RNA Recognition by a Bent α-Helix Regulates Transcriptional Antitermination in Phage λ†

Leila Su,‡ James T. Radek,‡ Klaas Hallenga,§ Patrick Hermanto,‡ Ging Chan,‡ Laura A. Labeots,‡,|| and Michael A. Weiss*,‡

Departments of Biochemistry & Molecular Biology, Chemistry, and Radiology, The University of Chicago, Chicago, Illinois 60637-5419

Received June 12, 1997; Revised Manuscript Received July 17, 1997⊗

ABSTRACT: A novel RNA recognition motif is characterized in an arginine-rich peptide. The motif, derived from λ transcriptional antitermination protein N, regulates an RNA-directed genetic switch. Its characterization by multidimensional nuclear magnetic resonance (NMR) demonstrates specific RNA-dependent folding of N- and C-terminal recognition helices separated by a central bend. The biological importance of the bent α-helix is demonstrated by mutagenesis: binding is blocked by substitutions in the N peptide or its target (the *boxB* RNA hairpin) associated *in vivo* with loss of transcriptional antitermination activity. Although arginine side chains are essential, the peptide is also anchored to *boxB* by specific nonpolar contacts. An alanine in the N-terminal helix docks in the major groove of the RNA stem whereas a tryptophan in the C-terminal helix stacks against a purine in the RNA loop. At these positions all 19 possible amino acid substitutions have been constructed by peptide synthesis; each impairs binding to *boxB*. The pattern of allowed and disallowed substitutions is in accord with the results of random-cassette mutagenesis *in vivo*. The helix−bend−helix motif rationalizes genetic analysis of N-dependent transcriptional antitermination and extends the structural repertoire of arginine-rich domains observed among mammalian immunodeficiency viruses.

Protein−nucleic acid recognition is mediated by conserved motifs (Harrison, 1991). Whereas protein−DNA recognition is well characterized at high resolution, principles of protein−RNA recognition are less well understood (Draper, 1995). Of particular biological interest are arginine-rich motifs (ARMs;[1] Lazinski *et al.*, 1989), observed among ribosomal proteins, viral coat proteins, and transcription factors, including the Tat and Rev proteins of mammalian immunodeficiency viruses (Tan & Frankel, 1995). Studies of Tat and Rev have revealed that analogous peptide sequences can employ diverse mechanisms of recognition (Puglisi *et al.*, 1995; Battiste *et al.*, 1994, 1996; Ye *et al.*, 1995, 1996). Underlying such diversity is the costabilization of novel RNA and protein structures. The present study focuses on the ARM motif of the λ transcriptional antitermination protein N (Lazinski *et al.*, 1989; Tan & Frankel, 1995; Van Gilst *et al.*, 1997; Cilley & Williamson, 1997).

Expression of delayed-early genes in phage λ requires that RNA polymerase (RNAP) read through ρ-dependent and ρ-independent terminators [Roberts, 1969; for reviews see Das (1993) and Greenblatt *et al.* (1993)]. Antitermination is regulated by the λ N protein (an immediate-early gene product; Franklin, 1985a,b) in concert with host factors (Figure 1A; Olsen *et al.*, 1982; Schauer *et al.*, 1987; Whalen *et al.*, 1988; Mason *et al.*, 1991, 1992; Li *et al.*, 1992; DeVito & Das, 1994). The antitermination activity of N is directed by an RNA enhancer in nascent messages (N-utilization *nut* sites; Salstrom & Szybalski, 1978; de Crombrugghe *et al.*, 1979; Barik *et al.*, 1987; Whalen & Das, 1990; Patterson *et al.*, 1994). The *nut* site consists of a 5′ single-stranded RNA element (designated *boxA*; Olson *et al.*, 1982; Friedman & Olson, 1983), a short single-stranded linker, and a 3′ hairpin (*boxB*). *BoxB* (hairpin in Figure 1B) is the site of N binding (Salstrom & Szybalski, 1978; Nodwell & Greenblatt, 1991; Chattopadhyay *et al.*, 1995a; Mogridge *et al.*, 1995). The RNA-binding domain of N consists of an arginine-rich peptide (ARM residues 1−22; Lazinski *et al.*, 1989; Tan & Frankel, 1995). Mutations in either *boxB* or the N ARM can reduce the efficiency of N-dependent antitermination (Doelling & Franklin, 1989; Franklin, 1993). Peptide models of the N RNA-binding domain have previously been characterized (Figure 1C; Lazinski *et al.*, 1989; Tan & Frankel, 1995; Van Gilst *et al.*, 1997; Cilley & Williamson, 1997).

In this paper the structure of the N ARM is characterized by multidimensional NMR spectroscopy and peptide mutagenesis. Whereas the free peptide is disordered, upon

† This work was supported in part by Grant 2959AR2 from the Council for Tobacco Research (M.A.W.), National Institutes of Health Grant RO1 CA63485-01 (M.A.W.), and the Illinois Chapter of the American Cancer Society (K.H.). M.A.W. is an Established Investigator of the American Heart Association, a Lucille Markey Scholar, and a Bane Scholar at The University of Chicago.

* To whom correspondence should be addressed at The University of Chicago (email: maweiss@midway.uchicago.edu).

‡ Departments of Biochemistry & Molecular Biology and Chemistry.

§ Department of Radiology.

|| Present address: Olson & Hierl, Ltd., Patent Law, 20 North Wacker Drive, Chicago, IL 60606.

⊗ Abstract published in *Advance ACS Abstracts,* September 15, 1997.

[1] Abbreviations: 1D, one dimensional; 2D, two dimensional; ARM, arginine-rich motif; CD, circular dichroism; DNA, deoxyribonucleic acid; GMSA, gel-mobility-shift assay; Fmoc, fluorenylmethoxycarbonyl; HBH, helix−bend−helix; HPLC, high-performance liquid chromatography; HTH, helix−turn−helix; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; RNA, ribonucleic acid; TOCSY, total correlation spectroscopy. Amino acids and RNA bases are designated by standard three- and one-letter codes. The term "bend" in description of induced peptide structure is descriptive only and is not meant to designate a specific interhelical angle.

S0006-2960(97)01408-6 CCC: $14.00

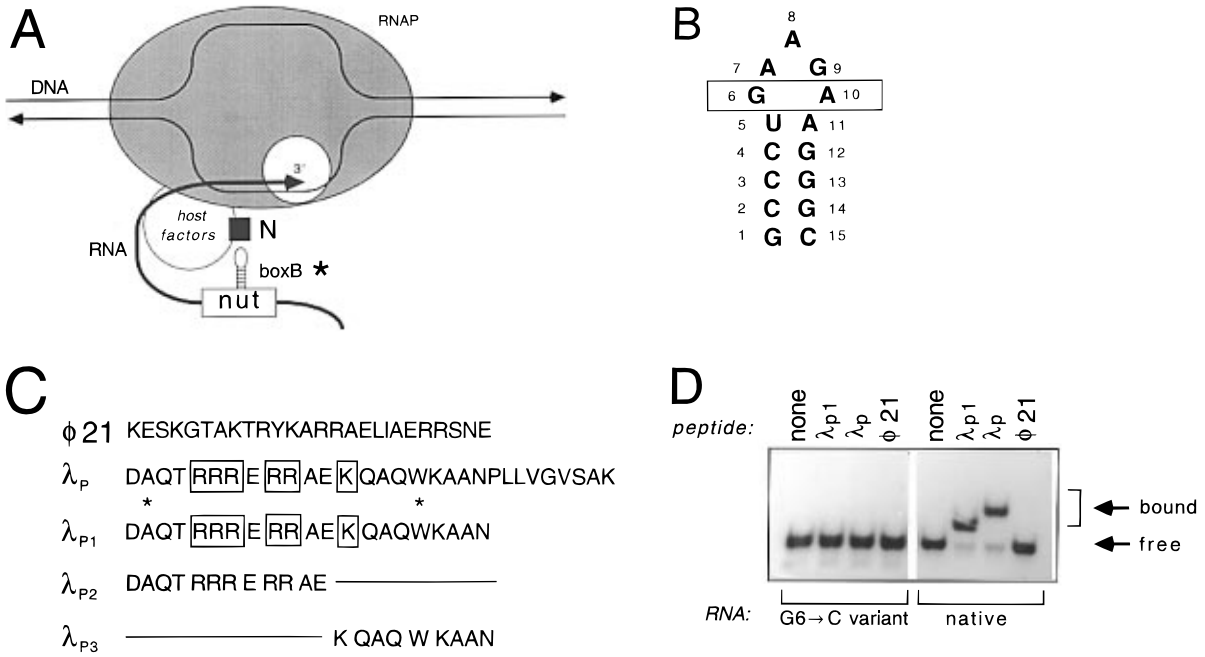


FIGURE 1: Overview of λ N regulatory system. (A) The N protein binds to RNA enhancer elements (*nut* sites; asterisk indicates *boxB* RNA hairpin) and to the RNAP holoenzyme to enable transcriptional elongation past ρ-dependent and intrinsic terminators. Not shown: *nut boxA* RNA motif and interactions of the N complex with host factors NusA, NusB, NusG, and S10. (B) Fifteen-base model of nut_L box B, with numbering scheme as shown. NMR of free RNA verifies stem structure with disordered loop. Box indicates purine−purine base pair (sheared GA base pair; Su *et al.*, 1997). (C) Synthetic N peptides used in the present study are designated by phage of origin, λ or ϕ21. Asterisks (Ala3 and Trp18) and boxed residues in λ_P and λ_{P1} are required for high-affinity RNA binding (see Figure 2). Peptide C-termini are blocked by amidation. Peptides exclude initiator methionine; peptides λ_P, λ_{P1}, and λ_{P2} hence begin with residue 2 as defined by codon number (Franklin, 1985a, 1993). (D) Gel-retardation study demonstrates specific binding of λ_P and λ_{P1} to *boxB* RNA but not to the G6 → C variant site, which is inactive in transcriptional antitermination (Doelling & Franklin, 1989). ϕ21 control peptide binds to neither RNA site. Peptides were made 20 nM in 50 mM KCl and 20 mM Tris−acetate buffer at pH 7.9 and 4 °C.

binding to *boxB* the domain becomes well organized. The RNA-bound structure consists of N- and C-terminal recognition helices separated by a noncanonical bend. This structure differs from that induced in the analogous ARM motifs of Tat and Rev, virulence factors encoded by mammalian immunodeficiency viruses (Puglisi *et al.*, 1995; Battiste *et al.*, 1994, 1996; Ye *et al.*, 1995, 1996). The biological relevance of these results is supported by mutagenesis and its correlation with genetic analysis (Doelling & Franklin, 1989; Franklin, 1993; Van Gilst *et al.*, 1997). In particular, the helix−bend−helix is anchored to the RNA by nonpolar amino acid side chains critical to transcriptional antitermination *in vivo.* In the N-terminal helix an alanine docks in the major groove of an RNA stem whereas in the C-terminal helix the indole ring of tryptophan stacks against a purine of an RNA loop. At these positions all 19 possible amino acid substitutions have been constructed in the peptide; each decreases specific RNA binding. Our results define a novel structural motif of RNA recognition (Draper, 1995), rationalize patterns of allowed sequence substitutions in a transcriptional antitermination factor (Franklin, 1993), and extend the repertoire of inducible ARM structures.

MATERIALS AND METHODS

RNA Synthesis. RNA was prepared as described (Davis, 1995). For binding studies the crude product was purified from denaturing polyacrylamide gel electrophoresis. Preparative purification was accomplished using ion-exchange HPLC. Purity (>98%) was assessed by HPLC and gel electrophoresis with ^{32}P autoradiography.

Peptide Synthesis. Peptides (N residues 2−22 with C-terminal amide; λ_{P1}) were prepared using Fmoc chemistry, lyophilized, desalted, and purified by RP-HPLC. Fidelity of synthesis was verified by mass spectrometry. Peptide variants were obtained from Chiron, Inc. (San Diego, CA), purified, and characterized as above. Peptide concentration was determined by optical density; purity (>98%) was assessed by HPLC and mass spectrometry.

Biosynthetic Labeling. A synthetic N gene fragment (residues 2−22) was prepared by solid-phase DNA synthesis and cloned into an overexpression plasmid encoding a thrombin-cleavable fusion protein (King & Weiss, 1993). The biosynthetic peptide (23 residues; Figure 5) contains two nonnative N-terminal residues and is designated λ_{P1*}. Incorporation of ^{15}N and ^{13}C was accomplished by growth of *Escherichia coli* in M9 minimal medium. Synthetic and biosynthetic peptides exhibit similar affinities for *boxB* as demonstrated by GMSA.

Gel-Mobility-Shift Retardation Assays. Peptide−RNA binding complexes were incubated at 4 °C in 20 mM Tris−acetate and 50 mM potassium acetate at pH 7.9 (as measured at room temperature). Free and bound species were resolved at 4 °C using a 10% gel [20:1 acrylamide:bis(acrylamide)] in 10 mM Tris−acetate (or Tris−borate) and 0.1 mM EDTA at pH 7.3 (measured at room temperature).

Optical Spectroscopy. CD spectra were obtained using an Aviv spectropolarimeter; deconvolution was estimated as described (Yang *et al.*, 1986). Binding buffer consists of 10 mM potassium phosphate (pH 7.4), 100 mM KCl, and 0.1 mM ethylenediaminetetraacetic acid. Fluorescence spectra were obtained using a SPEX steady-state fluorometer with an excitation wavelength of 295 nm. Correction of the inner filter effect was based on control experiments in 2 M KCl, in which no binding occurs (Supporting Information).

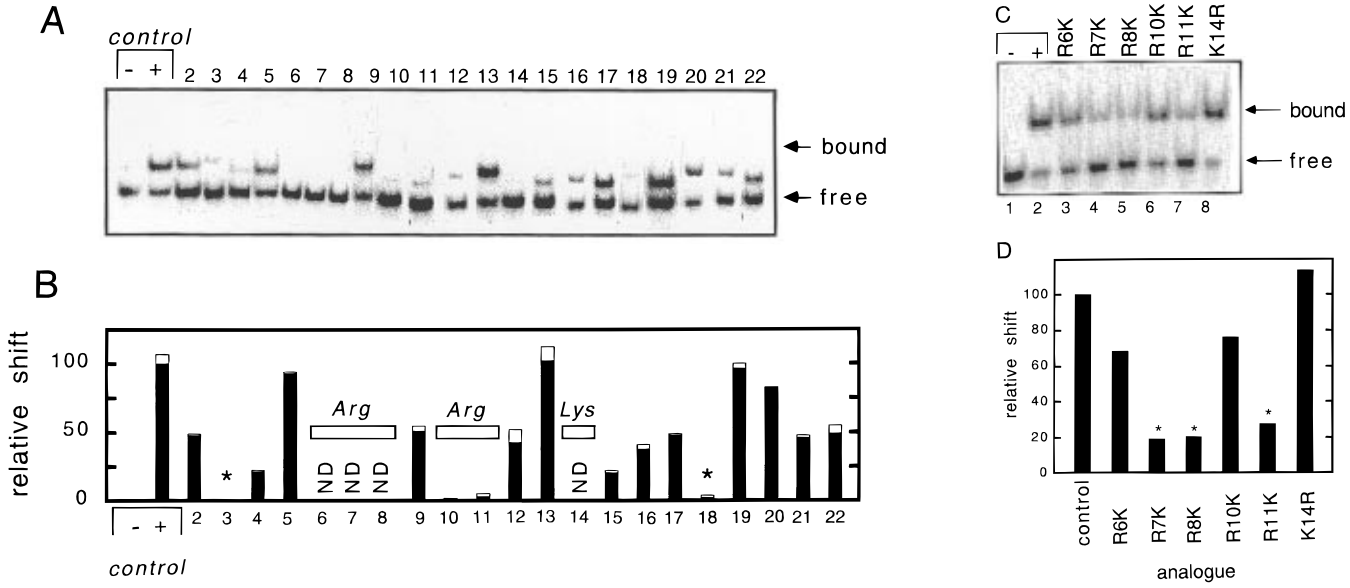


FIGURE 2: Scanning mutagenesis. (A) Gel-shift experiment showing effects of single amino acid substitutions in λ_{P1} at successive positions 2−22. Control indicates wild-type peptide. Non-alanine residues were replaced by alanine; the five native alanines were replaced by glycine (peptide positions 3, 12, 16, 20, and 21). (B) Quantitation of gel showing mean of three trials (black bar) and standard deviation (white bar). Critical residues are Ala3 and Trp18 (asterisks) in addition to five arginines and one lysine (open rectangles). ND indicates no detectable binding. Conditions were as in Figure 1D. Asterisks indicate sites of approximate 50-fold perturbation. (C and D) Gel-retardation experiment showing effects of interchanging arginine and lysine in λ_{P1} at positions shown. Control indicates native λ_{P1}. (D) Quantitation of gel as normalized to native peptide (100%). Asterisks indicate sites of approximate 5-fold perturbation. Conditions were as in Figure 1 except that the peptide concentrations were 10 nM.

NMR Spectroscopy. ^{1}H-NMR spectra were obtained at 400, 500, and 600 MHz. Spectra in H_2O were obtained without solvent presaturation by use of laminar-shaped pulses and pulse-field gradients (Patt, 1992; Smallcombe, 1992; Dhalluin *et al.*, 1995). RNA base and H1′ resonance assignments were obtained as described (Varani & Tinoco, 1991). Assignment of the peptide backbone and non-arginine side chains was obtained by combined application of triple-resonance [gradient-assisted HNCA, HN(CO)CA, and CB-CA(CO)NH] and double-resonance (HMQC, constant-time HSQC, 2D HMQC-NOESY, 2D HMQC-TOCSY, 3D NOESY-HSQC, and 3D TOCSY-HSQC) methods (Bax, 1994; Vuister & Bax, 1992; Hallenga & Lippens, 1995). NOESY spectra with mixing times of 30, 70, 100, 150, 200, 250, and 300 ms were obtained. $^3J_{\alpha NH}$ coupling constants were obtained from DQF-COSY spectra by the method of Redfield and Dobson (1990). Protection of exchangable side chain protons was assessed by comparison of HSQC spectra obtained with and without presaturation (Sklenar *et al.*, 1993). NOEs between the labeled peptide and unlabeled RNA were distinguished by the combined use of isotopic half-filters (Otting *et al.*, 1990; Folmer *et al.*, 1995). A chemical shift index was employed as described (Figure 5; Wishart & Sykes, 1994).

RESULTS

Characterization of a Minimal Model. Structural studies employ a 15-base RNA hairpin (*nut*$_L$ *boxB*; Figure 1B) and the N-terminal 21 residues of the N protein without initiator methionine (designated λ_{P1} in Figure 1C; Tan & Frankel, 1995).[2] The peptide−RNA dissociation constant (K_d 6 nM) is similar to that of the intact protein (Van Gilst *et al.*, 1997); no enhancement is observed upon lengthening the peptide to 30 residues (λ_P; Figure 1C,D). Deletion analysis demonstrates attenuation of RNA binding with removal of any N- or C-terminal amino acid (Supporting Information). No binding is detectable by GMSA of isolated N- or C-terminal fragments (λ_{P2} and λ_{P3} in Figure 1C). No enhancement of affinity is observed on lengthening the RNA stem; deletion of one or more 3′ or 5′ bases from *boxB* leads to no detectable peptide binding by GMSA. The specificity of λ_{P1} and related N peptides is well characterized (Lazinski *et al.*, 1989; Tan & Frankel, 1995; Van Gilst *et al.*, 1997; Cilley & Williamson, 1997). λ_{P1} binds with >50-fold decreased affinity to the cognate DNA hairpin and to certain variant RNA sites inactive in transcriptional antitermination (as illustrated by the substitution G6 → C in Figure 1D; Tan & Frankel, 1995; Cilley & Williamson, 1997). Further, the analogous ARM of phage ϕ21 N protein (sequence given in Figure 1C) does not bind to λ *boxB* RNA (Figure 1D; Tan & Frankel, 1995). Specific binding of the λ N peptide is more readily disrupted by substitutions in *boxB* than is that of the N protein, presumably due to additional nonspecific RNA contacts by the intact polypeptide (Cilley & Williamson, 1997).

Alanine Scanning Mutagenesis Correlates with Genetic Analysis. To identify peptide residues critical for *boxB* binding, each (non-alanine) amino acid was singly changed to alanine; the five native alanines were singly changed to glycine. Variants were tested for specific RNA-binding activity by GMSA (Figure 2A,B). Found to be of critical importance ($\Delta\Delta G$ > 2 kcal/mol) are each arginine (R6, R7, R8, R10, and R11), one lysine (K14), and two neutral side chains (A3 and W18). Small decrements in affinity ($\Delta\Delta G$ < 0.5 kcal/mol) are observed following substitution of carboxamide side chains (Q4, Q15, Q17, and N22) by alanine

[2] Peptide numbers refer to the codon number including the initiator methionine (Franklin *et al.*, 1985b); the N-terminal residue of the peptide is hence designated D2, the second residue A3, *etc.* The same numbering scheme is used in the biosynthetic peptide, which contains two N-terminal non-native residues (GS; Figure 5).

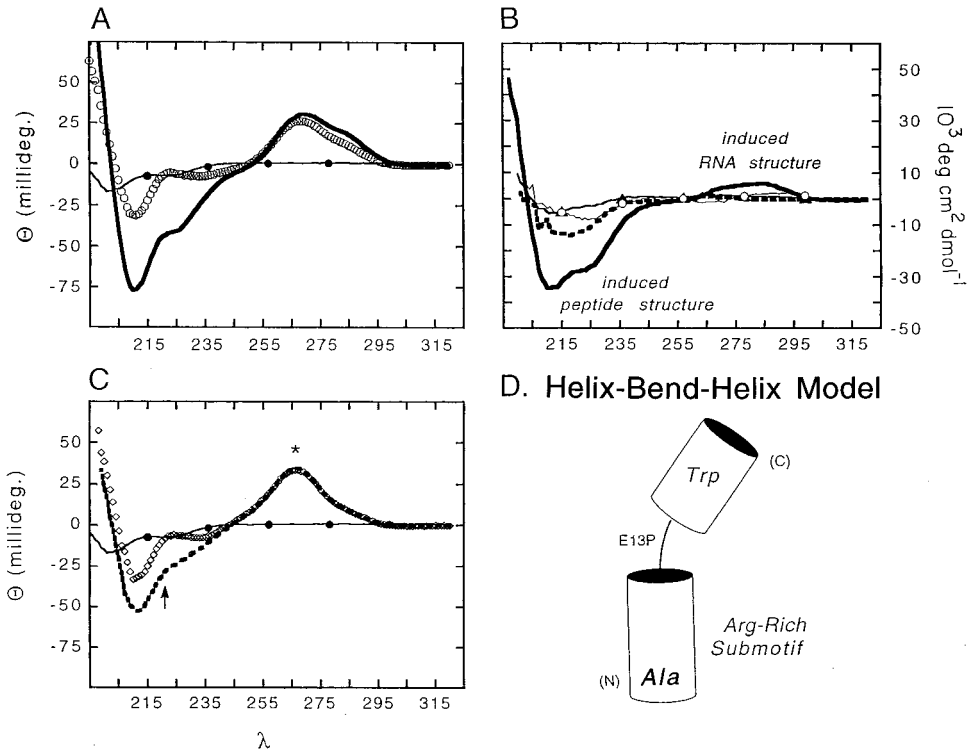


FIGURE 3: Circular dichroism. (A−C) Studies of λ_{P1}. Binding of λ_{P1} to native RNA site is shown in (A): spectra of free RNA (open circles), free peptide (closed circles), and specific complex (thick line). (C) Binding of λ_{P1} to C6 variant RNA: spectra of free RNA (open diamonds), free peptide (closed circles), and distorted complex (filled circles). The asterisk indicates absence of change in the RNA band; the arrow indicates blunting of induced α-helical band. Difference spectra are shown in (B): the thick line indicates the difference between the CD spectrum of the specific complex and the sum of the spectra of the isolated parts. In dashed line is shown the analogous difference spectrum for the G6C variant complex. In thin line is shown the corresponding difference spectra in 2 M KC (no binding). A control difference spectrum using an analogous ϕ21 N peptide (Figure 1C) is also shown (open circles; no structural change detected). Peptide concentrations in each case were 50 μM in 100 mM KCl, 10 mM potassium and phosphate (pH 7.4), and 0.1 mM EDTA at 4 °C. (D) Cylinder model of the proposed helix−bend−helix model of induced peptide structure. Critical alanine and tryptophan residues (see Figure 2) are shown; the site of functional E13P substitution (see Figure 6; Franklin, 1993) is shown in the bend.

and of three alanines (A12, A16, and A20) by glycine. Because absence of a shifted band could represent either decreased affinity or kinetic instability (Cilley & Williamson, 1997), a fluorescence quenching assay (Van Gilst *et al.*, 1997) was employed to corroborate >50-fold decreased binding following K14A and each Arg → Ala substitution (data not shown). To distinguish between general electrostatic interactions and specific contacts by arginine or lysine, each was substituted in turn by the other as described by Tao and Frankel (1993). Fivefold reduction in binding is observed following such interchanges at position 7, 8, or 11 whereas the remaining interchanges are well tolerated (Figure 2C,D). We speculate that Arg7, Arg8, and Arg11 provide base- or architecture-specific contacts. The overall results of scanning mutagenesis and the particular importance of arginines 7, 8, and 11 are in agreement with genetic analysis of the the N protein by amber suppression and random-cassette mutagenesis (Franklin, 1993). Such correlation provides evidence that the peptide mimics the ARM domain of the native protein.[3]

An α-Helical Motif Is Induced on Specific RNA Binding. CD spectra of λ_{P1}, *boxB* RNA, and their specific complex are shown in Figure 3A. The peptide exhibits a largely random-coil spectrum; the α-helix content is <10% at 4 °C (Tan & Frankel, 1995; Van Gilst *et al.*, 1997). That a change in peptide conformation occurs on RNA binding is suggested by the CD spectrum of the specific complex (solid line in Figure 3A). Analysis of the difference spectrum (solid line in Figure 3B)[4] predicts that the bound peptide has an α-helix content of 80% (16 residues). The nonhelical component is predicted to comprise a turn (20%); no random coil or β-strand is predicted (Yang *et al.*, 1986). A change in RNA-specific CD bands in the near-ultraviolet (similar to that in the α-helical Rev−*RRE* complex; Tan & Frankel, 1994) is also observed (labeled in Figure 3B). The CD transitions are each blunted by a G6 → C substitution in the RNA (respectively indicated by an arrow and asterisk in Figure 3C) at peptide and RNA concentrations sufficiently high (50 μM) to allow formation of a variant equimolar complex (K_d 0.5 μM). This substitution is associated with complete loss of N-dependent transcriptional antitermination *in vivo* (Doelling & Franklin, 1989). These results suggest that the induced structure of the peptide is RNA-sequence-specific and is related to biological activity. Induction of an α-helical motif is in accord with recent CD studies of the intact N protein (like λ_{P1}, a random coil in isolation) and its specific *boxB* complex (also containing an estimated 16 helical residues; Van Gilst *et al.*, 1997).

[3] The small decrement in binding associated with substitution of noncontact alanines by glycine (Figure 2) is presumably due to partial destabilization of the bound α-helices. No intermolecular NOEs are observed from these alanine methyl groups to the RNA.

[4] CD difference spectra (Figures 3 and 6) are defined as the difference between the spectrum of a complex and the sum of the spectra of the isolated components.

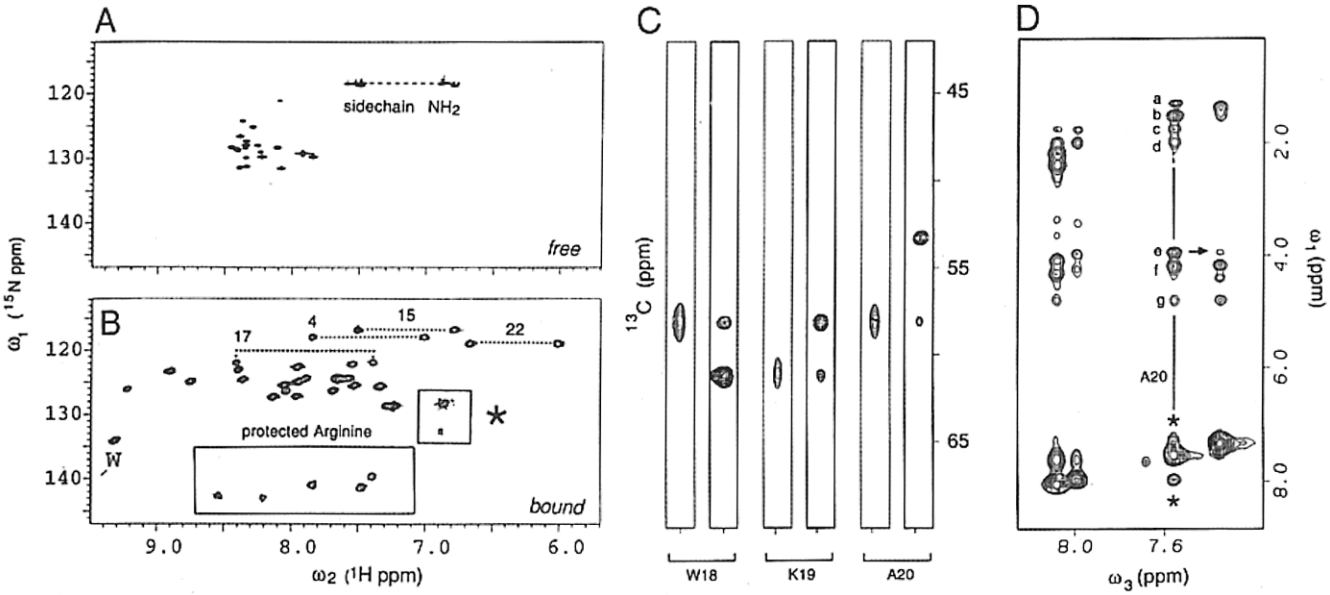


FIGURE 4: (A and B) Comparison of free and bound λ_{P1*} provides an NMR HMQC "footprint". (A) 400 MHz $^1H-^{15}N$ HMQC spectrum (with solvent presaturation) of free peptide λ_{P1*} containing uniform ^{15}N labeling demonstrates the narrow range of chemical shifts of peptide NH resonances (central cluster) and unresolved carboxamide side chain NH_2 resonances of Gln (three residues: Q4, Q15, and Q17) and Asn (N22) (red dotted line). No arginine $^{15}N-H$ correlations are observed due to rapid solvent exchange. (B) Corresponding 400 MHz HMQC spectrum (with solvent presaturation) of the specific complex reveals dramatic enhancement of chemical shift dispersion due to peptide folding. In the black box are five protected Arg $^{15}N_\epsilon H$ resonances (one per side chain); in the blue box (asterisk) are weak Arg $^{15}N_\zeta H_2$ resonances (two per side chain). The latter are enhanced by gradient HSQC methods (Vuister & Bax, 1992; Hallenga & Lippens, 1995), indicating three or four very well protected arginine guanidinium NH_2 groups in the peptide−RNA interface. The indole NH resonance of W18 is labeled W in (B); it is not seen in (A) due to its location downfield of the spectral region shown. Spectra were obtained at 25 °C and pH 6; solvent water was presaturated for 1 s. The arginine resonances are folded from higher field (^{15}N transmitter offset at 125 ppm with a sweep width of 52.5 ppm). Actual Arg $^{15}N_\epsilon$ chemical shifts are 88−92 ppm, and Arg $^{15}N_\zeta$ chemical shifts are 75−82 ppm. (C) Representative regions of 3D HNCA (black) and CBCA(CO)NH (red) spectra at 500 MHz illustrating the sequential main chain assignment W18-K19-A20. CBCA(CO)NH (red) panels contain sequential connectivities ($^{13}C_{\alpha(i,i-1)}$, $^{15}N_i$, H_{Ni}); HNCA contains both intraresidue and sequential correlations (Bax, 1994). (D) Representative plane of the 3D ^{15}N NOESY-HSQC spectrum at 600 MHz showing the ^{15}N-filtered $^1H-^1H$ NOESY plane (editing frequency 125.2 ppm). Asterisks indicate d_{NN} NOEs between residues 19 and 20 (above diagonal) and residues 20 and 21 (below diagonal). The arrow indicates the sequential $d_{\alpha N}$ NOE between residues 19 and 20. Cross peaks a−f are assigned: (a) A20-H_N/A20-CH_3; (b) A20-H_N/A16-CH_3; (c and d) A20-H_N/Q17-H_β; (e) intraresidue H_N-H_α, (f) A20-H_N/A21-H_α; and (g) H_2O exchange cross peak.

Multidimensional NMR Demonstrates a Helix−Bend−Helix Motif. The secondary structure of the bound peptide λ_{P1*} (cylinder model in Figure 3D) was determined by heteronuclear 3D NMR spectroscopy following uniform labeling of the peptide with ^{13}C and ^{15}N (see Materials and Methods). The free peptide exhibits near-random-coil ^1H, ^{13}C, and ^{15}N chemical shifts (illustrated in Figure 4A). Marked dispersion of backbone and side chain chemical shifts occurs on RNA binding (Figure 4B). Illustrated in panels A and B of Figure 4 are the large ^1H and ^{15}N complexation shifts of Q4, Q17, and N22 (dotted red and blue lines connect pairs of NH_2 resonances); the complexation shift of Q15 is by contrast negligible. Analogous dispersion is observed among exchangable arginine resonances in the complex (boxed in Figure 4B).

Sequential assignment was obtained by standard double- and triple-resonance methods (Bax, 1994) as illustrated in panels C and D of Figure 4. Because the RNA in these experiments is unlabeled, such methods permit selective observation of the bound peptide. Analysis of peptide-specific NOEs, coupling constants, and diagnostic $^1H_\alpha$ and $^{13}C_\alpha$ chemical shifts (Wyshart & Sykes, 1994) reveals the presence of N- and C-terminal helices (Figure 5). An interruption in the α-helical motif occurs between residues 9 and 12, as indicated by a discontinuity in the pattern of helix-related NOEs, backbone coupling constants ($^3J_{\alpha NH}$), and $^{13}C_\alpha$ and H_α chemical shifts (Wishart & Sykes, 1994). Although the string of helix-related $d_{(i,i+3)}$ NOEs is broken from residues 12 to 14, weak NOEs are observed from residues 9 to 12 ($d_{N\alpha(i,i+3)}$ and $d_{N\beta(i,i+3)}$) and 12 to 15 ($d_{\alpha N(i,i+3)}$); further, an (i,i+4) spans the region from residues 9 to 13 ($d_{\alpha N(i,i+4)}$). These data, although insufficient to specify the precise structure of the intervening region, indicate the presence of a non-α-helical bend. Observation of three successive large $^3J_{\alpha NH}$ couplings (residues 9−11; Figure 5) indicates that the interruption is nonrandom but not a canonical β-turn (*e.g.*, type I, type II, *etc.*; Wüthrich, 1986). Unlike the classical HTH, whose supersecondary structure contains a β-turn and is stabilized by helix−helix packing (Harrison & Aggarwal, 1990), no long-range NOEs are observed between helices.

Bent α-Helix Motif Rationalizes Allowed Proline Substitutions. Observation of a bent α-helix motif rationalizes the location of proline substitutions associated with loss or retention of activity. Whereas the substitutions Thr5 → Pro (in the N-terminal helix between the critical side chains of A3 and the arginine cluster; AQ**P**RRR), E9 → P (between the two arginine clusters; RRR**P**RR), and W18P (in the C-terminal helix) lead to loss of function (Franklin, 1993; Tan & Frankel, 1995), genetic studies indicate retention of high biological activity following proline substitution at position 13 and partial activity following substitution at position 12 or 15 (Franklin, 1993). These results suggest that proline in or just C-terminal to the bend permits

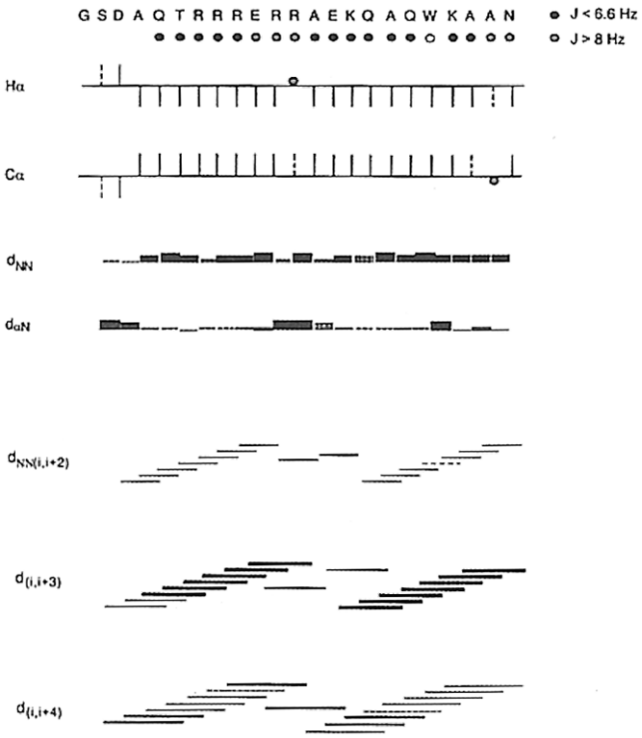


FIGURE 5: Secondary structural analysis of bound peptide λ_{P1*}. Summary of $^3J_{\alpha N}$ coupling constants, chemical shift indices (Wishart & Sykes, 1994), and sequential assignment in standard format (Wuthrich, 1986). The sequence of the labeled peptide is shown at the top. Shown in red are values inconsistent with a continuous α-helix: large $^3J_{\alpha N}$ coupling constants, anomalous H_α or $^{13}C_\alpha$ chemical shifts, small d_{NN} NOE intensities, and large $d_{\alpha N}$ NOE intensities. Dashed lines indicate borderline chemical shifts (rows 3 and 4) or NOEs that are ambiguous due to overlap. Relative NOE intensities are schematically indicated by the thickness of the line. The two contiguous strings of strong (i,i+3) helix-related NOEs are shown in blue and green. Medium-range NOEs maintained in the central bend are shown in black. Helix-related (i,i+3) NOEs include $d_{\alpha N(i,i+3)}$, $d_{\beta N(i,i+3)}$, and other backbone−side chain contacts; likewise, (i,i+4) NOEs include $d_{\alpha N(i,i+3)}$, $d_{\beta N(i,i+3)}$, and other backbone−side chain contacts.

maintenance of native peptide−RNA contacts. Investigation of the E13P analog of λ_{P1} in fact demonstrates native RNA affinity (inset in Figure 6A); moreover, the induced peptide structure (as monitored by CD; Figure 6B and difference spectrum in Figure 6D) is similar to that of the native complex. The E13P analog also exhibits native quenching of the intrinsic fluorescence of Trp18 (Figure 6A; Van Gilst *et al.*, 1997), shown below to correlate with stacking of the indole ring at the peptide−RNA interface.

Peptide Folding Is Coupled to a Specific RNA Structure. Removal of the exocyclic amino group of G6 by inosine substitution has recently been shown to preclude specific protein binding (Van Gilst *et al.*, 1997). Inosine substitution likewise leads to no detectable binding of λ_{P1} by GMSA (inset in Figure 6A). Weak binding (K_d 500 nM) is nevertheless observed by fluorescence and is associated with an abridgement of peptide-induced fit: the bound peptide exhibits incomplete quenching of Trp18 (Figure 6A) and an attenuated α-helical transition (Figure 6C; difference spectrum in Figure 6D). Because no intermolecular NOEs are observed between the G6 amino group and the peptide (see below), this perturbation is indirect: absence of the G6 amino group destabilizes an induced RNA structure necessary for native peptide folding (a sheared GA base pair; see Discussion).

Peptide−RNA Contacts. Sites of contact between the peptide and RNA were identified by isotope-filtered NMR experiments (Otting *et al.*, 1990), designed to resolve NOEs between ^{13}C- or ^{15}N-attached protons in a labeled peptide and ^{12}C- or ^{14}N-attached protons in the unlabeled RNA (Folmer *et al.*, 1995). Contacts to bases involve the side chains of Ala3, Trp18, and an unassigned arginine. Backbone contacts are made by the side chains of additional arginines (unassigned), Gln4, Gln17, and Asn22. The importance of these side chains in N-dependent antitermination has been demonstrated *in vivo* by mutagenesis (Franklin, 1993). The peptide−RNA interface is thus remarkable for a combination of charged, polar and nonpolar base contacts as proposed on the basis of sequence comparisons (Lazinski *et al.*, 1989) and inferred from genetic analysis (Franklin, 1993). Base- and ribose-specific contacts in the stem are restricted to major groove protons; no contacts are made to $H_{1'}$ anomeric ribose protons or adenine H_2 protons (A10 and A11) in the minor groove.

Base-specific interactions by peptide side chains are defined by observation of direct NOEs at low mixing times (30−70 ms; see Materials and Methods). Such data position the Ala3 methyl group in the RNA major groove between C2 and C3 (in each case adjoining pyrimidine protons H_5 and H_6; Figure 7E); the complementary guanines (G13 and G14) are not contacted. NOEs are also observed between the methyl group and unassigned ribose protons. The methyl resonance of Ala3 exhibits a negligible complexation shift (<0.05 ppm in magnitude), consistent with absence of side chain intercalation. These observations place the N-terminal portion of the peptide in the RNA major groove at the base of the stem. The guanidinium group of an unassigned arginine is near the major groove protons U5-H_6 and G6-H_8; no side chains are detected near the G6 amino group involved in GA base pairing (Su *et al.*, 1997).

The indole ring of Trp18 contacts the base at A7 (Figure 7E). The C-terminal region of the peptide is thus near the RNA pentaloop. Large upfield complexation shifts are observed among Trp18 indole resonances (Supporting Information). These shifts presumably reflect the aromatic ring current of A7; upfield complexation shifts are likewise observed in the resonances of A7, presumably due in part to the reciprocal ring current of Trp18. Stacking is demonstrated by observation of contacts between corresponding protons in respective five- and six-membered ring systems (A7-H_2/indole-H_6 and A7-H_8/indole-H_2; respectively indicated by an arrow and asterisk in Figure 7D). Smaller NOEs are observed to adjoining indole protons (A7-H_2/indole-H_7 and A7-H_2/indole-H_5; red box in Figure 7D). No contacts are observed between the side chain of Trp18 and other bases, including G6 or A8. Trp18 is thus not between A7 and another base: the indole ring *overlies* A7 but does not intercalate. The NOEs between Trp18 and A7 differ only slightly from those expected from precise vertical stacking (A7-H_2/indole-H_5 and A7-H_8/indole-H_2; arrow and asterisk in Figure 7C). A7 itself stacks over A8 at the crux of a GNRA-like RNA fold (Su *et al.*, 1997).

Specificity of Nonpolar Contacts. To investigate the specificity of the alanine- and tryptophan-binding sites in *boxB*, the 19 possible amino acid substitutions were singly introduced at these peptide positions and tested by GMSA.

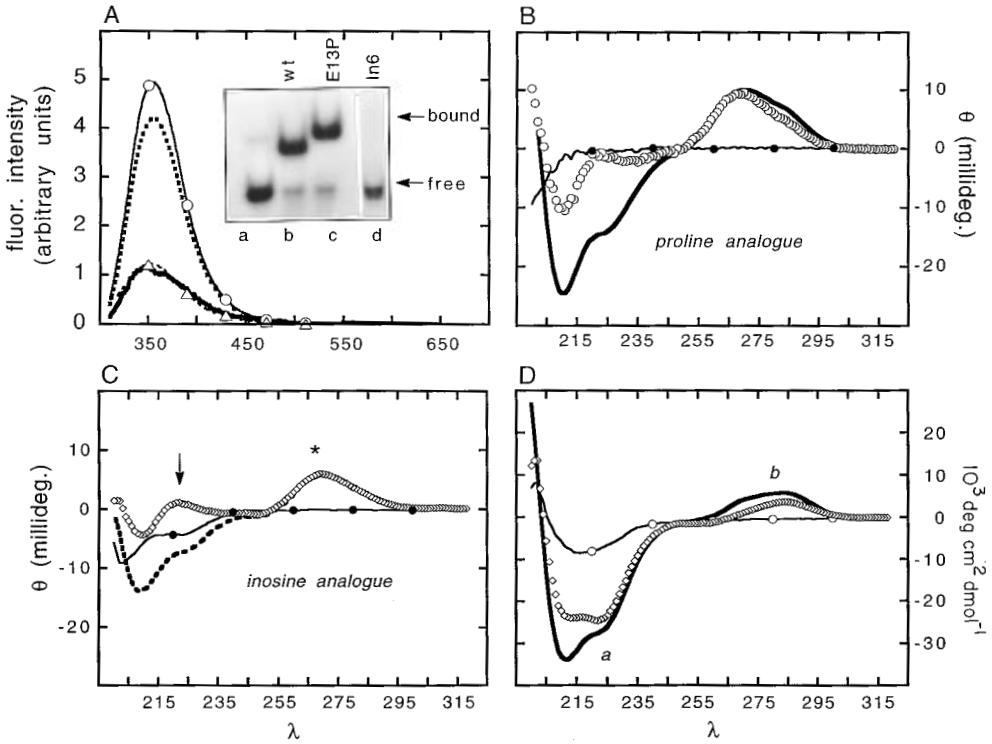


FIGURE 6: (A) Fluorescence emission spectra demonstrate almost complete quenching of Trp18 in the native and E13P λ_{P1} complexes (superimposed thick line and dashed line at the bottom indicated by triangles) relative to free peptides (open circles). Low-affinity binding of the native peptide to the G6I RNA site is associated with attenuated quenching (dotted line). Peptide and RNA concentrations were 1 μM. The G6I variant is approximately 75% bound under these conditions, indicating that its extrapolated extent of quenching in the bound state is incomplete. Inset: RNA-binding assay (GMSA) showing native binding by wild-type and E13P variant λ_{P1} peptides (lanes b and c); no binding is detectable to the G6I variant RNA site (lane d). The concentration of λ_{P1} is 20 nM. (B) Binding of the E13P λ_{P1} variant peptide induces an α-helical transition similar to that of the native peptide (Figure 3A). (C) Binding of the λ_{P1} to the G6I variant RNA induces an attenuated α-helical transition similar to that in the G6C complex (Figure 3C). No change in RNA-specific bands (255−300 nm) is observed. (D) CD difference spectra calculated for the native complex (thick line; a), the E13P complex (diamonds), and the G6I complex (thin line; circles). The change between the native and E13P double-difference spectra in region a is due largely to a change in the spectra of the free peptides; the change in region b is assigned to RNA structure.

At position 3 (Figure 8A,B) the highest affinity is conferred by the native side chain (alanine); near-native affinity is conferred by serine (80%) followed by threonine and valine (25%). The majority of substitutions yield little or no specific RNA binding at peptide concentrations of 10 and 50 nM. The pattern of active and inactive substitutions is in accord with genetic analysis (Franklin, 1993). Analogous results are observed at peptide position 18 (Figure 7A,B). As expected for coplanar stacking of an indole ring, each substitution was observed by GMSA to decrease formation of a shifted complex; eight (Asn, Asp, Cys, Glu, Gly, Ser, Pro, and Thr) lead to no detectable binding (Figure 7B). Of active variants, highest apparent affinities are achieved by aromatic side chains and in particular by tyrosine. Aliphatic side chains are associated with >10-fold decrements in apparent affinity. Lysine and arginine confer low but detectable binding, presumably due to formation of compensating electrostatic interactions.

DISCUSSION

A general feature of protein−RNA recognition is costabilization of novel binding surfaces (Tan & Frankel, 1995). Induced fit is associated with a diverse repertoire of nonstandard base pairing and backbone organization (Puglisi *et al.*, 1995; Battiste *et al.*, 1994, 1996; Ye *et al.*, 1995, 1996; Allain *et al.*, 1996). Disordered regions of a protein can likewise adopt well organized structures on RNA binding. Such induced peptide structures are well characterized in the Tat and Rev proteins of mammalian immunodeficiency viruses. These proteins, virulence factors in viral pathogenesis, contain analogous ARM domains (Lazinski *et al.*, 1989). The structures of specific complexes nevertheless reveal different modes of RNA recognition. Whereas Tat folds as a β-hairpin (Puglisi *et al.*, 1995; Ye *et al.*, 1995), Rev forms a continuous α-helix (Battiste *et al.*, 1996; Ye *et al.*, 1996). The present study focuses on a novel peptide structure induced in the ARM domain of the λ transcriptional antitermination factor N.

The N peptide, largely disordered when free (Tan & Frankel, 1995), is predominantly α-helical in a specific RNA complex as described in studies of the intact N protein (Van Gilst *et al.*, 1997). The α-helical transition (as monitored by CD) is incomplete (Figure 3), suggesting that the "recognition α-helix" is interrupted. This possibility is supported by the genetic data of Franklin (1993): a continuous α-helix seems inconsistent with the distribution of critical side chains as inferred from random-cassette mutagenesis and in particular with observation of functional proline substitutions near the center of the motif (positions 12, 13, and 15; indicated in Figure 9). Multidimensional NMR studies explicitly demonstrate RNA-dependent folding of a bent α-helix in the major groove of an RNA stem and contiguous loop. The central bend, not a canonical β-turn, is characterized by three successive large coupling constants with maintenance of weak (i,i+3) contacts (Figure 5). One face of the entire structure is in close contact with RNA.

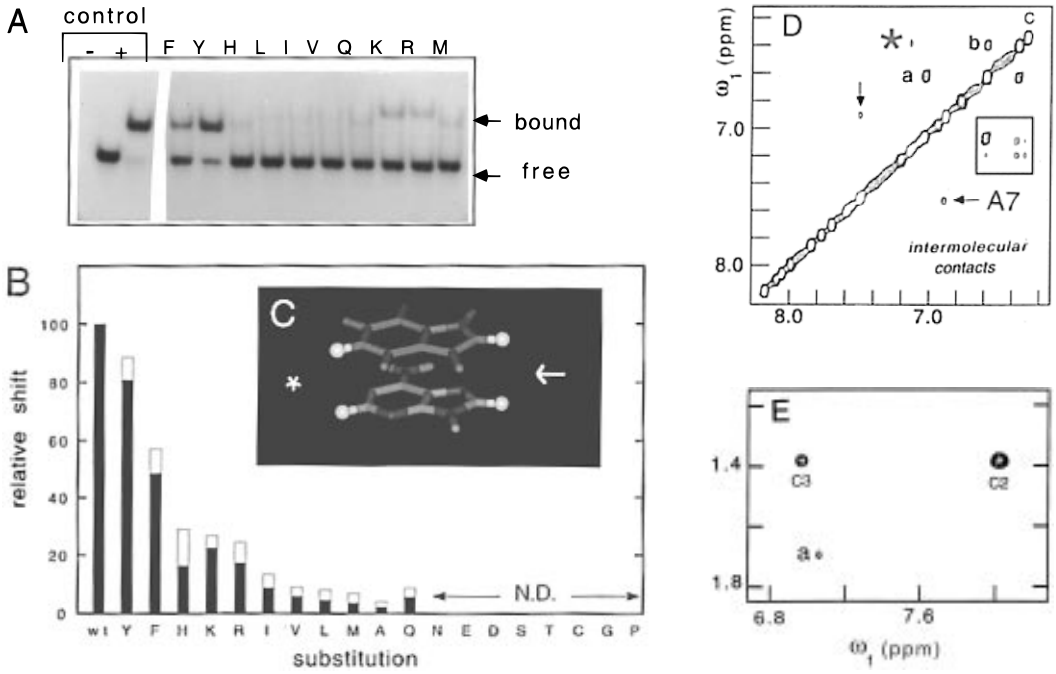


FIGURE 7: (A and B) Mutational analysis of Trp18 in N peptide λ_{P1}. Gel-retardation experiment showing reduction in RNA binding by selected substitutions at position 18. Control indicates native peptide (Trp18). Analogs without detectable binding are not shown. The respective peptide and RNA concentrations were 20 and 0.5 nM. (B) Quantitation of the gel showing mean (solid bar) and standard deviation (open bar) of three repetitions of the experiment. N.D. indicates no detectable binding. (C) Inset: hypothetical model of vertical stacking between the indole ring (green and blue; top) and adenine (red and blue; bottom). The asterisk indicates the predicted proximity of Trp18-H_5 and A7-H_2 (white balls); the arrow indicates the proximity between Trp18-H_2 and A7-H_8. Nitrogen atoms are shown in blue. (D) Direct intermolecular NOEs are observed between Trp18 and A7 in the RNA−λ_{P1} complex at low mixing time (70 ms at 25 °C). The asterisk indicates contact between Trp18-H_6 and A7-H_2; arrows indicate contact between Trp18-H_2 and A7-H_8; the indole H_5−H_6 NOE is labeled a and the H_6−H_7 NOE is labeled b. The pattern of adenine−indole NOEs indicates that the five-membered ring of the purine overlies the five-membered ring of the indole; likewise stacked are respective six-membered rings. At lower contours (red box) are observed additional NOEs between A7-H_2 and indole H_5 and H_7. Also seen at low contour are weak NOEs between the respective H_2 protons of A7 and A8 (not shown). (E) Direct contacts between the methyl resonance of Ala3 (horizontal axis) and H_6 protons of C3 and C2 (vertical axis) rationalize effects of mutagenesis *in vivo* (Franklin, 1993). Experimental conditions: 25 °C and a mixing time of 70 ms at 600 MHz.

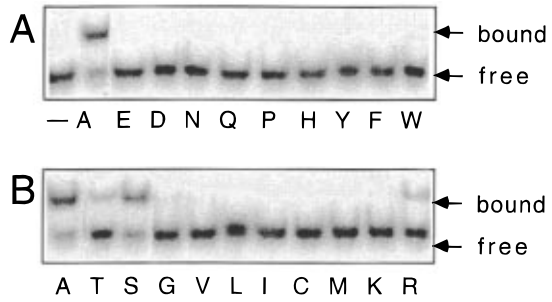


FIGURE 8: Mutational analysis of N peptide λ_{P1} at residue 3 (Ala3). (A and B) All 19 substitutions were constructed in the peptide and screened for specific RNA binding by GMSA. The peptide concentration was in each case 10 nM. The highest affinity is achieved by the native peptide. The relative affinity of the Ser3 variant is 80%, and Arg and Thr are 25%. At this or higher (50 nM) peptide concentrations no shifted band is observed on substitution of Ala3 by Asn, Asp, Gln, Glu, Ile, Lys, and Phe.

We imagine that a bent α-helix enables the peptide to follow the contours of a bipartite RNA surface of an RNA hairpin (Chattopadhyay *et al.,* 1995a; Mogridge *et al.,* 1995). Docking of the N-terminal α-helix in the major groove of an RNA stem is analogous to the binding mechanism of the arginine-rich HIV-1 Rev peptide (Battiste *et al.*, 1994, 1996; Ye *et al.,* 1996).

Arginines in the N protein are clustered in the N-terminal α-helix and central bend. As expected from analysis of the intact protein (Lazinski *et al.,* 1989; Franklin, 1993), these side chains are essential for the specific RNA-binding activity of λ_{P1} (Figure 2A). The observed pattern of functional Arg−

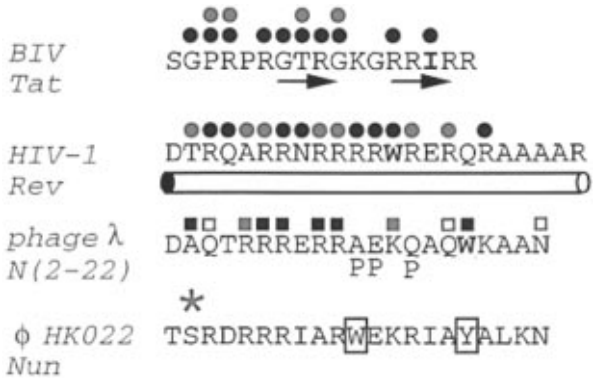


FIGURE 9: Amino acid sequences of arginine-rich motifs of BIV Tat protein, HIV-1 Rev protein, phage λ N protein, and phage HK022 Nun protein. The structures of BIV Tat− and HIV-1 Rev−RNA complexes have been determined and reveal β-sheet and α-helical binding motifs, respectively. Strands and helices are indicated by arrows and a cylinder, respectively. Red circles indicate backbone contacts, and blue circles indicate base-specific contacts. Analogous squares above the λ sequence indicate proposed backbone and base-specific contacts; open squares indicate that the proposed contact is of minor importance in affinity (as inferred by alanine mutagenesis; Figure 2). Use of aliphatic or aromatic side chains in the peptide−RNA interface is indicated by red letters. The three P's beneath the λ sequence indicate sites of conservative proline substitution (Franklin, 1993). The red asterisk indicates corresponding Ala3 and Ser3 in λ N and HK022 Nun, respectively; red boxes in the Nun sequence indicate sites of possible contact by aromatic side chains.

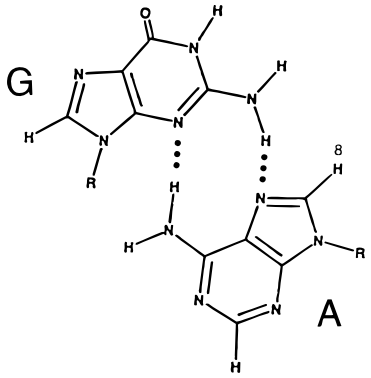
Lys interchanges (Figure 2C) provides an "electrostatic footprint" different from that of Tat or Rev (Tao & Frankel,

1993; Tan & Frankel, 1994), suggesting a distinct three-dimensional array of base- or architecture-specific arginine interactions. ^{15}N and ^{13}C labeling of the N peptide has enabled observation of individual guanidinium groups; on RNA binding, each side chain $N_\epsilon H$ is protected (on the time scale of 0.1−1 s) from solvent exchange (Figure 4A,B). Arginines contribute both base- and backbone-specific RNA contacts; one (unassigned) lies in the major groove between U5 and G6. By analogy to the HIV-1 Rev complex (Battiste *et al.*, 1996; Ye *et al.*, 1996), we propose that the latter guanidinium group contributes base- and groove-specific hydrogen bonds to base-specific hydrogen bond acceptors: the carbonyl function of U5 and to the *aza* nitrogen (N_7) of G6. Binding of the N protein and derived peptides is specific for U and G at these positions (Chattapadhyay *et al.*, 1995a; Tan & Frankel, 1995; Cilley & Williamson, 1997).

Genetic analysis has highlighted the importance the nonpolar side chains of Ala3 and Trp18 (Franklin, 1993). Each is found to make base-specific contacts (Figure 7D,E). The methyl group of Ala3 packs in the major groove between C2 and C3 whereas the indole ring of Trp18 overlies the purine ring of A7. We imagine that such packing (like that of Ile79 in the BIV Tat complex; Chen & Frankel, 1995; Puglisi *et al.*, 1995; Ye *et al.*, 1995) permits the close apposition of nonpolar peptide and RNA surfaces. The indole ring of tryptophan resembles a purine: its stacking directly extends the pattern of base stacking in the bound RNA (Su *et al.*, 1997). The *in vitro* specificity of the alanine−cytosine and -tryptophan−adenine interactions was investigated by analysis of peptide substitutions. At each position effects of the 19 possible amino acid substitutions were tested by GMSA. Allowed substitutions at peptide position 3 are Ser > Val, Thr (Figure 6), in accord with genetic analysis (Franklin, 1993). We imagine that the −OH groups of Ser and Thr can form hydrogen bonds with either water or ribose in the RNA major groove. Larger side chains (such as Cys, Leu, and Ile) are presumably disallowed by steric clash, polarity (Asn and Gln), or charge (Asp and Glu). It is interesting that in studies of N variants containing multiple simultaneous substitutions (obtained by random-cassette mutagenesis; Franklin, 1993) unfavorable substitutions at codon 3 were obtained by functional selection only in combination with additional arginines elsewhere in the motif (such as E13R). We speculate that the latter substitutions provided second-site suppression of the unfavorable side chain at residue 3: the additional arginines enhance specific or nonspecific RNA-binding activity. Such electrostatic compensation has been observed in studies of operator and nonoperator DNA binding by variants of phage λ repressor isolated by genetic suppression (Nelson & Sauer, 1986).

At peptide position 18 highest affinities were achieved by peptide variants containing aromatic amino acids; the order of affinity is Trp ≅ Tyr > Phe > His (Figure 7A,B). This trend is in overall accord with genetic analysis of codon 18 by *amber* suppression: only Tyr provides a functional replacement of tryptophan (Franklin, 1993). Further, retention of residual RNA-binding activity by a subset of variant peptides is broadly consistent with the ability of a codon 18 λN_{am} phage to form plaques (intermediate or normal in size) on some suppressor strains of *E. coli* (inserting Ala, Arg, His, Phe, or Tyr) but not on others (inserting Cys, Gln, Glu, Leu, Lys, Pro, or Ser; Franklin, 1993). Although the relative affinities of peptide analogs are in qualitative agreement with genetic results, it is not clear why small differences *in vitro* (such as a 2-fold difference between the affinity of Tyr and Phe analogs; Figure 7A) are associated with large differences in biological activity. It is possible that such differences reflect a second function of the complex, in particular, binding of NusA (Mogridge *et al.*, 1995). If Trp18 and A7 jointly contribute to a NusA-binding surface required for assembly of a processive antitermination complex, then the difference in the biological activities of Tyr18 and Phe18 N variants (Franklin, 1993) could be due to the participation of the Tyr18 *p*-OH group (a potential hydrogen bond donor and acceptor) in NusA recognition. This hypothesis predicts that second-site revertants of the inactive Phe18 N variant are likely to occur in NusA.

The dependence of peptide folding on RNA structure is illustrated by a seemingly conservative substitution: replacement of G6 by inosine. Although inosine differs from guanine only in the absence of the exocyclic amino group, the RNA substitution precludes specific binding and induced fit of the N protein (Van Gilst *et al.*, 1997). The G6I substitution likewise abridges induced fit of the N peptide (Figure 6). Isotope-filtered NMR experiments nevertheless indicate no contacts between the native G6 amino group and the peptide. We note that, of possible GA pairing schemes at neutral pH, *only the sheared geometry utilizes guanine's exocyclic amino group in base pairing*.[5]



In the absence of this hydrogen bond we speculate that a sheared IA base pair in *boxB* would be unstable. Loss of specific binding by the G6I analog therefore suggests that a physical "bridge" of particular geometry is required for the costabilization of active peptide and RNA structures. We demonstrate elsewhere (Su *et al.*, 1997) that binding of λ_{P1} indeed stabilizes a sheared GA base pair in *boxB*. Analogous purine−purine base pairs or base triples occur in Tat and Rev complexes; their stability is also coupled to peptide binding (Puglisi *et al.*, 1995; Battiste *et al.*, 1994, 1996; Ye *et al.*, 1995, 1996). Such noncanonical base pairs serve to orient a specific RNA surface.

Highlighted in Figure 9 are side chain−RNA contacts in the N peptide as identified by NMR or as suggested by alanine scanning mutagenesis.[3] For comparison, a summary is provided of base and backbone contacts by the arginine-rich motifs of BIV Tat and HIV-1 Rev. In each case the contacts are extensive, reflecting deep penetration of the

[5] A variant sheared GA pairing scheme occurs at the GNRA-like fold of an RNA aptomer (Jiang *et al.*, 1996) and is excluded here by a distinct pattern of long-range NOEs (Su *et al.*, 1997).

β-hairpin (Tat) or α-helix (Rev) into a widened RNA major groove. We anticipate that the N−RNA complex will exhibit similar overall features, especially in the docking of the arginine-rich N-terminal helix and central bend in the stem of *boxB*. Also shown in Figure 9 is the arginine-rich sequence of the transcriptional termination protein Nun of phage HK022 (Hung & Gottesman, 1995; Chattopadhyay *et al.*, 1995b). This peptide is also specific for λ *nut boxB*. Although Nun and N have opposite functions (termination and antitermination; Hung & Gottesman, 1995), the arginine-rich motif in the intact N protein may be replaced by the HK022 Nun motif without change in RNA-binding specificity or function (Henthorn & Friedman, 1996). Interestingly, the Nun arginine-rich motif contains serine at position 3 (asterisk in Figure 9), which is the most active of the 19 non-alanine substitutions tested in the N peptide (Figure 8). The Nun motif also contains aromatic side chains (boxed in Figure 9). These sequence features suggest that Nun and N exhibit similar mechanisms of RNA recognition (Chattopadhyay *et al.*, 1995b; M. Gottesman, personal communication).

The RNA-binding helix−bend−helix motif defined in this paper is reminiscent of the classical DNA-binding helix−turn−helix (HTH; Harrison & Aggarwal, 1990). Extension of the HTH to RNA recognition has recently been suggested by its characterization in the unbound structure of a ribosomal protein (Xing *et al.*, 1997). The biological importance of HTH−RNA interactions is further suggested by its proposed role in modulating the classical operator-binding activity of the λ repressor (Retallack & Friedman, 1995) and the gene-regulatory activity of a *Drosophila* homeodomain (Rivera-Pomar *et al.*, 1996; Dubnau & Struhl, 1996). In neither case is the structural role of the HTH in RNA binding well characterized. Dual use of a structural motif in DNA and RNA recognition occurs among Zn finger proteins (Draper, 1995) but is also not well understood. We emphasize that the N motif differs from the classical HTH in four key respects. First, the N structure is induced on RNA binding and does not preexist in a stably folded glubular domain. Second, the bend does not conform to sequence patterns conserved in the HTH (Harrison & Aggarwal, 1990; Harrison, 1991). Third, no long-range NOEs occur between N- and C-terminal helices of the HBH to define a supersecondary structure. Finally, in the N motif both N- and C-terminal side chains make contacts to RNA bases whereas in the HTH base-specific contacts are restricted to the C-terminal helix. Despite these differences, it is intriguing to speculate that nucleic acid-binding motifs have their evolutionary origins in the transition between RNA and protein worlds (Joyce, 1991). The simplest proteins might have been autonomous peptide motifs, whose folding was directed by interaction with the RNA itself. We imagine that such motifs were later incorporated by gene assembly into larger proteins. This viewpoint suggests that ubiquitious ARM domains of ribosomal proteins, viral coat proteins, and transcription factors (Lazinkski *et al.*, 1989; Tan & Frankel, 1995) represent an early innovation of the RNA−protein world. The classical HTH element, now stably imbedded in the diverse tertiary structure of larger proteins (Harrison & Aggarwal, 1990), may represent the subsequent radiation in the DNA world of a primordial bipartite RNA-binding motif.

In summary, regulation of transcriptional elongation provides a major control point in prokaryotic operons (Yanofsky, 1988; Das, 1993). That analogous regulation exists in eukaryotes is supported by observation of intrinsic pause sites in nuclear proto-oncogenes and by the characterization of host and viral transcriptional elongation factors (Maldonado & Reinberg, 1995), including the Tat and Rev proteins of mammalian immunodeficiency viruses (Puglisi *et al.*, 1995; Battiste *et al.*, 1994, 1996; Ye *et al.*, 1995, 1996). The λ N ARM (Lazinski *et al.*, 1989; Tan & Frankel, 1995) and its interaction with *boxB* (Van Gilst *et al.*, 1997) provide a model of RNA-mediated transcriptional antitermination in the viral developmental program (Roberts, 1969). Characterization of a novel helix−bend−helix motif rationalizes genetic analysis of N-dependent transcriptional antitermination and extends the structural repertoire of arginine-dependent protein−RNA recognition.

ACKNOWLEDGMENT

We thank A. Das, N. Franklin, M. Gottesman, A. Jancsó, and S. Kron for discussion and D. N. Jones for assistance with NMR spectroscopy.

SUPPORTING INFORMATION AVAILABLE

Two figures showing (i) control fluorescence−quenching experiments and (ii) decrements in specific RNA binding on deletion of N- or C-terminal residues from peptide λ_{P1} and seven tables providing assignment of ^{1}H, ^{13}C, and ^{15}N chemical shifts of the bound peptides and selective complexation shifts (9 pages). Ordering information is given on any current masthead page.

REFERENCES

Allain, F. H., Gubser, C. C., Howe, P. W., Nagai, K., Neuhaus, D., & Varani, G. (1996) *Nature 380*, 646−650.
Barik, S., Ghosh, B., Whalen, W., Lazinski, D., & Das, A. (1987) *Cell 50*, 885−899.
Battiste, J. L., Tan, R., Frankel, A. D., & Williamson, J. R. (1994) *Biochemistry 33*, 2741−2747.
Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D., & Williamson, J. R. (1996) *Science 273*, 1547−1551.
Bax, A. (1994) *Curr. Opin. Struct. Biol. 4*, 738−744.
Chattopadhyay, S., Garcia-Mena, J., DeVito, J., Wolska, K., & Das, A. (1995a) *Proc. Natl. Acad. Sci. U.S.A. 92*, 4061−4065.
Chattopadhyay, S., Hung, S. C., Stuart, A. C., Palmer, A. G., III, Garcia-Mena, J., Das, A., & Gottesman, M. E. (1995b) *Proc. Natl. Acad. Sci. U.S.A. 92*, 12131−12135.
Chen, L., & Frankel, A. D. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 5077−5082.
Cilley, C. D., & Williamson, J. R. (1997) *RNA 3*, 57−67.
Das, A. (1993) *Annu. Rev. Biochem. 62*, 893−930.
Davis, R. H. (1995) *Curr. Opin. Biotechnol. 6*, 213−217 .
de Crombrugghe, B., Mudryj, M., DiLauro, R., & Gottesman, M. (1979) *Cell 18*, 1145−1151.
DeVito, J., & Das, A. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 8660−8664.
Dhalluin, C., Wieruszelski, J. M., & Lippens, G. M. (1995) *J. Biomol. NMR 5*, 327−331.
Doelling, J. H., & Franklin, N. C. (1989) *Nucleic Acids Res. 17*, 5565−5577.
Draper, D. E. (1995) *Annu. Rev. Biochem. 64*, 593−620.
Dubnau, J., & Struhl, G. (1996) *Nature 379*, 694−699.
Folmer, R. H. A., Hilbers, C. W., Konings, R. N. H., & Hallenga, K. (1995) *J. Biomol. NMR 5*, 427−432.
Franklin, N. C. (1985a) *J. Mol. Biol. 181*, 75−84.
Franklin, N. C. (1985b) *J. Mol. Biol. 181*, 85−91.
Franklin, N. C. (1993) *J. Mol. Biol. 231*, 343−360.
Friedman, D. I., & Olson, E. R. (1983) *Cell 34*, 143−149.

Greenblatt, J., Nodwell, J. R., & Mason, S. W. (1993) *Nature 364*, 401−406.
Hallenga, K., & Lippens, G. M. (1995) *J. Biomol. NMR 5*, 59−66.
Harrison, S. C. (1991) *Nature 353*, 715−719.
Harrison, S. C., & Aggarwal, A. K. (1990) *Annu. Rev. Biochem. 59*, 933−969.
Henthorn, K. S., & Friedman, D. I. (1996) *J. Mol. Biol. 257*, 9−20.
Hung, S. C., & Gottesman, M. E. (1995) *J. Mol. Biol. 247*, 428−442.
Jiang, F., Kumar, R. A., Jones, R. A., & Patel, D. J. (1996) *Nature 382*, 183−186.
Joyce, G. F. (1991) *New Biol. 3*, 399−407.
King, C. Y., & Weiss, M. A. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 11990−11994.
Lazinski, D., Grzadzielska, E., & Das, A. (1989) *Cell 59*, 207−218.
Li, J., Horwitz, R., McCracken, S., & Greenblatt, J. (1992) *J. Biol. Chem. 267*, 6012−6019.
Maldonado, E., & Reinberg, D. (1995) *Curr. Opin. Cell Biol. 7*, 352−361.
Mason, S. W., & Greenblatt, J. (1991) *Genes Dev. 5*, 1504−1512.
Mason, S. W., Li, J., & Greenblatt, J. (1992) *J. Biol. Chem. 267*, 19418−19426.
Mogridge, J., Mah, T.-F., & Greenblatt, J. (1995) *Genes Dev. 9*, 2831−2845.
Nelson, H. C., & Sauer, R. T. (1986) *J. Mol. Biol. 192*, 27−38.
Nodwell, J. R., & Greenblatt, J. (1991) *Genes Dev. 5*, 2141−2151.
Olson, E. R., Flamm, E. L., & Friedman, D. I. (1982) *Cell 31*, 61−70.
Otting, G., Qian, Y. Q., Billeter, M., Muller, M., Affolter, M., Gehring, W. J., & Wuthrich, K. (1990) *EMBO J. 9*, 3085−3092.
Patt, S. L. (1992) *J. Magn. Reson. 95*, 94−102.
Patterson, T. A., Zhang, Z., Baker, T., Johnson, L. L., Friedman, D. I., & Court, D. L. (1994) *J. Mol. Biol. 236*, 217−228.
Puglisi, J. D., Chen, L., Blanchard, S., & Frankel, A. D. (1995) *Science 270*, 1200−1203.
Redfield, C., & Dobson, C. M. (1990) *Biochemistry 29*, 7201−7209.
Retallack, D. M., & Friedman, D. I. (1995) *Cell* 83, 227−235.
Rivera-Pomar, R., Niessing, D., Schmidt-Ott, U., Gehring, W. J., & Jackle, H. (1996) *Nature 379*, 746−749.
Roberts, J. W. (1969) *Nature 224*, 1168−1174.
Salstrom, J. S., & Szybalski, W. (1978) *J. Mol. Biol. 124*, 195−221.
Schauer, A. T., Carver D. L., Bigelow, B., Baron, L. S., & Friedman, D. I. (1987) *J. Mol. Biol. 194*, 679−690.
Sklenar, V., Piotto, M., Leppik, R., & Saudek, V. (1993) *J. Magn. Reson. 102*, 241−247.
Smallcombe, S. H. (1992) *J. Am. Chem. Soc. 115*, 4776−4785.
Su, L., Radek, J. T., Labeots, L. A., Hallenga, K., Hermanto, P., Chen, H., Nakagawa, S., Zhao, M., Kates, S., & Weiss, M. A. (1997) *Genes Dev. 11*, 2214−2226.
Tan, R., & Frankel, A. D. (1994) *Biochemistry 33*, 14579−14585.
Tan, R., & Frankel, A. D. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 5282−5286.
Tao, J., & Frankel, A. D. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 1571−1575.
Van Gilst, M., Rees, W. A., Das, A., & von Hippel, P. H. (1997) *Biochemistry 36*, 1514−1524.
Varani, G., & Tinoco, I., Jr. (1991) *Q. Rev. Biophys. 24*, 479−532.
Vuister, G., & Bax, A. (1992) *J. Magn. Reson. 98*, 428−435.
Whalen, W. A., & Das, A. (1990) *New Biol. 2*, 975−991.
Whalen, W., Ghosh, B., & Das, A. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 2494−2498.
Wishart, D. S., & Sykes, B. D. (1994) *J. Biomol. NMR 4*, 171−180.
Wüthrich, K. (1886) *2D-NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York.
Xing, Y., Guha Thakurta, D., & Draper, D. E. (1997) *Nat. Struct. Biol. 4*, 24−27.
Yang, J. T., Wu, C.-S. C., & Martinez, H. M. (1986) *Methods Enzymol. 130*, 208−269.
Yanofsky, C. (1988) *J. Biol. Chem. 263*, 609−612.
Ye, X., Kumar, R. A., & Patel, D. J. (1995) *Chem. Biol. 2*, 827−840.
Ye, X., Gorin, A., Ellington, A. D., & Patel, D. J. (1996) *Nat. Struct. Biol. 3*, 1026−1033.

BI971408K