Structural mimicry in the phage ϕ 21 N peptide–*boxB* RNA complex

CHRISTOPHER D. CILLEY and JAMES R. WILLIAMSON

Department of Molecular Biology, Department of Chemistry, and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037, USA

ABSTRACT

We determined the solution structure of a 22-amino-acid peptide from the amino-terminal domain of the bacteriophage ϕ 21 N protein in complex with its cognate 24-mer *boxB* RNA hairpin using heteronuclear magnetic resonance spectroscopy. The N peptide binds as an α -helix and interacts predominately with the major groove side of the 5' half of the *boxB* RNA stem-loop. This binding interface is defined by surface complementarity of polar and nonpolar interactions, and little sequence-specific recognition. The ϕ 21 *boxB* loop (CUAACC) has hydrogen bond and backbone torsions typical of the "U-turn" motif, as well as base stacking of the last 4 nt, and a hydrogen bonded C:C pair closing the loop. The exposed face of the ϕ 21 *boxB* loop, in complex with the N peptide, is strikingly similar to the GNRA tetraloop-like folds of the related λ and P22 bacteriophage N peptide–*boxB* RNA complexes. The N peptide–*boxB* complexes of the various phage, while individually distinct, provide similar structural features for interactions with the *Escherichia coli* host factors to enable antitermination.

Keywords: Transcription antitermination; bacteriophage N protein; NMR spectroscopy; U-turn; protein/RNA interactions

INTRODUCTION

The Lambda family of bacteriophage (λ , ϕ 21, and P22) use a temporally programmed sequence of gene expression during infection that is controlled, in part, by termination and antitermination of transcription. These bacteriophage all use N proteins, N-utilization (*nut*) sites in the nascent mRNA, and transcription terminators, to regulate early transcription. The interaction between the phage-encoded N protein and the RNA *nut* site plays a central role in the organization of viral and host factors (Nus proteins) required for formation of a termination resistant form of the RNA polymerase (RNAP) transcription complex (Das 1993; Greenblatt et al. 1993; Friedman and Court 1995).

The *nut* site is composed of a single-stranded 5'–*boxA* site and a 3'–*boxB* RNA hairpin, separated by a variable length linker region of 8–14 nt (Fig. 1A,B; Szybalski et al. 1986). The organization of the genome is conserved between phage types, but the sequences of the RNA *nut* sites and N proteins are diverse (Franklin 1985a). The *boxA* sequences are highly homologous between the related phage and have

been shown to bind host-encoded NusB and ribosomal protein S10 (Olson et al. 1984; Horwitz et al. 1987; Nodwell and Greenblatt 1993). The *boxB* hairpins consist of short stems (5–7 bp) and 5- or 6-nt loops and have little sequence homology between the three phages. Point mutations in λ *boxB* loop sequences have been shown to abolish antitermination in vivo (Doelling and Franklin 1989; Chattopadhyay et al. 1995) and N binding in vitro (Cilley and Williamson 1997).

The three related phage N proteins are all small (96–107 amino acids), basic proteins that only function with their cognate *nut-boxB* RNAs (Lazinski et al. 1989). Amino acid sequence alignment of the N proteins from λ , ϕ 21, and P22 shows an amino-terminal, arginine-rich, 18-amino-acid region of homology with four invariant residues (Fig. 1C; Franklin 1985b). In vivo and in vitro studies show peptides from this region specifically bind the appropriate *boxB* hairpins (Lazinski et al. 1989; Tan and Frankel 1995), with affinities similar to full-length N protein (Cilley and Williamson 1997). The N protein binds to one face of the *boxB* hairpin (Chattopadhyay et al. 1995), whereas *Escherichia coli* elongation factor NusA binds the opposite face (Mogridge et al. 1995) and interacts directly with N (Van Gilst and von Hippel 1997).

The solution structures of the N peptide–*boxB* RNA complex have been solved for the phages λ (Legault et al. 1998)

Reprint requests to: James R. Williamson, The Scripps Research Institute, 10550 North Torrey Pines Rd., MB-33, La Jolla, CA 92037, USA; e-mail: jrwill@scripps.edu.

Article and publication are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.2189203>.

FIGURE 1. Alignments of the three phage *nut* RNAs and N proteins. (A) Schematic of the secondary structure of the *nut* site RNA highlighting the *boxA* (red), linker (black), and *boxB* (blue) domains. (B) Sequence alignment of the three phage *nut* RNAs. The *boxB* loop nucleotides are shown in blue italics. (C) Sequence alignment of the amino-terminal domains of the three phage N proteins. The four conserved residues are highlighted in green.

and P22 (Cai et al. 1998). Both structures show a bent α -helical peptide bound to one face of an RNA hairpin loop. A number of polar and nonpolar interactions are observed with few, if any, involved in sequence-specific contacts in either of the structures. Each hairpin has a pentanucleotide loop that adopts a GNRA-type tetraloop fold with the fourth or third nucleotide flipped out of the loop in λ and P22, respectively.

Here, we describe the solution structure of the ϕ 21 N-*boxB* consisting of an amino-terminal 22-amino-acid N peptide in complex with a 24-nt *boxB* RNA by NMR methods. The *boxB* hairpin has an A-form helical stem with a hexanucleotide loop (5'-CUAACC-3') with the last four bases continuously stacking on the 3' half of the hairpin stem. The loop residues U11-A12-A13 form a "U-turn," first described in yeast tRNA^{phe} (Quigley and Rich 1976) and also observed in ribosomal RNA loops (Fountain et al. 1996; Huang et al. 1996; Conn et al. 1999; Zhang et al. 2001), and the hammerhead ribozyme (Pley et al. 1994). The U-turn consensus sequence, UNR, is characterized by a hydrogen bond between the U imino proton and the 5-phosphate of the nucleotide following the UNR, and a second hydrogen bond between the U 2' hydroxyl and the N7 of the R nucleotide. U-turns are commonly flanked by non-canonical Y:Y base pairs (Auffinger and Westhof 1999; Guttell et al. 2000). A core 13-amino-acid region of N ^{ϕ 21} peptide binds as an α -helix and interacts predominately with both the major groove side of the 5' half of the ascending upper stem, and the loop of the *boxB* hairpin. Electrostatic and hydrogen bond interactions exist all along the phosphodiester backbone of the RNA.

This structure completes the family of three λ -related phage N-*boxB* protein-RNA structures that have been genetically characterized. All three peptides bind the RNA helical stems using polar and nonpolar contacts that are almost superimposable. Peptide interactions with their respective RNA hairpin loops diverge from this structural similarity, and all three peptides show loop-specific interactions. The pentanucleotide loops of λ and P22 adopt GNRA-type tetraloop structures with exclusion of 1 of the 5 nt. The ϕ 21 *boxB* hexanucleotide loop contains a U-turn that is structurally related to the GNRA tetraloop (Jucker and Pardi 1995), a mismatched C10:C15 base pair, and stacking of C14, giving the ϕ 21 complex a structure similar in shape and exposed surface groups to the λ N peptide-*boxB* complex. Comparison of the ϕ 21 complex to those of λ and P22 reveals how phage type-specificity is maintained in the presence of conserved host factor interactions, through similar peptide-RNA stem binding motifs coupled with loop-specific interactions that yield similar RNA loop structures.

RESULTS

Minimal peptide/RNA complex

The N-*nut* complexes from λ and P22 were minimized to short N peptides containing the amino-terminal basic domain and short RNA hairpins corresponding to *boxB* (Tan and Frankel 1995). As a starting point for defining a similar minimal system for ϕ 21, we began with the 41 amino-terminal residues of ϕ 21 N protein and a 49-nt *nut* site RNA. The affinity of this interaction was determined using a polyacrylamide coelectrophoresis (PACE) assay (Cilley and Williamson 1999) with a $K_{d,app}$ of 135 nM. The ϕ 21 *nut* RNA was minimized to just the *boxB* hairpin, and this minimal 24-nt hairpin binds to the 41-mer peptide with the same affinity, 82 nM (Fig. 2A).

To identify the minimal N ^{ϕ 21} peptide, a series of N-terminal and C-terminal deletion peptides were prepared (Fig. 2B) and assayed. The resulting 22-amino-acid peptide spans the basic region homologous to λ and P22 bound to the *boxB* RNA hairpin and has a $K_{d,app}$ of 200 nM, less than twofold down from the ϕ 21 N(1-41)-*nut* RNA complex.

The three N-*nut* complexes exhibit phage type specificity in vivo. Each N protein has a strong preference for its cognate RNA, and it is the amino terminus of the N protein that confers this type specificity (Lazinski et al. 1989). We determined the complete matrix of in vitro binding affinities of each N peptide for each *boxB* RNA using the PACE assay (Fig. 3; Table 1). Binding affinity in vitro reflects type specificity, consistent with the in vivo experiments. The three cognate N peptide-*boxB* RNA complexes have a 20-fold range in binding affinities. The individual peptides and RNAs demonstrate a wide range of binding specificity for cognate versus noncognate complexes. The N ^{λ} peptide and

FIGURE 2. *boxB* RNA and N peptide minimization for NMR structural studies. (A) The full *nut* site was truncated and tested for binding against the full-length 41-amino-acid N ^{$\phi 21$} peptide. The binding affinities of the RNAs were directly compared against one another in a single PACE gel. (B) Deletions were made from both ends of the 41-amino-acid peptide. The relative binding affinity of all of the peptides for *boxB* ^{$\phi 21$} were compared in a single PACE gel in which each peptide was cast in the gel at 300 nM concentration. The migration distance of *boxB* ^{$\phi 21$} RNA in each lane was compared to that of the N ^{$\phi 21$} 41-amino-acid peptide to evaluate loss of binding affinity.

the *boxB* ^{λ} RNA each show about a 300-fold range in binding affinities (Fig. 3A). In contrast, the N^{P22} peptide and *boxB*^{P22} RNA are the least discriminating, showing at most a fourfold preference for cognate sequences (Fig. 3B), whereas N ^{$\phi 21$} peptide and *boxB* ^{$\phi 21$} RNA have a more intermediate 15-fold specificity (Fig. 3C).

NMR structure determination

The solution structure of the $\phi 21$ N peptide–*boxB* RNA complex was solved with standard heteronuclear NMR experiments, using both ¹⁵N and ¹³C, ¹⁵N-labeled peptide and RNA (see Material and Methods; Table 2). Complex formation was monitored using ¹⁵N- or ¹³C-HSQC experiments. The ¹⁵N-HSQC spectra contains a single set of peaks for a 1:1 complex of ¹⁵N-peptide and unlabeled *boxB* RNA, distinct from the set of peaks for the ¹⁵N-peptide only (Fig. 4A). Sequential and unambiguous assignment of all of the peptide ¹H, ¹³C, and ¹⁵N backbone and side chain chemical shifts was achieved using triple resonance through-bond correlation experiments (Table 2). The resonances of the backbone amides were well dispersed in both the ¹H and ¹⁵N dimensions, and there were few problems with spectral overlap in these experiments. The *boxB* ribose spin systems were assigned using three-dimensional HCCH-COSY, HCCH-TOCSY, and HCCH-COSY-TOCSY experiments

(Hu et al. 1998). These through-bond experiments were combined with sequential assignment of the RNA from two-dimensional (2D) NOESY experiments in both H₂O and D₂O to allow for the complete assignment of the nucleotide spin systems. A pair of 2D HCN spectra were collected to directly establish ribose-to-base connectivities (Table 2). These experiments correlate a base N1/N9 to the ribose H1' or the base H6/H8 to the N1/N9. These connectivities served as a check on the sequential assignments from NOESY analysis. Finally, an HNN-COSY experiment (Dingley and Grzesiek 1998) provided direct evidence for the base pairing in the stem.

Distance restraints for structure calculations were derived from a variety of NOESY experiments (Table 2). Use of

FIGURE 3. Binding data from PACE experiments for each of the three $\phi 21$ N peptides with the three *boxB* RNAs. The N ^{λ} (2–19), N^{P22}(2–30), and N ^{$\phi 21$} (1–41) experiments are graphed in A, B, and C, respectively. The *boxB* ^{λ} , *boxB*^{P22}, and *boxB* ^{$\phi 21$} RNAs in all three graphs are represented by an open circle, a triangle, and a square, respectively. The y-axis, Θ , is the normalized fraction bound. The error bars are estimates from the χ^2 of the least-squares fit to the data. All of these PACE binding assays were run at 25°C, with no salt and no detergent.

TABLE 1. The N peptides and *boxB* RNAs discriminate among noncognate partners.

	N ^A (2–19)	N ^{P22} (2–30)	N ^{Φ21} (1–41)
λ <i>boxB</i> R	4 ± 1	79 ± 18	990 ± 309
P22 <i>boxB</i> L	132 ± 13	55 ± 13	216 ± 66
Φ21 <i>boxB</i> L	1225 ± 87	139 ± 30	82 ± 15

All values are $K_{d,app}$ (nM) determined using the PACE assay at 25°C, no salt and no detergent.

labeled peptide and RNA samples allowed for the determination of a number of NOEs (Table 3, Panel A), often with the same NOE present in both peptide-only and RNA-only spectra (Fig. 4B). The N peptide showed sequential NOE connectivities in ¹⁵N-NOESY spectra typical of an α-helical structure. Similarly, the *boxB* RNA had intra- and internucleotide NOE patterns indicative of A-form helical regions of a stem-loop structure. Peptide backbone torsion angle restraints and RNA ribose sugar pucker restraints were determined from HNHA and COSY experiments, re-

spectively (Tables 2, 3, Panel A). A schematic of the RNA–RNA and peptide–RNA NOEs shows the density of distance restraints, especially in the loop region (Fig. 5A). The core, ordered region of the peptide–RNA complex is evident in the superposition of the 14 low energy structures (Fig. 5B). The average structure of this low energy ensemble was minimized to obtain the final solution structure (Fig. 6A,B; Table 3, Panel B).

Peptide structure in complex

The N^{Φ21} peptide in complex with the *boxB* RNA forms an α-helix spanning residues Ala13–Ala26, with the remaining residues in an extended conformation. The measured φ and ψ angles for the amino acids Ala13–Ala26 fall within the α-helical regions of the Ramachandran plot (Morris et al. 1992). The conserved residues Ala13, Arg16, Arg20, and Arg21 of the N peptide family do not appear to lie on a single face of the α-helix, but they are all oriented towards the binding surface of the *boxB* RNA (Fig. 6A). The slight curvature of the helical axis does not disrupt the helical connectivities. No apparent hydrogen bonds between side

TABLE 2. Data acquisition parameters for N^{Φ21} peptide–*boxB* RNA heteronuclear NMR experiments.

Experiment	F3	F2	F1	T _{mix}	Reference
N^{Φ21} Peptide					
¹⁵ N-HSQC	1024, 5406, ¹ H	256, 578, ¹⁵ N	—	—	—
¹³ C-HSQC	1024, 6006, ¹ H	2048, 48,052, ¹³ C	—	—	—
HNCO	1024, 7000, ¹ H	128, 1800, ¹³ C	64, 2000, ¹⁵ N	—	Muhandiram and Kay 1994
HNCA	1024, 7000, ¹ H	128, 7000, ¹³ C	64, 1600, ¹⁵ N	—	Ikura and Bax 1992
HCCH-TOCSY	1024, 5200, ¹ H	168, 2800, ¹ H	96, 3800, ¹³ C	—	Clore et al. 1990
HNHA	2048, 7900, ¹ H	128, 5800, ¹ H	48, 1600, ¹⁵ N	—	Vuister and Bax 1993
CBCA(CO)NH	512, 4000, ¹ H	128, 10,000, ¹³ C	60, 1250, ¹⁵ N	—	Grzesiek and Bax 1992
HNCACB	512, 4808, ¹ H	100, 10,000, ¹³ C	60, 1250, ¹⁵ N	—	Muhandiram and Kay 1994
HBHA(CO)NH	1024, 4808, ¹ H	198, 4041, ¹ H	58, 1250, ¹⁵ N	—	Grzesiek and Bax 1992
H(C)(CO)NH	1024, 4808, ¹ H	140, 2846, ¹ H	60, 1250, ¹⁵ N	—	Montelione et al. 1992
(H)C(CO)NH	1024, 4808, ¹ H	120, 10,000, ¹³ C	60, 1250, ¹⁵ N	—	Logan et al. 1992
¹⁵ N-NOESY-HSQC	1024, 6410, ¹ H	256, 8333, ¹ H	78, 1524, ¹⁵ N	50, 145, 250 ms	Mori et al. 1995
¹³ C, ¹⁵ N-NOESY-HSQC (CN-NOESY)	1024, 6006, ¹ H	192, 6006, ¹ H	24, 3850, ¹³ C	150 ms	Pascal et al. 1994
<i>boxB</i>^{Φ21} RNA					
NOESY-watagate	2048, 12,500, ¹ H	1024, 12,500, ¹ H	—	50, 145, 250 ms	—
¹⁵ N-HSQC	2048, 8802, ¹ H	256, 422, ¹⁵ N	—	—	—
¹³ C-HSQC	1024, 7180, ¹ H	512, 5405, ¹³ C	—	—	—
dqf-COSY	768, 4805, ¹ H	1024, 4805, ¹ H	—	—	Rance et al. 1983
HCCH-COSY	512, 3623, ¹ H	200, 2381, ¹ H	48, 6667, ¹³ C	—	Clore et al. 1990
HCCH-TOCSY	512, 3623, ¹ H	200, 2381, ¹ H	116, 3333, ¹³ C	—	Clore et al. 1990
HCCH-COSY-TOCSY	256, 1800, ¹ H	128, 1800, ¹ H	60, 5300, ¹³ C	—	Hu et al. 1998
¹⁵ N-NOESY-HSQC	1024, 9615, ¹ H	256, 6024, ¹ H	100, 333, ¹⁵ N	200 ms	Mori et al. 1995
¹³ C-NOESY-HSQC	1024, 7184, ¹ H	256, 5435, ¹ H	128, 5435, ¹³ C	250 ms	Clore et al. 1990
HNN-COSY	2048, 14,000, ¹ H	64, 6087, ¹⁵ N	—	—	Dingley and Grzesiek 1998
2D-HCN	1024, 640, ¹ H	256, 2800, ¹⁵ N	—	—	Sklenar et al. 1993
2D-filtered/edited NOESY	1024, 12,820, ¹ H	1024, 7142, ¹ H	—	200 ms	Ikura and Bax 1992
2D-filtered/edited NOESY	1024, 12,820, ¹ H	954, 7142, ¹ H	—	200 ms	Zwahlen et al. 1998
					Ikura and Bax 1992
					Zwahlen et al. 1998

Data acquisition parameters including number of points, sweep width, and nuclei are listed for each dimension F1, F2, and F3.

FIGURE 4. Peptide and RNA NMR data. All experiments were carried out at 25°C. (A) ^{15}N -HSQC on labeled peptide both free (black) and bound (red) to unlabeled *boxB* RNA. Peaks are labeled with their amino acid and sequence number for the backbone amide (black), arginine epsilon (blue), and arginine eta (green) protons. The epsilon and eta protons are folded in the ^{15}N dimension by 32.9 ppm. (B) Identification of intermolecular NOEs. An ω_1 - ^{13}C -Filtered, ω_2 - ^{13}C -Edited NOESY-HSQC spectrum recorded on an unlabeled *boxB* RNA in complex with a fully ^{13}C , ^{15}N -labeled N peptide. Strong crosspeaks between the peptide (blue) and RNA (red) are noted.

chains can be resolved in the N peptide structure. In the α -helical portion of the peptide, all side chains with potential hydrogen bond donor moieties are oriented on one face of the helix, whereas the potential hydrogen bond acceptor side chains lie mostly on the opposite face. All of the peptide side chains that gave rise to NOE distance restraints (Fig. 5A) lie along the 5'-side of the major groove of the upper 4 bp of the stem and behind the loop of *boxB* RNA (Fig. 6B).

The nonhelical regions of the N peptide, Glu8 to Thr12 and Glu27 to Arg29, are randomly oriented away from the *boxB* RNA. These residues had few interresidue NOEs and no NOEs to the *boxB* RNA. Deletion analysis showed that a peptide truncated at Thr12, just before the conserved Ala13, bound the *boxB* RNA with less affinity than a peptide truncated at Glu8 (Fig. 2B). The first seven amino acids of full-length N $^{\phi 21}$ contain a number of hydrophobic residues, including two valines, an isoleucine, and a tryptophan (Fig. 1C). Although not important for RNA binding, these residues may play a role in subsequent N protein folding, or contribute to protein–protein interactions in the complete antitermination complex.

boxB structure in complex

The $\phi 21$ *boxB* RNA in complex with the N peptide adopts a stem-loop structure (Fig. 6A). The stem of the *boxB* RNA contains seven Watson–Crick base pairs and two canonical wobble U:G base pairs, U3:G22 and U9:G16. The stem is

essentially A-form helical, with no significant perturbations of ribose sugar puckers or base glycosidic torsion angles.

In the hexanucleotide loop of *boxB* RNA (CUAACC), the four bases on the 3' end stack upon the 3' strand of the helical stem. Cytidine-15, located at the 3' end of the hexanucleotide loop, stacks on G16, the 3' base in the U9:G16 base pair at the top of the stem. This stacking in the hexanucleotide loop continues upward with nucleotides C14, A13, and A12, extending the topology of an A-form helix up to the tip of the loop. The U11–A12–A13 U-turn facilitates reversal of the direction of the phosphate backbone. Although the imino proton of U11 is not observed in the NMR experiments, 11 of the 14 structures show hydrogen bonds between this imino and the phosphate of C14, a hallmark of the U-turn motif. Similarly, a hydrogen bond between the 2' hydroxyl of U11 and the N7 of A13 is observed in 13 of the 14 low energy structures. In 12 of the 14 low energy structures, the N4-amino group of C15 can hydrogen bond to either the N3 or O2 of C10. The C10:C15 pair stacks on the U9:G16 base pair. This C10:C15 pairing can only be inferred indirectly because there are no observed resonances from any of the N4-amino protons, indicating the lack of a stable hydrogen bond.

Electrostatic and hydrogen bond interactions

One-third of all residues in the N $^{\phi 21}$ peptide are either lysine or arginine. The N $^{\phi 21}$ peptide belongs to a loose class

TABLE 3. Summary of restraints used for molecular modeling and the resulting structural statistics for the N^{φ21} peptide–*boxB* RNA complex.

Panel A	# of Restraints	Restraints/Residue
Distance		
Peptide intraresidue	158	7.2
Peptide interresidue	192	8.7
RNA Hydrogen bonds	48	2.0
RNA intraresidue	286	11.9
RNA interresidue	328	13.7
Peptide–RNA	52	1.1
Total distance	1016	22.1
Torsion		
Glycosidic bond	18	0.9
Sugar pucker	120	5.0
Peptide ϕ	15	0.7
Peptide ψ	14	0.6
Total torsion	167	3.6
Total all	1183	25.7
Panel B	[SA]	$\langle SA \rangle_r$
NOE violations		
R.m.s.d. of violations	0.002 ± 0.019	0.002
Number > 0.1 Å	2.60 ± 1.82	4
Maximum violation	0.41 ± 0.25	0.77
Torsion violations		
R.m.s.d. of violations	0.028 ± 0.013	0.165
Number > 0.2°	4.86 ± 1.56	7
Maximum violation (°)	1.62 ± 0.68	5.82
R.m.s.d. from ideal covalent geometry		
Bond length (Å)	0.0114 ± 0.0001	0.0115
Bond angle (°)	2.25 ± 0.04	2.34
R.m.s.d. (Å) of [SA] from $\langle SA \rangle$		
All heavy	2.77 ± 0.87	
Core ^a heavy	1.16 ± 0.20	

[SA] is the ensemble of 14 lowest energy structures.

 $\langle SA \rangle$ is the average structure of obtained by averaging the coordinates of the [SA]. $\langle SA \rangle_r$ is the $\langle SA \rangle$ that has been restrained energy minimized.^aCore refers to the ordered region of the peptide–RNA complex defined as amino acids 13–25 and nt 7–18.

of protein sequences containing the “arginine-rich motif” (Weiss and Narayana 1998). Members of this “class” have ultimately shown little sequence or structural-specific homology, but all use arginine side chains in binding and recognition of the RNA major groove. In the N^{φ21} peptide–*boxB* RNA complex, all but one of the five arginines and one of the three lysines make extensive contacts with the phosphate backbone (Fig. 7A). The guanidinium groups of the conserved arginines, Arg16, Arg20, and Arg21, are within 5 Å of the phosphate backbone on the 5′-ascending stem of the *boxB* RNA. The guanidinium group of Arg28 is more distant from the phosphate of A12, but could still contribute a favorable electrostatic interaction in the complex. The amino group of Lys18 is within 5 Å of the phosphate group of C14 on the opposite side of the groove. Analysis of the family of low energy structures indicates that some of these

guanidinium and amino groups may be making stronger ionic interactions, or even hydrogen bonds, to the phosphate backbone. Although not as apparent in the average structure (Fig. 7A), in all 14 structures the guanidinium group of Arg16 is <2.4 Å from the nonbridging phosphate oxygens of C7 and C8, in many cases close to both phosphate groups. Similarly, the amino group of Lys18 is close to the phosphate of C14 in 13 of the 14 structures. In 12 of the 14 structures, the Tyr17-H η is <2.4 Å from the C10 phosphate group. Tyrosine has been shown to have a high propensity for forming hydrogen bonds in protein–RNA complexes (Jones et al. 2001).

No clear sequence-specific base interactions are seen in the peptide–RNA complex. In all 14 of the low energy structures, the amino group of Lys14 penetrates into the major groove and makes close contacts to one or more of the Hoogsteen faces of G16, G17, and G18. The lack of precision in the placement of Lys14 from the NOE data makes it impossible to definitively assign its role in any sequence-specific recognition of the *boxB* RNA. Collectively, these arginine and lysine side chains, together with the contribution of Tyr17, create a unique positive surface on one side of the N^{φ21} peptide α -helix that interacts with the negatively charged RNA phosphodiester backbone and contributes to recognition of *boxB* RNA.

All of the conserved arginines in the N^{φ21} peptide are within electrostatic and hydrogen bonding distances to the *boxB* RNA. From examination of the structure of the complex, we believe the conserved Arg21 plays a role in specific recognition of *boxB* RNA. Arginine-21 rests in the curve of the phosphate backbone from C10 to A13 at the top of the *boxB* loop. The H ϵ and H η hydrogens are all close enough to the phosphate backbone to provide a framework around which the backbone can organize. All of the other arginines that make contacts to the phosphate backbone do so by simply extending out towards a particular phosphate or pair of phosphates without any other contact with the RNA. The ¹⁵N-HSQC of the peptide in the complex shows that the Arg21-H ϵ is shifted downfield in both the ¹H and ¹⁵N dimensions from all of the other Arg-H ϵ resonances, indicative of a unique structural or chemical environment. In addition to its extensive ionic interactions, the aliphatic part of the Arg21 side chain makes van der Waals contacts with the base H5/H6 protons of U11. The ability of arginine to make a variety of polar and nonpolar interactions to accommodate a particular binding surface on an RNA is a common motif (Legault et al. 1998; Weiss and Narayana 1998).

van der Waals contacts

With ~1100 Å² of buried surface area between the N^{φ21} peptide–*boxB* RNA complex, polar and nonpolar interactions play an important role in binding. Beginning with Ala13, with each four-residue turn of the α -helix to Ile25,

FIGURE 5. The NMR derived restraints and resulting low energy structures. (A) Schematic of distance restraints. Internucleotide RNA-RNA restraints are indicated by dashed lines between protons (colored dots). Peptide-RNA restraints (solid arrows) indicate one or more NOEs between the peptide side chains and the RNA ribose (pentagon) or base (rectangle). (B) Superposition of the 14 lowest energy structures of the ϕ 21 N peptide-*boxB* RNA complex, drawn in red and blue, respectively. The ordered regions of both the RNA (C7 to G18) and the peptide (Ala13 to Ile25) are shown and the alignment was calculated using all of the heavy atoms in this region. For clarity, only the backbone atoms of the peptide are shown.

the peptide side chains make van der Waals contact to the 5' bases and riboses of the *boxB* RNA (Fig. 7B). The methyl group of Ala13 lies between C7 and C8, packing against both the ribose and base H5/H6 protons of these nucleotides. This alanine is conserved among the family of N proteins, and mutagenesis of this position in the N ^{λ} system showed that only alanine and serine maintained wild-type binding affinity for the *boxB* RNA (Su et al. 1997). Tyrosine-17 is one turn up the α -helix and lies between U9 and C10, the next two bases up the stem from the Ala13 position. In addition to the hydrogen bond interaction of the Tyr17 hydroxyl group, the aromatic ring lies against the edges of the H5/H6 base protons of U9 and C10, and the ribose of U9. Further along the α -helix, Arg21 is positioned to make van der Waals contacts with the U11 base H5/H6 protons of the *boxB* loop, as discussed above. Finally, Ile25 makes extensive van der Waals contacts with the ribose of A12, at the very top of the *boxB* hairpin (Fig. 7B). Although the interactions of Arg21 with U11 may play a role in supporting the shape of the RNA backbone in the loop, the close contact of Ile25 may support the back side of the *boxB*

loop on the 3' face. Lysine-18 may also play a role in recognition and stabilization of the *boxB* loop, through potentially important van der Waals contacts between the Lys18-H γ and H δ groups and the ribose of A13.

DISCUSSION

Overview of ϕ 21, λ , and P22 N peptide-*boxB* structures

The NMR solution structures of the N peptide-*boxB* RNA complexes for phage λ (Legault et al. 1998; Scharpf et al. 2000) and phage P22 (Cai et al. 1998) have been previously determined. All three peptides bind as α -helices. The N ^{λ} peptide has a pronounced bend in the α -helix of $\sim 120^\circ$ as a result of deviations of the backbone ϕ and ψ angles for Arg11. This bend has been shown to be important for N ^{λ} binding (Tan and Frankel 1995; Su et al. 1997) and anti-termination in vivo (Franklin 1993; Su et al. 1997), and is required for bringing Trp18 ^{λ} into position for stacking onto the *boxB* ^{λ} tetraloop. The N^{P22} peptide α -helix is also bent,

FIGURE 6. (A) Stereo view of the minimized, average structure of the complex. The heavy atoms and phosphate backbone ribbons of the RNA (blue) are shown in complex with the peptide (red). The conserved residues (Ala13, Arg16, Arg20, and Arg21) are highlighted (green). (B) View of minimized, average structure looking down the α -helical axis. RNA bases and backbone phosphates along the groove contacted by the peptide are highlighted (cyan), as are side chains that had NOEs (distance restraints) to the RNA (green).

but not as dramatically as that of N^{Δ} . In contrast, the $N^{\phi 21}$ peptide α -helix is almost linear. This result is consistent with the NMR-derived NOE connectivities and backbone torsion angles. In all three complexes, the conserved alanine and conserved arginines play important roles in binding their cognate *boxB* RNAs, though not always identical in each structure. In addition to the conserved residues, each peptide has a unique subset of residues that participate in the peptide–RNA interaction.

All three *boxB* RNAs form hairpin structures, and the stems of each are essentially A-form helical RNA. The se-

quences of the pentanucleotide loops of λ (GAAGA) and P22 (GACAA) are similar, and, in fact, both the λ and P22 *boxB* loops adopt a fold that is almost identical to a GAAA-type GNRA tetraloop (Cai et al. 1998; Legault et al. 1998). In each case, one of the five nucleotides in the loop is excluded from the tetraloop fold. The *boxB* $^{\lambda}$ loop excludes the fourth nucleotide, G, whereas in P22 it is the third nucleotide, C. In a GNRA tetraloop, the first G and last A form a sheared G–A base pair (type XI; Saenger 1984) resulting in a severe change in direction of the phosphate backbone between the first G and the second nucleotide (N) in the loop. The last three nucleotides in the GNRA sequence stack sequentially on the 3' stem below the loop. The structure of a GAAA tetraloop has been solved by NMR (Heus and Pardi 1991) and X-ray crystallography (Scott et al. 1995; Cate et al. 1996). The r.m.s.d. values of superpositions of the λ and P22 GAAA tetraloop folds with GAAA tetraloops from crystal structures were 1.4 and 0.8 Å, respectively. The N^{Δ} peptide makes no contacts with the bulged G (fourth nucleotide in the loop). It was shown that mutation of the fourth position had no effect on N binding, and mutations that abolish the “GNRA” sequence motif diminish N^{Δ} binding (Doelling and Franklin 1989; Cilley and Williamson 1997). The bulged C (third nucleotide in the loop) in the *boxB* P22 of the structure makes extensive hydrophobic interactions with the peptide (Cai et al. 1998). In the $\phi 21$ complex, only Arg21 $^{\phi 21}$ and Ile25 $^{\phi 21}$ have any significant contacts with the loop. The side chain of Arg21 $^{\phi 21}$ binds in the sharp turn in the phosphate backbone of nucleotides C10, U11, A12,

and A13 (Fig. 7A), and makes hydrophobic contact with U11 (Fig. 7B). Just above Arg21 $^{\phi 21}$, Ile25 $^{\phi 21}$ packs against the ribose of A12 at the top of the loop. The $\phi 21$ peptide has few interactions with nucleotide bases in the *boxB* $^{\phi 21}$ RNA loop, in contrast to the base stacking and extensive base hydrophobic packing seen in the λ and P22 structures.

Stem binding motif

The binding of $N^{\phi 21}$ peptide to its cognate *boxB* hairpin has a bipartite character where the residues in the lower half of

FIGURE 7. A Connolly surface over the *boxB* RNA helps illustrate the close contacts within the peptide–RNA complex. In both panels, the RNA phosphate and oxygen atoms are colored white with the rest of the nucleotide atoms in blue. The peptide backbone is shown as a red ribbon. Peptide side chains are labeled using the single-letter for the amino acid and peptide sequence number. The RNA phosphates are numbered along the phosphate backbone. (A) Electrostatic and hydrogen bond interactions with the phosphate backbone. Ball and stick representations of amino acid side chains with potential phosphate interactions. (B) van der Waals contacts in the complex. Yellow CPK representations are used for the side chains that pack against the *boxB* RNA.

the α -helix interact with the *boxB* stem, and the upper part of the α -helix interacts with the *boxB* loop. A similar division can be made in the N^{λ} and N^{P22} peptide–*boxB* complexes. The bend in the N^{λ} peptide α -helix occurs at Arg11 $^{\lambda}$ (Legault et al. 1998). The slight bend in the N^{P22} peptide α -helix is centered between Arg6 P22 and Glu8 P22 (Cai et al. 1998). All three N peptides interact with their cognate *boxB* stems in these complexes in very similar ways (Fig. 8). Interestingly, a sequence alignment of the three N peptides shows that there exists the greatest degree of homology in the first half of the amino acid sequences. Between the conserved alanine and the last of the three conserved arginines, a total of nine residues, there is only one position that is not the same in at least two of the three sequences (Fig. 8). The high degree of sequence homology, combined with similar α -helical secondary structure results in the three N peptides being almost identical in their binding and recognition of the *boxB* stem, with a pairwise r.m.s.d. of important side chains and nucleotides of 1.41 ± 0.22 (detailed in the Fig. 8 legend).

The conserved alanine, Ala13 $^{\phi 21}$, packs against two 5'-nucleotides (C7 and C8) that are in the same relative position in the *boxB* stems of all three phages (cyan coloring in Fig. 8). This is a 5'C-G step in λ and ϕ 21, and 5'G-C in the P22. Mutation of the λ 5'C-G step to 5'G-C only diminished in vivo antitermination activity approximately sixfold (Chattopadhyay et al. 1995). This would imply that al-

though the interaction of this conserved alanine with the *boxB* stem at this position contributes to N binding, it is more likely a nonspecific than sequence-specific interaction.

The first of the three conserved arginines, Arg16 $^{\phi 21}$, makes electrostatic interactions with the phosphate backbone (blue in Fig. 8). Like the conserved alanine, these interactions occur in the same position along the *boxB* stems. In both the λ and ϕ 21 complexes, the guanidino groups of this arginine make close contact with the phosphate backbone. In the P22 structure, as reported (Cai et al. 1998), this same arginine is not as strongly oriented towards the phosphate backbone. An Arg6 $^{\lambda}$ $\rightarrow$ Ala mutation at this position in N^{λ} reduces binding by >20-fold (Su et al. 1997) and abolishes antitermination activity in vivo (Franklin 1993).

The next residue, Tyr17 $^{\phi 21}$, is instead an arginine in both λ and P22. However, similar contacts are observed in each structure with the pair of 5'-nucleotides (U9 and C10) above the alanine step. The Tyr17 $^{\phi 21}$, Arg7 $^{\lambda}$, and Arg6 P22 all

make hydrogen bond or electrostatic interactions with the phosphate backbone, and all three have van der Waals contacts with the bases and ribose sugars (green in Fig. 8). The arginines in λ and P22 at this position can also make additional electrostatic interactions through the H ϵ hydrogen that Tyr17 $^{\phi 21}$, with its single polar group, is unable to duplicate.

Finally, the second conserved arginine, Arg20 $^{\phi 21}$, makes electrostatic contacts with the phosphate backbone in the same region in the three structures. In all three structures, this arginine appears to have a weaker, more distant interaction than any of the other arginines. In only 10 of the 14 low energy structures, the Arg20 $^{\phi 21}$ hydrogen bonds to the phosphate of C8 or U9. It is important to note that alanine substitutions at this position in N^{λ} had deleterious effects on binding and antitermination activity (Franklin 1993).

These striking similarities end as each of the N peptides transitions from binding the stem to binding the loop. The three N peptides differ in their recognition of and interactions with their cognate *boxB* loops. It was proposed that the more conserved parts of the three N peptides would have similar modes of binding the *boxB* RNA stem (Cai et al. 1998; Legault et al. 1998). This hypothesis is strengthened by the addition of the $N^{\phi 21}$ peptide–*boxB* structure. The amino acid sequences of λ and P22 in the stem binding part of the peptide are phylogenetically conserved. $N^{\phi 21}$ helps to identify the conserved position and types of interactions that are important for *boxB* stem binding.

FIGURE 8. Two views of the superposition of the *boxB* stem-peptide recognition regions from all three phage. An amino acid sequence alignment of the three N peptides is at the bottom. In the molecular models, the three *boxB* RNA hairpins (blue ribbon for the phosphate backbone and purple spheres for the actual phosphate atoms) and the three peptide backbones (shown as colored ribbons) have been superimposed using the phosphate backbone and heavy atoms of the cyan and green colored nucleotides along with the backbone and side chain heavy atoms from the conserved alanine (cyan) and arginines (light blue with violet N η spheres). The pairwise r.m.s.d. is 1.41 ± 0.22 angstroms. The coloring of the side chains and bases in the molecular models corresponds to the coloring in the sequence alignment at bottom. The nucleotides on the 3' half of the *boxB* stem have been removed for greater clarity.

Structural mimicry within the *boxB* loops

Binding interactions between the N peptides and *boxB* loops diverges sharply from the structural homology evident in the peptide-stem interactions (Fig. 8). The third conserved arginine in all three complexes binds the *boxB* loop; there is not the same degree of overlap seen for the side-chains-involved stem binding part of the peptide (highlighted in red, orange, and yellow in Fig. 8). The pronounced bend in the N $^{\lambda}$ α -helix accommodates the stacking of Trp18 $^{\lambda}$ on top of the AAA stack of the tetraloop. This interaction is a critical determinant in λ recognition because neither of the other N peptides has an aromatic residue near this position in its primary sequence, and mutation of this position diminishes binding of the N $^{\lambda}$ peptide (Su et al. 1997) and abolishes antitermination in vivo (Franklin 1993). The majority of residues between Arg11 $^{\lambda}$ and Lys19 $^{\lambda}$ are involved in electrostatic and van der Waals interactions with the GAAA tetraloop formed by *boxB $^{\lambda}$. The N P22 peptide interacts with its cognate *boxB* through extensive van der Waals interactions with a cytidine nucleotide*

that is excluded from the GAAA tetraloop structure of *boxB* P22 (Cai et al. 1998). The N P22 has far fewer interactions with the tetraloop structure of its *boxB* than does N $^{\lambda}$ peptide. Despite the differences in peptide-loop interactions, the N P22 and N $^{\lambda}$ peptide-*boxB* complexes have essentially the same GAAA tetraloop structure, with the bases of the stacked adenosines on the opposite face from the N peptides.

The structure determined in this work shows that the *boxB* $^{\phi 21}$ loop, in complex with the N $^{\phi 21}$ peptide, does have structural homology with the GNRA-type tetraloops of *boxB $^{\lambda}$ and *boxB* P22 . The GNRA-like fold is mimicked in *boxB* $^{\phi 21}$ by formation of a U-turn and stacking of the last four nucleotides in the loop (AACC) onto the 3' stem, similar to the AAA stacking in tetraloops of λ and P22. The degree of structural similarity is evident in a superposition of the stacked AACC of *boxB* $^{\phi 21}$ and the stacked Trp18:AAA of *boxB $^{\lambda}$ (Fig. 9). It is striking that the Trp18 $^{\lambda}$ amino acid occupies a similar position to the A12 $^{\phi 21}$ nucleotide in the *boxB* $^{\phi 21}$ complex and has a similar structural geometry. The face of the two loops that would be presented to NusA or other bacterial host factors are very similar in both the geometry and the constellation of hydrogen bond donors and acceptors. This is consistent with the hypothesis that the N-*boxB* complex interacts with some other factor or factors in the antitermination complex. This degree of structural similarity is a surprising coincidence given the disparate RNA loop sequences and the very different ways in which they bind to their cognate N proteins, as discussed above. Despite having very different sets of interactions with their *boxB* loops, these three peptides ultimately organize the 3'-half of their *boxB* hairpins to be structurally and chemically similar.**

The three peptide-RNA structures help explain the binding specificity and affinity of the N peptide-*boxB* RNAs (Table 1). The λ and P22 *boxB* RNAs contain the requisite sequences for forming a GNRA tetraloop, each flipping out one base in the process. The bulged C (third nucleotide) in the *boxB* P22 may have steric clashes with the N $^{\lambda}$ peptide stacking of the critical Trp18 $^{\lambda}$ onto the tetraloop, a critical binding interaction. In contrast, the *boxB* $^{\lambda}$ could substitute for the *boxB* P22 by flipping out the third nucleotide (A instead of C) in the context of the extensive hydrophobic packing of the N P22 peptide. Two factors could contribute to the weak binding of N $^{\lambda}$ peptide to *boxB* $^{\phi 21}$ RNA. First, the top of the stacked nucleotides in *boxB* $^{\phi 21}$ is further away from the stem and thus not as accessible for Trp18 $^{\lambda}$ stacking. Second, A12 in *boxB* $^{\phi 21}$ occupies the position where Trp18 $^{\lambda}$ would stack, resulting in steric clash.

N-*boxB* in a biological context

The *boxB* hairpin does more than simply anchor N protein to the RNA transcript and help deliver the N protein to RNA polymerase. Although the mechanism of N-mediated

FIGURE 9. Two views of the superposition of the λ and ϕ 21 *boxB* loops. The peptides are shown as ribbons with only the λ -Trp18 side chain drawn in. The *boxB* loops of λ and ϕ 21 (cyan and blue, respectively) are shown using heavy atoms and a phosphate backbone ribbon. The λ and ϕ 21 structures were superimposed using the heavy atoms of ϕ 21 A12, A13, C14 and C15 with the λ Trp18, A9, A10, and A12 with a resultant r.m.s.d. of 1.9 Å (106 atoms). Potential hydrogen bond donors and acceptors in both structures are highlighted as spheres (green and purple, respectively).

RNA binding by NusA is not known, there is evidence that NusA interacts with both the 3'-half of the *boxB* hairpin (Mah et al. 2000) and directly with N protein ($K_d \approx 70$ nM), and that the NusA-binding and RNA-binding domains of N are distinct (Van Gilst and von Hippel 1997). A mutation of guanosine to cytidine at the fourth position in the *boxB* $^\lambda$ loop or substitution of the GAAGA with a GAAA tetraloop sequence has no effect on N^λ binding in vitro, but abolishes binding of NusA to an N^λ –*boxB* complex (Mogridge et al. 1995; Legault et al. 1998). Similar mutations abolish antitermination activity in vivo and with a minimal N, NusA, RNAP antitermination system, and indicate that it is the 3' half of the *boxB* stem that is being recognized (Chattopadhyay et al. 1995). This evidence suggests that *boxB* interacts with other elements of the transcription antitermination complex. NusA has an S1 homology region, which is found in proteins that can nonspecifically bind to RNA (Bycroft et al. 1997). This S1 domain is activated to bind RNA by both N protein and the α subunit of *E. coli* RNA polymerase (Mah et al. 2000). Though there has been no direct evidence of a NusA–*boxB* interaction in the absence of N protein, NusA is a likely candidate for interaction with the *boxB* hairpin.

Normally, NusA plays a role in pausing of the transcription complex. Interaction of NusA with N and *boxB* RNA reverses that role (Mah et al. 2000), diminishing pausing and increasing processivity of the transcription complex. Evidence suggests that NusA binds termination hairpins on the nascent transcript and allosteric effects the RNA polymerase active site (Toulkhonov et al. 2001). What roles do N and *boxB* play in potentially sequestering this function of NusA? Additionally, it has been shown that NusA cannot

bind the λ N peptide–*boxB* complex if the fourth nucleotide (A or G) is removed or is mutated to a C or U, and supershifts with NusA are sensitive to the identity at this position. In the λ *boxB*–N structure, this nucleotide is extruded out of the tetraloop and has no apparent interactions with either the *boxB* RNA or the N peptide. Presumably this bulged nucleotide is available for interactions with NusA or other factors. Neither the ϕ 21 nor the P22 *boxBs* have a purine base bulged out of the side of the loop. How is this seemingly critical determinant of the N^λ –*boxB* $^\lambda$ –NusA interaction provided for in the other two phage?

Addition of a third member to the family of λ phage *boxB*–N structures has helped clarify general schemes for α -helix:hairpin complex formation. It is clear that these phage N proteins use very similar modes of interaction to bind the

hairpin stem, but diverge significantly in binding and recognition of the hairpin loop, easily discriminating cognate from noncognate. As the ϕ 21 N–*boxB* structure so elegantly shows, these RNA loops and α -helical peptides form complexes using different strategies while still presenting almost identical faces with which to interact with their host.

MATERIALS AND METHODS

Sample preparation

Peptides for these studies were obtained by using a bacterial expression system in which the coding sequence for the peptide was fused to the carboxy terminus of a poly-His-tagged TrpLE leader polypeptide (Schumacher et al. 1996). This system allowed for rapid, high-level expression and purification of both unlabeled and uniformly isotopically enriched (^{15}N , ^{13}C) peptides. Fusion peptides were purified from inclusion bodies from *E. coli* (JM109) and subsequently subjected to cyanogen bromide cleavage in 70% formic acid to remove the TrpLE leader sequence, followed by reverse phase HPLC purification. The final product was dialyzed extensively against water and lyophilized.

All RNAs were prepared by transcription from synthetic DNA templates using T7 RNA polymerase (Wyatt et al. 1991). The template strand, or bottom strand, of the RNA used for NMR analysis was 5'-GGCTCACCCGGTTAGAGGTGAACCTATAGTGAGTCGTATTA-3'. Use of 4-methyl-indole at the 5'-end of the template DNA causes a dramatic decrease in the amount of N + 1 and higher add-on transcripts (Moran et al. 1996). The optimized transcription conditions used for the $N^{\phi 21}$ *boxB* RNA were: 40 mM Tris-HCl (pH 8.1), 1 mM spermidine, 0.01% Triton X-100, 15 mM DTT, 80 mg/mL PEG-8000, 300 nM top strand, 300 nM template strand, 10 mM NTPs (total), 7.8 mM MgCl_2 , and 0.4 μL

T7 RNA polymerase (optimized amount) per 50 μ L of transcription volume. Transcription reactions were run for 4 h at 37°C, then extracted with phenol to remove the T7 RNA polymerase protein and ethanol precipitated. The crude RNA was purified using denaturing preparative gel electrophoresis, electroeluted using an Elutrap (Schleicher and Schuell), and ethanol precipitated. The purified RNA was extensively dialyzed against ddH₂O, and lyophilized. For the most accurate determination of the RNA concentration, a small sample was hydrolyzed in NaOH (pH 12), neutralized with HCl, and the UV absorbance at 260 nm was used to calculate the concentration.

Peptide and RNA stocks were stored at –20°C in a lyophilized state until used. Peptide and RNA samples for NMR were resuspended with 500 μ L of ddH₂O and dialyzed in a microdialysis chamber (BRL) with a 500 MWCO membrane for 12 to 24 h each against 0.5 M NaCl, 10 mM EDTA, then 0.1 mM EDTA, and finally two changes of ddH₂O. Samples were lyophilized and then resuspended in N⁴²¹ NMR buffer: 25 mM D₆(98%)-succinate (pH 6.0; Cambridge Isotopes), 2 mM NaCl, 0.2 mM EDTA (pH 8.0), 0.05 mM Na-azide, 10% D₂O. NMR sample volumes were 600 μ L. Samples were exchanged into D₂O as needed by lyophilization and resuspension in 99.9% D₂O (Cambridge Isotopes). This was repeated three times, and after the third lyophilization, the sample was resuspended in “100%” D₂O to 600 mL and transferred to an NMR sample tube.

NMR assignments

A number of heteronuclear NMR experiments were run to determine peptide and RNA ¹H, ¹⁵N, and ¹³C resonance assignments, as well as generate restraints for molecular modeling (Table 2). These included NOE distance restraints, torsion angle restraints, sugar pucker and glycosidic bond angle restraints, RNA base pairing, and peptide backbone hydrogen bond restraints (Table 3, Panel A).

Molecular modeling

The molecular modeling of the N⁴²¹ peptide–boxB RNA complex was done in three steps. First, a complete intramolecular restraint set was generated for each molecule, and then peptide and RNA structures were generated separately using ab initio simulated annealing (SA) starting from a random extended structure in CNSolve (Brunger et al. 1998). For both the peptide and the RNA, constrained (torsion) dynamics was used at 50,000 K. Though this approach is more computer intensive than restrained (Cartesian) dynamics, it demonstrates a better rate of convergence of the calculated structures (Rice and Brunger 1994). A total of 100 structures each of the peptide and RNA were generated.

In the second step, each of the 20 lowest energy peptide and 20 lowest energy RNA structures were combined in single PDB files, in all 400 possible combinations. The RNA was held at the origin and the peptide was randomly rotated and moved 100 Å away in a random direction from the origin. These 400 possible “complexes” were docked using CNSolve. The objective of the docking was to bring the peptide and RNA together without dramatically perturbing their folded structures from the first round of SA, so the temperatures for the docking were set much lower than in the initial calculations (1000 versus 50,000 K).

The 100 lowest energy docked structures were minimized by two rounds of low-temperature annealing using Sander, a module of AMBER (Pearlman et al. 1995). As with the docking, the temperature was kept low (1000 K). Our experience has been that structural minimization using AMBER yields better and more consistent results for nucleic acid, and protein–nucleic acid complexes than Xplor or Discover. We also wanted to take advantage of the recent improvements in the Generalized Born model for solvation that is incorporated into AMBER. Structure refinement using implicit solvent is computationally expensive, but results in both faster convergence and higher quality structures than in vacuo calculations (Bashford and Case 2000). The 14 lowest energy structures were used for generating an average structure, which was energy minimized using a conjugate gradient. The structural statistics for the ensemble and the minimized, average structure indicate that, at least within the core peptide–RNA-binding region, the solution structure converges well (Table 3, Panel B).

The atomic coordinates for the ensemble of low energy structures, as well as that for the minimized, average structure have been deposited with the Protein Data Bank, and assigned PDB ID code 1NYB.

ACKNOWLEDGMENTS

This work has been supported by grants from the NIH (GM-53320) and the Skaggs Institute for Chemical Biology. The authors gratefully acknowledge the assistance of John Chung of the Buddy Taub NMR center at The Scripps Research Institute.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact.

Received December 3, 2002; accepted February 14, 2003.

REFERENCES

- Auffinger, P. and Westhof, E. 1999. Singly and bifurcated hydrogen-bonded base-pairs in tRNA anticodon hairpins and ribozymes. *J. Mol. Biol.* **292**: 467–83.
- Bashford, D. and Case, D.A. 2000. Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **51**: 129–52.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. 1998. Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**: 905–21.
- Bycroft, M., Hubbard, T.J., Proctor, M., Freund, S.M., and Murzin, A.G. 1997. The solution structure of the S1 RNA binding domain: A member of an ancient nucleic acid-binding fold. *Cell* **88**: 235–42.
- Cai, Z., Gorin, A., Frederick, R., Ye, X., Hu, W., Majumdar, A., Ketani, A., and Patel, D.J. 1998. Solution structure of P22 transcriptional antitermination N peptide–boxB RNA complex. *Nat. Struct. Biol.* **5**: 203–212.
- Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R., and Doudna, J.A. 1996. Crystal structure of a group I ribozyme domain: principles of RNA packing. [see comments]. *Science* **273**: 1678–1685.
- Chattopadhyay, S., Garcia-Mena, J., DeVito, J., Wolska, K., and Das, A. 1995. Bipartite function of a small RNA hairpin in transcription antitermination in bacteriophage λ . *Proc. Natl. Acad. Sci.* **92**: 4061–4065.

- Cilley, C.D. and Williamson, J.R. 1997. Analysis of bacteriophage N protein and peptide binding to boxB RNA using polyacrylamide gel coelectrophoresis (PACE). *RNA* **3**: 57–67.
- . 1999. PACE analysis of RNA–peptide interactions. *Methods Mol. Biol.* **118**: 129–141.
- Clore, G.M., Bax, A., Driscoll, P.C., Wingfield, P.T., and Gronenborn, A.M. 1990. Assignment of the side-chain ^1H and ^{13}C resonances of interleukin-1b using double- and triple-resonance heteronuclear three-dimensional NMR spectroscopy. *Biochemistry* **29**: 8172–8184.
- Conn, G.L., Draper, D.E., Lattman, E.E., and Gittis, A.G. 1999. Crystal structure of a conserved ribosomal protein-RNA complex. *Science* **284**: 1171–1174.
- Das, A. 1993. Control of transcription termination by RNA-binding proteins. In *Annual Review of Biochemistry*, pp. 893–930, Annual Reviews. Palo Alto, CA.
- Dingley, A.J. and Grzesiek, S. 1998. Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide 2JNN couplings. *J. Am. Chem. Soc.* **120**: 8293–8297.
- Doelling, J.H. and Franklin, N.C. 1989. Effects of all single base substitutions in the loop of boxB on antitermination of transcription by bacteriophage λ 's N protein. *Nucleic Acids Res.* **17**: 5565–5577.
- Fountain, M.A., Serra, M.J., Krugh, T.R., and Turner, D.H. 1996. Structural features of a six-nucleotide RNA hairpin loop found in ribosomal RNA. *Biochemistry* **35**: 6539–6548.
- 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.
- . 1985a. Conservation of genome form but not sequence in the transcription antitermination determinants of bacteriophages λ , ϕ 21 and P22. *J. Mol. Biol.* **181**: 75–84.
- . 1985b. “N” transcription antitermination proteins of bacteriophages λ , ϕ 21 and P22. *J. Mol. Biol.* **181**: 85–91.
- Friedman, D.I. and Court, D.L. 1995. Transcription antitermination: The λ paradigm updated. *Mol. Microbiol.* **18**: 191–200.
- Greenblatt, J., Nodwell, J.R., and Mason, S.W. 1993. Transcriptional antitermination. *Nature* **364**: 401–406.
- Grzesiek, S. and Bax, A. 1992. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114**: 6291–6293.
- Gutell, R.R., Cannone, J.J., Konings, D., and Gautheret, D. 2000. Predicting U-turns in ribosomal RNA with comparative sequence analysis. *J. Mol. Biol.* **300**: 791–803.
- Heus, H.A. and Pardi, A. 1991. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **253**: 191–194.
- Horwitz, R.J., Li, J., and Greenblatt, J. 1987. An elongation control particle containing the N gene transcriptional antitermination protein of bacteriophage λ . *Cell* **51**: 631–641.
- Hu, W., Kakalis, L.T., Jiang, L., Jiang, F., Ye, X., and Majumdar, A. 1998. 3D HCCH-COSY-TOCSY experiment for the assignment of ribose and amino acid side chains in ^{13}C labeled RNA and protein. *J. Biomol. NMR* **12**: 559–564.
- Huang, S., Wang, Y.X., and Draper, D.E. 1996. Structure of a hexanucleotide RNA hairpin loop conserved in ribosomal RNAs. *J. Mol. Biol.* **258**: 308–321.
- Ikura, M. and Bax, A. 1992. Isotope-filtered 2D NMR of a protein-peptide complex: Study of a skeletal muscle myosin light chain kinase fragment bound to calmodulin. *J. Am. Chem. Soc.* **114**: 2433–2440.
- Jones, S., Daley, D.T.A., Luscombe, N.M., Berman, H.M., and Thornton, J.M. 2001. Protein-RNA interactions: A structural analysis. *Nucleic Acids Res.* **29**: 943–954.
- Jucker, F.M. and Pardi, A. 1995. GNRA tetraloops make a U-turn. *RNA* **1**: 219–222.
- Lazinski, D., Grzadzinska, E., and Das, A. 1989. Sequence-specific recognition of RNA hairpins by bacteriophage antiterminators requires a conserved arginine-rich motif. *Cell* **59**: 207–218.
- Legault, P., Li, J., Mogridge, J., Kay, L.E., and Greenblatt, J. 1998. NMR structure of the bacteriophage λ N peptide/boxB RNA complex: Recognition of a GNRA fold by an arginine-rich motif. *Cell* **93**: 289–299.
- Logan, T.M., Olejniczak, E.T., Xu, R.X., and Fesik, S.W. 1992. Side chain and backbone assignments in isotopically labeled proteins from two heteronuclear triple resonance experiments. *FEBS Lett.* **314**: 413–418.
- Mah, T.F., Kuznedelov, K., Mushegian, A., Severinov, K., and Greenblatt, J. 2000. The α subunit of *E. coli* RNA polymerase activates RNA binding by NusA. *Genes & Dev.* **14**: 2664–2675.
- Mogridge, J., Mah, T.F., and Greenblatt, J. 1995. A protein–RNA interaction network facilitates the template-independent cooperative assembly on RNA polymerase of a stable antitermination complex containing the λ N protein. *Genes & Dev.* **9**: 2831–2845.
- Montelione, G.T., Lyons, B.A., Emerson, S.D., and Tashiro, M. 1992. An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically enriched proteins. *J. Am. Chem. Soc.* **114**: 10974–10975.
- Moran, S., Ren, R.X., Sheils, C.J., Rumney, St., Kool, E.T. 1996. Non-hydrogen bonding ‘terminator’ nucleosides increase the 3'-end homogeneity of enzymatic RNA and DNA synthesis. *Nucleic Acids Res.* **24**: 2044–2052.
- Mori, S., Abeygunawardana, C., Johnson, M.O.N., and Zijl, P.C.M.v. 1995. Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *J. Magn. Reson. B* **108**: 94–98.
- Morris, A.L., MacArthur, M.W., Hutchinson, E.G., and Thornton, J.M. 1992. Stereochemical quality of protein structure coordinates. *Proteins* **12**: 345–364.
- Muhandiram, D.R. and Kay, L.E. 1994. Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. B* **103**: 203–216.
- Nodwell, J.R. and Greenblatt, J. 1993. Recognition of boxA antiterminator RNA by the *E. coli* antitermination factors NusB and ribosomal protein S10. *Cell* **72**: 261–268.
- Olson, E.R., Tomich, C.S., and 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 λ . *J. Mol. Biol.* **180**: 1053–1063.
- Pascal, S.M., Muhandiram, D.R., Yamazaki, T., Forman-Kay, J.D., and Kay, L.E. 1994. Simultaneous acquisition of ^{15}N - and ^{13}C -edited NOE spectra of proteins dissolved in H_2O . *J. Magn. Reson. B* **103**: 197–201.
- Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.R., Cheatham III, T.E., DeBolt, S., Ferguson, D., Seibel, G., and Kollman, P. 1995. AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules. *Comput. Phys. Commun.* **91**: 1–41.
- Pley, H.W., Flaherty, K.M., and McKay, D.B. 1994. Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**: 68–74.
- Quigley, G.J. and Rich, A. 1976. Structural domains of transfer RNA molecules. *Science* **194**: 796–806.
- Rance, M., Sorensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R., and Wuthrich, K. 1983. Improved spectral resolution in cosy ^1H NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Commun.* **117**: 479–485.
- Rice, L.M. and Brunger, A.T. 1994. Torsion angle dynamics: Reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins* **19**: 277–290.
- Saenger, W. 1984. *Principles of nucleic acid structure* (ed. C.R. Cantor). Springer-Verlag, New York, NY.
- Scharpf, M., Sticht, H., Schweimer, K., Boehm, M., Hoffmann, S., and Rosch, P. 2000. Antitermination in bacteriophage λ . The structure of the N36 peptide-boxB RNA complex. *Eur. J. Biochem.* **267**: 2397–2408.

- Schumacher, T.N., Mayr, L.M., Minor Jr., D.L., Milhollen, M.A., Burgess, M.W., and Kim, P.S. 1996. Identification of D-peptide ligands through mirror-image phage display. *Science* **271**: 1854–1857.
- Scott, W.G., Finch, J.T., and Klug, A. 1995. The crystal structure of an all-RNA hammerhead ribozyme: A proposed mechanism for RNA catalytic cleavage. *Cell* **81**: 991–1002.
- Sklenar, V., Peterson, R.D., Rejante, M.R., and Feigon, J. 1993. Two- and three-dimensional HCN experiments for correlating base and sugar resonances in ^{15}N , ^{13}C -labeled RNA oligonucleotides. *J. Biomol. NMR* **3**: 721–727.
- Su, L., Radek, J.T., Hallenga, K., Hermanto, P., Chan, G., Labeets, L.A., and Weiss, M.A. 1997. RNA recognition by a bent α -helix regulates transcriptional antitermination in phage λ . *Biochemistry* **36**: 12722–12732.
- Szybalski, W., Brown, A.L., Hasan, N., Podhajska, A.J., and Soma-sekhar, G. 1986. Modular structure of the nut and gut antiterminations of transcription. Interactions between control elements of phage λ and construction of novel regulatory circuits. In *RNA polymerase and the regulation of transcription* (ed. W.S. Reznikoff), pp. 381–390. Elsevier Publishing, New York.
- Tan, R. and Frankel, A.D. 1995. Structural variety of arginine-rich RNA-binding peptides. *Proc. Natl. Acad. Sci. USA* **92**: 5282–5286.
- Touloukhonov, I., Artsimovitch, I., and Landick, R. 2001. Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. *Science* **292**: 730–733.
- Van Gilst, M.R. and 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.
- Vuister, G.W. and Bax, A. 1993. Quantitative J correlation: A new approach for measuring homonuclear three-bond $J(\text{H}^{\text{N}}\text{H}^{\alpha})$ coupling constants in ^{15}N -enriched proteins. *J. Am. Chem. Soc.* **115**: 7772–7777.
- Weiss, M.A. and Narayana, N. 1998. RNA recognition by arginine-rich peptide motifs. *Biopolymers* **48**: 167–180.
- Wyatt, J.R., Chastain, M., and Puglisi, J.D. 1991. Synthesis and purification of large amounts of RNA oligonucleotides. *Biotechniques* **11**: 764–769.
- Zhang, H., Fountain, M.A., and Krugh, T.R. 2001. Structural characterization of a six-nucleotide RNA hairpin loop found in *Escherichia coli*, r(UUAAGU). *Biochemistry* **40**: 9879–9886.
- Zwahlen, C., Vincent, S.J., and Kay, L.E. 1998. Analytical description of the effect of adiabatic pulses on IS, I2S, and I3S spin systems. *J. Magn. Reson.* **130**: 169–175.

RNA

A PUBLICATION OF THE RNA SOCIETY

Structural mimicry in the phage [phis]21 N peptide–*boxB* RNA complex

CHRISTOPHER D. CILLEY and JAMES R. WILLIAMSON

RNA 2003 9: 663-676

References

This article cites 58 articles, 12 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/9/6/663.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *RNA* go to:
<http://rnajournal.cshlp.org/subscriptions>