Chattopadhyay, S., Garcia-Mena, J., Devito, J., Wolska, K. and Das, A. Bipartite function of a small RNA hairpin in transcription antitermination in bacteriophage lambda. (1995) *Proc Natl Acad Sci U S A* **92**, 4061-4065.

39. Mogridge, J., Legault, P., Li, J., Van Oene, M.D., Kay, L.E. and Greenblatt, J. Independent ligand-induced folding of the RNA-binding domain and two functionally distinct antitermination regions in the phage lambda N protein. (1998) *Mol Cell* **1**, 265-275.

40. Barik, S., Ghosh, B., Whalen, W., Lazinski, D. and Das, A. An antitermination protein engages the elongating transcription apparatus at a promoter-proximal recognition site. (1987) *Cell* **50**, 885-899.

41. Mason, S.W. and Greenblatt, J. Assembly of transcription elongation complexes containing the N protein of phage lambda and the *Escherichia coli* elongation factors NusA, NusB, NusG, and S10. (1991) *Genes Dev* **5**, 1504-1512.

42. Toulokhonov, I., Artsimovitch, I. and Landick, R. Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. (2001) *Science* **292**, 730-733.

43. Marciniak, R.A., Calnan, B.J., Frankel, A.D. and Sharp, P.A. HIV-1 Tat protein trans-activates transcription *in vitro*. (1990) *Cell* **63**, 791-802.

44. Richter, S., Ping, Y.H. and Rana, T.M. TAR RNA loop: a scaffold for the assembly of a regulatory switch in HIV replication. (2002) *Proc Natl Acad Sci U S A* **99**, 7928-7933.

45. Steinmetz, E.J. and Brow, D.A. Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. (1996) *Mol Cell Biol* **16**, 6993-7003.

46. Steinmetz, E.J. and Brow, D.A. Control of pre-mRNA accumulation by the essential yeast protein Nrd1 requires high-affinity transcript binding and a domain implicated in RNA polymerase II association. (1998) *Proc Natl Acad Sci U S A* **95**, 6993-7003.

47. Steinmetz, E.J., Conrad, N.K., Brow, D.A. and Corden, J.L. RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. (2001) *Nature* **413**, 327-331.

48. Austin, R.J., Xia, T., Ren, J., Takahashi, T.T. and Roberts, R.W. Designed arginine-rich RNA-binding peptides with picomolar affinity. (2002) *J Am Chem Soc* **124**, 10966-10967.

49. Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. (1987) *Nucleic Acids Res* **15**, 8783-8798.

50. Rees, W.A., Weitzel, S.E., Das, A. and Von Hippel, P.H. Regulation of the elongation-termination decision at intrinsic terminators by antitermination protein N of phage lambda. (1997) *J Mol Biol* **273**, 797-813.

51. Peled-Zehavi, H., Smith, C.A., Harada, K. and Frankel, A.D. Screening RNA-binding libraries by transcriptional antitermination in bacteria. (2000) *Methods Enzymol* **318**, 297-308.