

## Antitermination in bacteriophage $\lambda$ The structure of the N36 peptide-*boxB* RNA complex

Manuela Schärpf, Heinrich Sticht, Kristian Schweimer, Markus Boehm\*, Silke Hoffmann and Paul Rösch

Lehrstuhl für Biopolymere der Universität Bayreuth, Germany

The solution structure of a 15-mer *nutRboxB* RNA hairpin complexed with the 36-mer N-terminal peptide of the N protein (N36) from bacteriophage  $\lambda$  was determined by 2D and 3D homonuclear and heteronuclear magnetic resonance spectroscopy. These 36 amino acids include the arginine-rich motif of the N protein involved in transcriptional antitermination of phage  $\lambda$ . Upon complex formation with *boxB* RNA, the synthetic N36 peptide binds tightly to the major groove of the *boxB* hairpin through hydrophobic and electrostatic interactions forming a bent  $\alpha$  helix. Four nucleotides of the GAAA pentaloop of the *boxB* RNA adopt a GNRA-like tetraloop fold in the complex. The formation of a GAAA tetraloop involves a loop-closing sheared base pair (G6-A10), base stacking of three adenines (A7, A8, and A10), and extrusion of one nucleotide (A9) from the loop, as observed previously for the complex of N(1–22) peptide and the *nutLboxB* RNA [Legault, P., Li, J., Mogridge, J., Kay, L.E. & Greenblatt, J. (1998) *Cell* 93, 289–299]. Stacking of the bases is extended by the indole-ring of Trp18 which also forms hydrophobic contacts to the side-chains of Leu24, Leu25, and Val26.

Based on the structure of the complex, three mutant peptides were synthesized and investigated by CD and NMR spectroscopy in order to determine the role of particular residues for complex formation. These studies revealed very distinct amino-acid requirements at positions 3, 4, and 8, while replacement of Trp18 with tyrosine did not result in any gross structural changes.

**Keywords:** bacteriophage lambda; antitermination; peptide–RNA complex; GNRA-tetraloop; NMR.

Bacteriophage  $\lambda$  N protein plays an essential role in transcriptional antitermination in the two phage early operons, which are critical for phage development. The inhibition of termination at intrinsic and rho-dependent terminators by N protein depends on recognition of an RNA-element called *nut* (N utilization) on the nascent phage transcript and on four *Escherichia coli* host factors (NusA, NusB, NusG, and ribosomal protein S10). Together they form a ribonucleoprotein complex that converts the RNA polymerase into a termination-resistant form upon binding [1,2].

The RNA enhancer elements (the *nut* sites *nutL* and *nutR*) are located on the DNA template between the early promoters and the target terminators of the *pL* and *pR* operons and consist of a 5′-single-stranded RNA element (*boxA*) and a 3′ hairpin (*boxB*) [3,4]. Genetic and biochemical studies suggest that the conserved sequence of *boxA* is recognized by a heterodimer composed of the host factors NusB and ribosomal protein S10 [5].  $\lambda$  *boxB* is a 15-mer RNA hairpin containing a purine-rich pentaloop (Fig. 1). Specific recognition of the five base pair stem and the five nucleotide loop of *boxB* by the N protein and the high affinity of the complex ( $K_d = 1.3$  nM) have been

shown previously [6,7]. RNase footprinting experiments and mutational data have revealed that the *boxB* hairpin forms the N-binding surface with its 5′ strand and the first, third, and fifth nucleotide of the loop [8–10]. A second functional surface is recognized by NusA in the assembly of the transcriptional antitermination complex [9]. The detailed mechanism of the antitermination system, however, is still not clear, although it is discussed that N protein in its complex with *boxB* RNA might prevent pausing of RNA polymerase [2].

The RNA-binding domain of N protein consists of an arginine-rich motif (residues 1–22) located at the NH<sub>2</sub>-terminus [7–11]. The arginine-rich motif of N protein binds to the *boxB* RNA hairpin with similar affinity ( $K_d = 20$  nM) and specificity as full-length N protein ( $K_d = 1.3$  nM). Mutations of critical positions within the *boxB* RNA lead to a more than 100-fold decrease in binding affinity for the N(1–22) peptide [10,11]. The small (107 residues) and very basic N protein of bacteriophage  $\lambda$  is disordered in solution according to CD spectroscopy, and its RNA-binding domain adopts its specific fold only upon binding to *boxB*. All structural changes occur exclusively within the N-terminal 36 residues of the N protein, while the other functional domains (NusA- and polymerase binding domains; residues 34–47 and 73–107) remain disordered [12,13].

The 3D structure of bacteriophage  $\lambda$  N(1–22) peptide in complex with *boxB* has recently been determined by NMR spectroscopy [14]. The solution structure of the complex exhibits a bent  $\alpha$  helix for the peptide, which binds exclusively to the 5′ strand of the *boxB* RNA and to the first three nucleotides of the loop. Four of the five residues of the *boxB* loop adopt a conformation very similar to that of a stable GNRA tetraloop [15] in which the fifth nucleotide is looped out. A similar structure has also been reported for the complex

Correspondence to P. Rösch: Lehrstuhl für Biopolymere der Universität Bayreuth, Universitätsstr. 30, 95447 Bayreuth, Germany.

Fax: + 49 921553544, Tel.: + 49 921553540,

E-mail: paul.roesch@uni-bayreuth.de

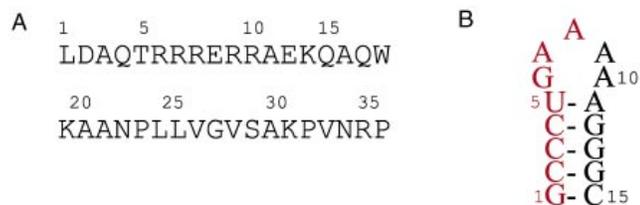
Abbreviations: *boxB*, *nutRboxB*, RNA from phage  $\lambda$ ; *nut*, N utilization.

\*Present address: National Institute for Child Health and Human Disease (NICHD), Cell Biology and Metabolism Branch (CBMB), National Institutes of Health (NIH), Bethesda, Maryland 20892, USA.

Note: web site available at <http://btcpxx.che.uni-bayreuth.de/>

(Received 23 November 1999, revised 23 February 2000, accepted

24 February 2000)



**Fig. 1.** Sequences of (A) the 36-mer N-terminal N peptide and (B) the 15-mer *nutR**boxB* RNA hairpin from bacteriophage  $\lambda$  used in this study. Nucleotides protected by N from RNase cleavage [8] are shown in red.

of a 20-mer basic peptide from the N protein of bacteriophage P22 in complex with its cognate 15-mer *boxB* RNA [16]. Both structures reveal a previously unknown mode of recognition of RNA and protein as a structural basis of transcriptional antitermination.

As peptides of 22 and 20 residues were used in both studies, little is known about the structural role of residues following the arginine-rich motif. Therefore, it is not clear whether these residues represent a simple 'linker' between the arginine-rich motif and the NusA binding domain of N or whether they are involved in RNA binding, either by direct interaction with *boxB* or indirectly by stabilizing residues that bind *boxB*.

Here we present the 3D solution structure of a 36-mer N36 peptide-*boxB* RNA complex from bacteriophage  $\lambda$  determined by heteronuclear NMR spectroscopy. In contrast to a previous structural study [14] that used the *nutL**boxB* site, the *nutR**boxB* site was used in the present work. The sites differ in one nucleotide of the *boxB* loop which is G9 in *nutL* and A9 in *nutR*.

## MATERIALS AND METHODS

### RNA synthesis

Unlabeled and  $^{13}\text{C}$ -,  $^{15}\text{N}$ -labeled 15-nucleotide *boxB* RNA were synthesized by *in vitro* transcription using T7 polymerase, a synthetic DNA template (5'-GCCCTTTTTCAGGGCTATAGT-GAGTCGTATTA-3') (MWG-BioTech, Ebersberg), and uniformly labeled nucleotide triphosphates (Campro scientific, Emmerich, Belgium). The RNA was purified as described previously [17,18]. Freeze dried *boxB* was resuspended in water, applied on a size-exclusion column (NAP, Pharmacia) for desalting, and freeze dried again.

### Expression and purification of N protein and N36 peptide

Unlabeled and  $^{15}\text{N}$ -labeled N protein and N(1–36) peptide (N36) were prepared from *Escherichia coli* BL21(DE3) grown in Luria–Bertani medium and minimal media with  $^{15}\text{NH}_4\text{Cl}$  as the only nitrogen source. N protein and N36 peptide were expressed as His-tag fusion proteins using the pET16b vector (Novagen). For the production of N36 peptide I37 of N products was changed to M37 using PCR-based site-directed mutagenesis, thus allowing cyanogen bromide cleavage at this position. Loss of Met1 during cleavage was compensated by introduction of Leu1.

Following lysis of the cells in buffer A (20 mM Tris/HCl, pH 8.0, 500 mM NaCl, 5 mM imidazol, 6 M urea) and centrifugation for 30 min at 100 000 g, the supernatant was added to a nickel-chelating column for purification (Chelating Sepharose, Fast Flow, Pharmacia). Fractions containing N protein and N36 peptide were then cleaved from the fusion

protein with cyanogen bromide (10 mg·mL<sup>-1</sup> per methionine) in 70% (v/v) formic acid at room temperature in the dark for 24 h. The mixture was freeze dried, resuspended in buffer A and applied to another nickel-chelating column. In the case of the N protein, the final purification step was a reversed-phase HPLC using a C<sub>18</sub> column (Macherey and Nagel). For the N36 peptide, the N-terminal cleavage product was purified from the C-terminal fragment and residual uncut N protein by RP-HPLC using a C<sub>4</sub> column (Vydag). The resulting N-terminal fragment eluted from the HPLC column as a single peak. Its molecular mass was checked by SDS/PAGE and MALDI-TOF MS.

The mutant peptides *mutW18Y*, *mutA3V* and *mutChi* (D2N/Q4K/R8H) were prepared by solid-phase synthesis (ZMBH, Heidelberg, Germany) and purified by RP-HPLC as described previously [19]. Fidelity of synthesis was verified by ESI-MS. *MutChi* represents a chimeric mutant that contains residues 1–8 from N-protein of bacteriophage P22 and residues 9–36 of N-protein of bacteriophage  $\lambda$ .

### Sample preparation

Complexes between N protein or N peptides and *boxB* were generated by gradual addition of concentrated ( $\approx$  4–8 mM) protein or peptide to *boxB* ( $\approx$  2–3 mM) with the stoichiometry monitored by following the intensity and number of the imino protons in the 1D NMR spectrum of free and bound RNA which are in slow exchange on the NMR time scale. Sample concentrations used for NMR spectroscopy were 2.2 mM for the ( $^{15}\text{N}$  N36)-*boxB* RNA complex, 2.0 mM for the N36 peptide-( $^{13}\text{C}$ ,  $^{15}\text{N}$ -*boxB*) complex and 1.0 mM for all unlabeled complexes of N36 peptide or mutant peptides and *boxB* in H<sub>2</sub>O/D<sub>2</sub>O (9 : 1, v/v, pH 6.5).

### CD spectroscopy

Far-UV CD spectra (190–250 nm) were recorded on a Jasco J600A spectropolarimeter in thermostatically controlled cells at 25 °C. Concentrations of N protein and N *boxB* complex were 10  $\mu\text{M}$ . Concentrations of the wild-type and mutant N36 peptides and their complexes with *boxB* were 40  $\mu\text{M}$ . For all spectra at least five scans were accumulated in water at pH 6 in a 0.05-cm cell. The CD spectra of the N protein and the peptides bound to the *boxB* RNA were determined by subtracting the spectrum of free *boxB* RNA at the appropriate concentration from the CD spectra of the complex.

### NMR spectroscopy

All NMR experiments were recorded at 25 °C on Bruker DRX 600 and DMX 750 spectrometers equipped with  $^1\text{H}/^{13}\text{C}/^{15}\text{N}$  probes and triple-axis pulsed-field gradient capabilities. Quadrature detection in the indirectly detected dimensions was obtained by the States-TPPI method [20] or the echo/antiecho-method [21–23], if coherence selection with gradients was employed.

Water suppression in experiments involving a  $^{15}\text{N}$  dimension was performed with the 3-9-19 binomial WATERGATE sequence [24]. Homonuclear double quantum filtered COSY [25] and all spectra involving a  $^{13}\text{C}$  dimension were recorded with coherence selection by magic angle gradients. The excitation sculpting module was used for water suppression in homonuclear  $^1\text{H}$ - $^1\text{H}$  TOCSY and NOESY experiments [26]. The DIPSI-2rc [27] sequence was used for proton TOCSY mixing and the DIPSI-2 [28,29] sequence was applied for  $^1\text{H}/^{13}\text{C}$  cross polarization and  $^{13}\text{C}$  TOCSY mixing. For

heteronuclear decoupling during acquisition the GARP sequence was used [30].

Homonuclear  $^1\text{H}$ - $^1\text{H}$  experiments were recorded at 600 MHz for the N36 mutants.

Data processing consisted typically of SVD-linear prediction with root-reflection in one heteronuclear dimension [31,32], apodization with 60–90° shifted squared sinebells, one zero-filling in all dimensions, and Fourier transformation. For constant time evolution periods mirror image linear prediction was employed [33]. Finally, baseline correction in the acquisition dimension was performed using a model free algorithm [34]. All NMR data were analyzed with the program packages NDEE and XNDEE (SpinUp Inc., Dortmund, Germany) augmented with in-house written routines.

Proton chemical shifts were referenced to external 2,2-dimethyl-2-silapentanesulfonate. The chemical shifts of  $^{13}\text{C}$  and  $^{15}\text{N}$  resonances were referenced indirectly using the  $\Xi$  ratios of the zero-point frequencies at 298 K: 0.10132905 for  $^{15}\text{N}/^1\text{H}$  and 0.25144952 for  $^{13}\text{C}/^1\text{H}$  [35].

### Experimental restraints for structure calculation

Intramolecular interproton distance restraints were obtained from 2D and 3D NOESY spectra recorded with mixing times of 75 ms and 150 ms. The intermolecular NOEs between *boxB* RNA and N36 peptide were identified using an indirect method of verification: all non peptide-peptide and non RNA-RNA NOEs were assumed to result from intermolecular contacts between peptide and RNA. These assignments were confirmed from the 3D NOESY spectra collected from a complex consisting of one unlabeled and one labeled component ( $^{15}\text{N}$ -labeled N36 or  $^{15}\text{N}$ ,  $^{13}\text{C}$  double labeled RNA, respectively). NOE intensities were estimated semiquantitatively on the basis of cross-peak intensities from spectra collected at 75-ms mixing time. The categories 'strong', 'medium' and 'weak' were converted in upper limit distance constraints of 3.0, 4.0 and 5.0 Å [36]. NOEs that were only visible in NOESY spectra with 150 ms mixing time were classified as 'very weak' with upper bounds of 6.0 Å or 7.0 Å [37]. In order to improve the convergence of the structure calculation, the lower bounds of all NOE restraints were set to 0 Å [38].

Presence of A-U and G-C Watson-Crick base pairs in the *boxB* stem and a sheared G-A base pair were confirmed from slow exchange of the hydrogen bonded imino protons, their downfield shifts, and their characteristic NOE pattern [39]. Canonical distance restraints were employed to define the hydrogen-bonding pattern and the planarity of these base pairs [40,41]. For all sugar puckers a C3'-endo conformation was deduced from the NOE pattern [42,43].

$^3J_{\text{HN}\alpha}$  coupling constants for the peptide were obtained from a 3D HNHA experiment [44].  $^3J_{\text{HN}\alpha}$  values < 6 Hz were converted to a  $\Phi$  angle of  $-60^\circ$  allowing deviations of  $\pm 20^\circ$  from the derived angle.

### Structure calculations

The structure of the complex was calculated by restrained molecular dynamics using the XPLOR package (v3.851) [45] and a calculation strategy similar to the one outlined previously [38].

A total of 100 initial structures with random  $\Phi$  and  $\Psi$  torsion angles for the peptide and with random backbone ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ , and  $\zeta$ ) and  $\chi$  angles for *boxB* were generated starting from random atomic coordinates [46,47]. The unassigned first (Leu1)

and last residue (homoserine 37) of the peptide were not included in the calculations.

The following simulated annealing protocol was used to determine the global fold: In the first 15 ps, molecular dynamics simulation at 1000 K only NOEs and the hydrogen bonds were used as experimental restraints and were represented by a soft-square well potential function. Then, 14 cycles of 1 ps dynamics each were performed at 1000 K, while the force constant for repulsive van der Waals interactions and the asymptote slope for the NOE potential were gradually increased. Finally, the system was cooled from 1000 K to 300 K (29 cycles of 0.5 ps MD each) followed by a 1000-step energy minimization.

In the following refinement, dihedral angle restraints were included for the sugar pucker and the  $\Phi$  angles of the peptide, and planarity restraints were used to maintain the planarity of all bases, the canonical base pairs, and the guanidino groups of all arginines. A 25-ps simulation was carried out at 1000 K in the first step of the refinement. A square well potential was applied to all distance restraints ( $k_{\text{noe}} = 50 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^{-2}$ ), and the force constants for the dihedral angle restraints were gradually increased from 5 to 55  $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{rad}^{-2}$  in steps of 10  $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{rad}^{-2}$ . The system was cooled to 300 K during the next 25 ps, followed by a final 1000 step energy minimization. The timesteps for integration were 1 fs and 0.5 fs for the 'global fold' and refinement simulations, respectively.

A total of 29 refined structures with no NOE violation greater than 0.3 Å and no torsion angle violation greater than  $2^\circ$  were accepted and retained for analysis. Geometry of the structures, structural parameters, and elements of secondary structure were analyzed using the programs PROCHECK [48] and NUCPLOT [49]. For the graphical presentation of the structures, SYBYL 6.5 (Tripos Ass.) and INSIGHT II (Molecular Simulations, Inc.) were used. The coordinates have been deposited in the Protein Data Bank (code: 1qfq).

## RESULTS AND DISCUSSION

### CD spectroscopy

The minimum near 200 nm in the far UV-CD spectra of free N protein and N36 peptide is characteristic for a random coil conformation (Fig. 2), consistent with the result of NMR spectroscopic studies. Difference CD spectra revealed that complex formation of the N protein and of the N36 peptide with *boxB* RNA resulted in conformational changes (Fig. 2) similar to those observed in previous studies for the respective complexes of the N protein, N(1–22) peptide and N(1–36) peptide [12,50]. The minimum of molar ellipticity is shifted to 205 nm for the N protein and to 208 nm for the N36 peptide. New minima at 222 nm and positive values for the molar ellipticity at 190 nm indicate  $\alpha$ -helical secondary structure formation in the N protein and the N36 peptide. The content of helical secondary structure in the N protein is 22% ( $\approx 23$  residues) and in the N36 peptide 63% ( $\approx 23$  residues) as estimated according to Holzwarth and Doty [51]. Thus, the structural changes in the N protein which occur upon binding to *boxB* RNA are mainly localized within the 36 N-terminal amino acids of the N protein, as indicated previously [12].

### NMR spectroscopy

The proportion of arginine and lysine residues in the N protein is 22%, and basic residues are found throughout the N protein

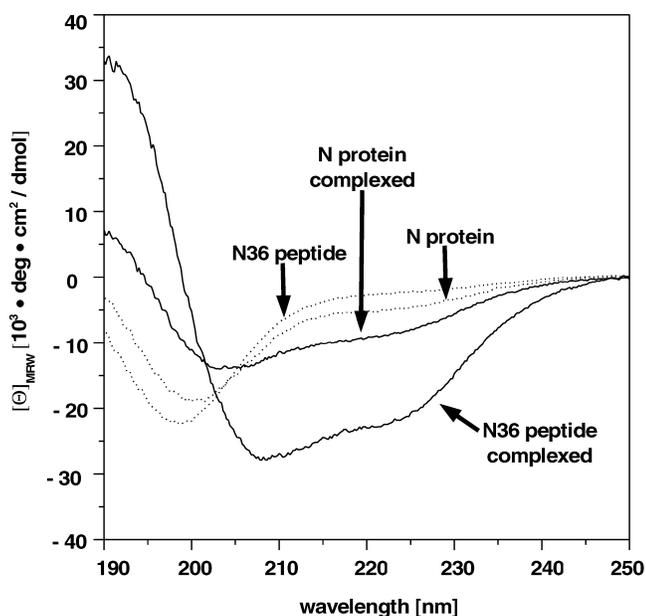


Fig. 2. Far-UV-CD spectra of N protein, N36 peptide, and their complexes with *boxB* RNA. The CD spectrum of 10  $\mu\text{M}$  N protein alone and in complex with *boxB* RNA minus the spectrum of RNA and the CD spectrum of 40  $\mu\text{M}$  N36 peptide alone and in complex with *boxB* RNA minus the spectrum of RNA are shown.

sequence. Therefore, numerous specific and nonspecific contacts between the positively charged side-chains of these amino acids and the negatively charged phosphodiester backbone of the RNA are possible, giving a likely explanation for complex aggregation at relatively low concentrations of 300  $\mu\text{M}$ . For this reason, a structure determination of the N-*boxB* complex by NMR spectroscopy was not possible, and NMR spectroscopic investigation was limited to the N36 peptide. During the titration of N36 peptide to *boxB* RNA, distinct signals for the imino protons of free and peptide-bound RNA were observed until equimolar concentrations of N36 peptide and *boxB* were reached, indicating that complex dissociation is slow on the NMR timescale. The most significant change in the chemical shifts of N36 peptide resonances was observed for the indole NH proton of Trp18 (10.12 p.p.m. vs. 9.31 p.p.m.; Fig. 3). Changes in the chemical shifts between the free and N36-bound *boxB* RNA were observed for the imino protons of G12, G13, and G14. For U5 and G6, new imino proton signals were visible upon complex formation (Fig. 3). The shifts and the new resonance signals for U5 and G6 indicate a stabilization of the apical U5-A11 base pair and suggest the formation of a novel base pair involving G6. In contrast, the absence of a signal for the imino proton of G1 indicates that no significant stabilization of the terminal base pair occurs on complex formation. These results support the proposed binding of the peptide to the upper part of the stem and the loop [8].

#### Assignment of the N36 peptide resonances and their analysis

The 3D  $^1\text{H}$ - $^{15}\text{N}$ -TOCSY-HSQC and NOESY-HSQC experiments allowed an almost complete assignment of all proton and nitrogen resonances of N36 peptide in complex with *boxB* RNA (data not shown). Elements of regular secondary structure were identified from the intensity of the short range NOEs, the  $^3J_{\text{HN}\alpha}$  coupling constants, and the chemical shift index (Fig. 4).

Patterns of short range NOEs  $d_{\alpha\text{N}}(i, i + 3)$  and  $d_{\alpha\text{B}}(i, i + 3)$ , typical for  $\alpha$  helices, were observed for residues 4–20. For the corresponding residues the NOE between sequential amide protons is always stronger than the sequential NOE between  $\text{H}_\alpha$  and amide protons. In addition, small  $^3J_{\text{HN}\alpha}$  coupling constants ( $< 6$  Hz) for 11 residues out of the first 20 amino acids and upfield chemical shifts for the  $\text{C}_\alpha$  protons give strong evidence for the presence of an  $\alpha$ -helix spanning residues Ala3 to Ala21 (Fig. 4). The chemical shift index and the pattern of medium range NOEs suggest an extension of the helix up to Leu25.

#### Assignment of the *boxB* resonances

Uniform  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeling of the *boxB* RNA allowed the assignment of all nonexchangeable protons and all carbon resonances of the peptide-bound *boxB* (data not shown). The assignment of the imino protons was based on sequential  $\text{H3}_i$ - $\text{H3}_{i+1}$  NOEs as well as on strong NOEs between the imino protons and the  $\text{NH}_2$  groups of the base-paired cytosines. These amino protons were identified by strong intrasidual NOEs to their own H5 protons. The assignment of the cytosine H5 and H6 protons from the 2D COSY was facilitated by their characteristic chemical shifts. The H5 proton from U5 was assigned using the different C5 carbon shift compared to the cytosines. The base and ribose protons were assigned using their typical intrasidual and sequential NOE pattern according to the strategy described in [52]. Proton and carbon assignments within each ribose spin system were deduced from 3D CCH-COSY, 3D HCCH-COSY, and 3D HCCH-TOCSY experiments. The connectivities between the ribose and their own as well as the sequential base were obtained from  $\text{H1}'$ -H6/H8 and  $\text{H1}'_{i+1}$  H6/H8 $_i$ -NOEs (Fig. 5). Interestingly, the NOE pattern of A8 suggests a direct connectivity to A10 (A8/ $\text{H1}'$ -A10/H8-NOE), indicative of an unusual conformation of A9.

#### Intermolecular NOEs

The observation of intermolecular NOEs indicates the presence of a well defined peptide-RNA interface in the complex. Fifteen intermolecular NOEs were observed between the methyl group of Ala3 and the nucleotides C2 and C3 (Fig. 6A). NOEs between the side-chain of arginines and base or sugar protons were detected as well, for example between

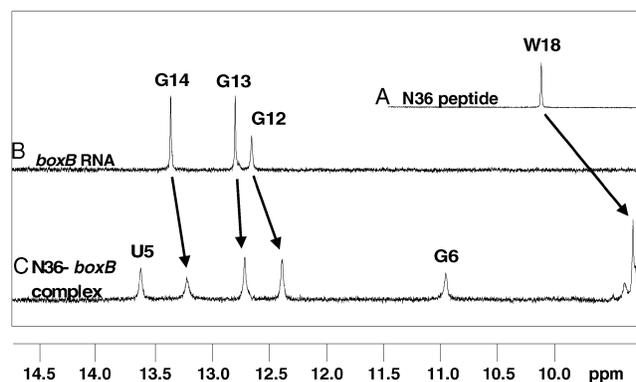
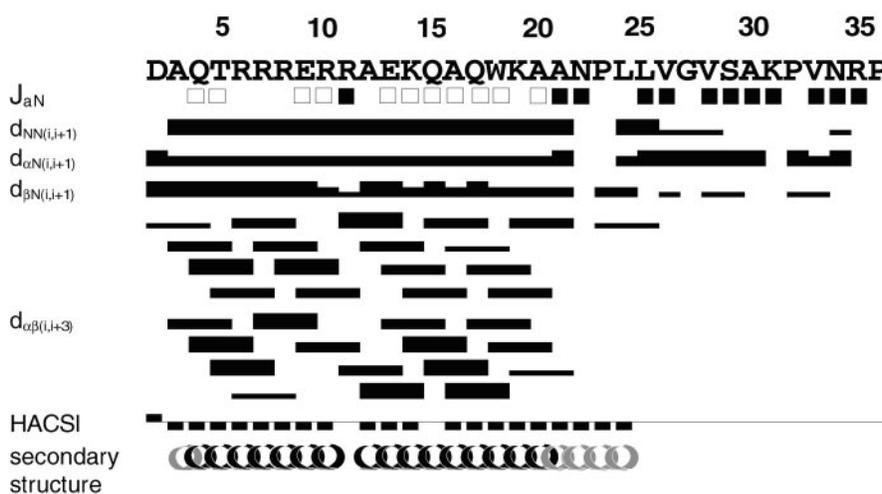


Fig. 3. Structural changes upon complex formation of N36 peptide and *boxB* RNA. 1D NMR spectra of the imino proton region of: (A) free N36 peptide (3.8 mM); (B) *boxB* RNA alone (750  $\mu\text{M}$ ); and (C) the complex between N36 peptide and *boxB* RNA (750  $\mu\text{M}$ ). All spectra were collected in  $\text{H}_2\text{O}$ , pH 6.5 at 25  $^\circ\text{C}$ .

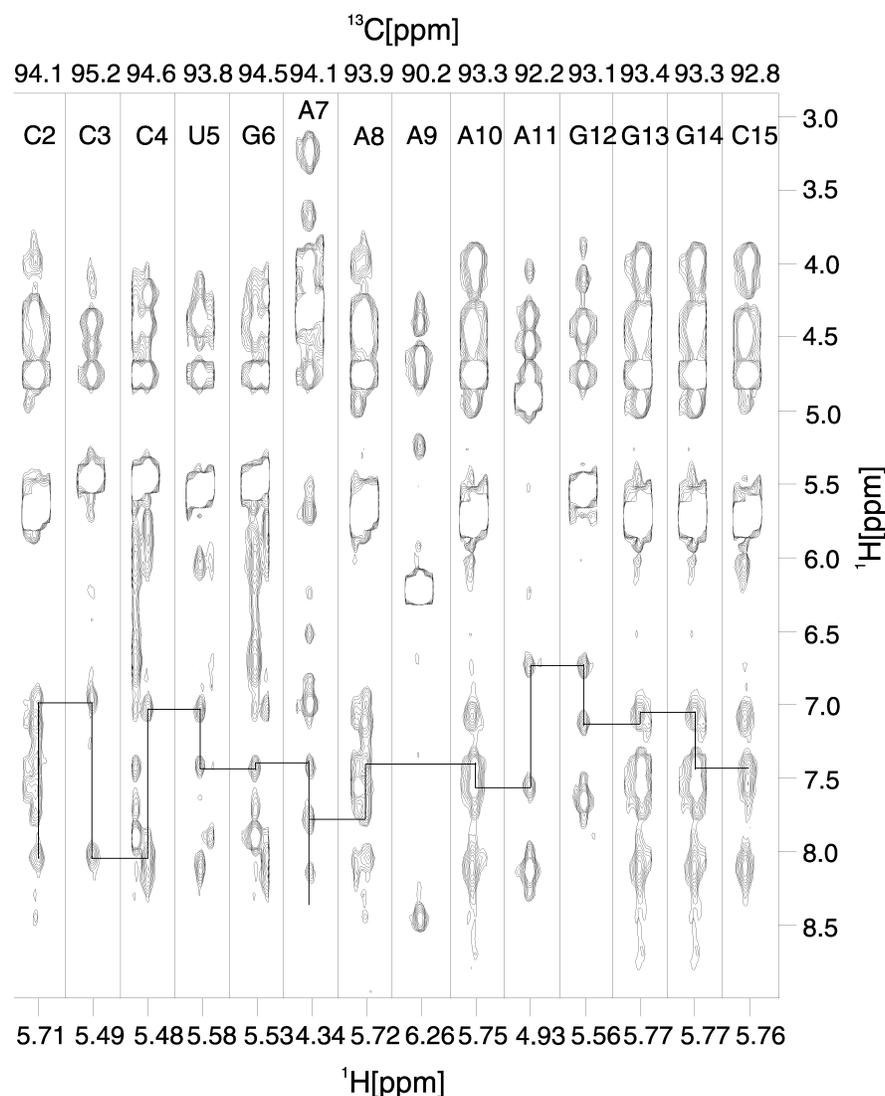
Fig. 4. Secondary structure of the N36 peptide. Summary of sequential and medium-range NOEs, vicinal backbone NH-C $\alpha$ H coupling constants ( $J_{\alpha N}$ ) and C $\alpha$ H chemical shift indices (HACSI) [70]. The relative strengths of the NOEs, categorized as strong, medium, and weak, by cross-peak intensity, are indicated by the width of the horizontal bars. Values of  $^3J_{HN\alpha} > 6$  Hz or  $\leq 6$  Hz are indicated by black and white squares, respectively. Positive and negative chemical shift indices are indicated by black rectangles above or below the axis, respectively. Regions of  $\beta$  sheet are typified by positive chemical shift indices, regions of  $\alpha$  helix by negative indices [70]. The secondary structure assignments are depicted in the bottom line.



Arg7 and C4 and C5, and between Arg8 and particularly the base protons from A8. Several intermolecular NOEs were observed between the side-chain protons of Gln15 and the ribose protons from A7 and A8, and between the side-chain

protons of Gln4 and the imino proton of G6. The side-chain protons of Lys14 and Lys 19 showed NOEs to the ribose of A7 and the base of A9, respectively. The largest number of intermolecular NOEs was detected between Trp18 and A7.

Fig. 5. Sequential connectivities between sugar H1' and base H6/H8 NOEs are traced in  $^1\text{H}$ - $^1\text{H}$  stripes at C6/C8 carbon chemical shifts taken from the  $^{13}\text{C}$ -edited NOESY-HSQC (150 ms mixing time) 3D data set of the complex with uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$  labeled *boxB* RNA. Nucleotides C2-C15 are shown.



**Table 1.** Experimental restraints for the structure determination of the N36 peptide-*boxB* RNA complex.

	Number
NOEs: total	1228
Peptide-intramolecular: total	490
Sequential $li - jl = 1$	209
Medium-range $li - jl = 2, 3, 4$	260
Long-range $li - jl > 4$	21
RNA-intramolecular: total	571
Intraresidual $li - jl = 0$	356
Sequential $li - jl = 1$	132
Medium-range $li - jl = 2, 3, 4$	37
Long-range $li - jl > 4$	46
Peptide-RNA: intermolecular	167
Dihedral angles	101
Peptide: phi angles	11
RNA: sugar-pucker <sup>a</sup>	90
RNA: hydrogen bonds <sup>b</sup>	32

<sup>a</sup> Six dihedral angle restraints were used for each *sugar-pucker*; <sup>b</sup> two distance restraints were used for each hydrogen bond.

Almost every proton of Trp18 is in close contact to a proton from A7 (Fig. 6B).

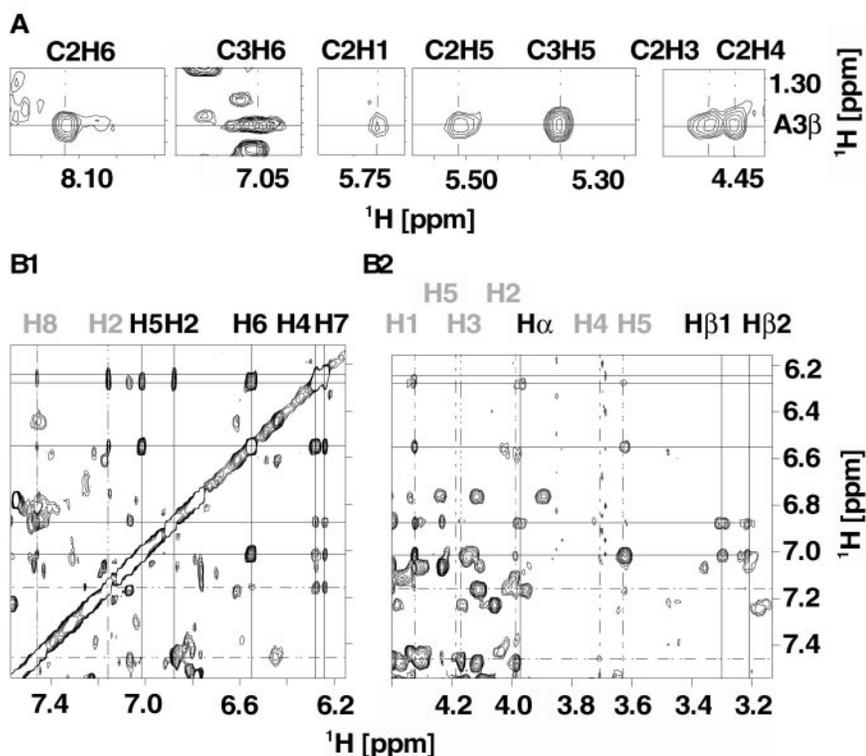
### Quality of the N36 peptide-*boxB* RNA complex structure

The calculation of the final structure was based on 1228 distance restraints, 101 dihedral angle restraints and 32 hydrogen-bond restraints (Table 1). A subset of 29 structures

with energies lower than  $440 \text{ kcal}\cdot\text{mol}^{-1}$  out of 100 structures was selected for further analysis. None of these structures showed NOE violations larger than  $0.3 \text{ \AA}$  or violations larger than  $2^\circ$  for the dihedral angles. The structure of the complex is well defined for the RNA and for the 27 N-terminal residues of the peptide (Fig. 7) with a rmsd of  $0.82 \text{ \AA}$  for all heavy atoms (residues G1-C15 and Asp2-Gly27). The C-terminal residues 28–36 were excluded from the calculation of the rmsd value, because they remain disordered after binding of the *boxB* RNA. Calculation of the rmsd for residues Ala3 Leu25 and C2-G14 only (Table 2) results in a decrease of the rmsd to  $0.74 \text{ \AA}$  for all heavy atoms of the complex, which can be attributed to the increased flexibility of the terminal base pair (G1-C15) and of residue Asp2. This rmsd value is of the same magnitude as for other peptide-RNA complexes of this size [14,16].

### Structure of N36 peptide in the complex

According to a PROCHECK analysis [48] of residues Asp2–Gly27 in the 29 accepted structures, almost all residues adopt energetically favourable backbone conformations: 83.4% of the residues are found in the most favoured regions, 15.9% in the allowed regions and only 0.7% in the additionally allowed regions of the Ramachandran plot. The N36 peptide forms an  $\alpha$  helix for residues Gln4–Ala20 which is bent at Arg11. The bend of roughly  $120^\circ$  does not require large deviations from ideal  $\alpha$  helix  $\Phi/\Psi$  angles. In addition, a turn or short helix was observed for residues Pro23–Val26, placing the side-chains of the hydrophobic amino acids Leu24, Leu25, and Val26 in close proximity to the aromatic ring of Trp18 (Fig. 7) as shown from 13 NOEs observed between those residues. The rmsd value for the backbone and all heavy atoms of the peptide (residues Ala3–Gly27) is  $0.23 \text{ \AA}$  and  $0.76 \text{ \AA}$ , respectively (Table 2).



**Fig. 6.** Intermolecular NOEs between N36 peptide and *boxB* RNA in the complex. NOEs between (A) the methyl group of Ala3 and nucleotides C2 and C3, (B1) the base of A7 (grey) and the indole ring protons of Trp18 (black) and (B2) additional NOEs of this protons to the ribose of A7 (grey) and the side-chain protons of Trp18 (black). Broken lines indicate resonances of the *boxB* RNA and solid lines resonances of the peptide.

**Table 2. Results of the structure calculations.** All values are averages over the family of 29 converged structures. Values in parenthesis indicate the standard deviation from this mean.

Average energies	Energy (kcal·mol <sup>-1</sup> )
$E_{\text{tot}}$	439.50 ( $\pm$ 13.28)
$E_{\text{bond}}$	18.46 ( $\pm$ 0.78)
$E_{\text{angle}}$	157.97 ( $\pm$ 4.41)
$E_{\text{improper}}$	54.72 ( $\pm$ 1.26)
$E_{\text{repel}}$	108.17 ( $\pm$ 7.64)
$E_{\text{NOE}}$	96.83 ( $\pm$ 4.98)
$E_{\text{dihedral}}$	2.33 ( $\pm$ 0.31)
$E_{\text{plane}}$	1.01 ( $\pm$ 0.74)
Rmsd from ideal distances	rmsd (Å)
NOE	0.039
Bond length	0.004
Rmsd from ideal angles	rmsd (°)
Bond angles	0.706
Improper angles	0.752
Dihedral angles	0.174
Rmsd among the family of 29 structures	rmsd (Å)
Peptide: backbone of Ala3–Gly27	0.23
Peptide: all heavy atoms of Ala3–Gly27	0.76
RNA: heavy atoms of G1–C15	0.76
RNA: heavy atoms of C2–G14	0.67
Complex: backbone of Ala3–Gly27 and G1–C15	0.73
Complex: all heavy atoms of Ala3–Gly27 and G1–C15	0.82

### Structure of *boxB* in the complex

Analysis of the torsion angles of the RNA [53] showed that all nucleotides of the stem adopt an A helical structure with 3'-endo conformation for the sugar puckers. As already observed in the N(1–22)/*boxB* complex [14], the GAAA pentaloop adopts a GNRA tetraloop conformation upon binding to the N36 peptide. The base of one loop nucleotide (A9) is flipped outside and the torsion angles for A7 are exactly the same as observed for the second nucleotide in a GNRA tetraloop (Fig. 8). As in the GAAA tetraloop structures [41,55,56], a sheared base pair (G6 A10) is formed between the first and last loop nucleotides in the *boxB* hairpin. This leads to a twisted phosphate backbone between nucleotide G6 and A7 and sequential stacking of the three purines (A7, A8, and A10) excluding the fourth nucleotide A9 (Fig. 8). The rmsd values for the bases of nucleotides G1–C15 and C2–G14 are 0.76 Å and 0.67 Å, respectively (Table 2).

### Interactions in the N36 peptide-*boxB* RNA complex

As expected from the large number of intermolecular NOEs, the interface between protein and RNA is well-defined in the complex.

The tight fit of the N36 peptide to the RNA becomes apparent from a representation that shows the Connolly surface of the RNA (Fig. 9A). The N36 peptide binds to the loop and to the upper part of the helical stem, including only the nucleotides on the 5'-site of the loop.

A detailed analysis of the intermolecular interactions (Fig. 9B) shows that most contacts to the RNA are observed for amino acids Ala3, Gln4, Arg7, Arg8, Arg11, Lys14, Gln15, Trp18, and Lys19.

The side-chains of the arginines and lysines form a positively charged surface that interacts with the negatively charged groups of the phosphodiester backbone of the RNA. Electrostatic interactions were observed between all arginines and lysines in the sequence region 2–20 and the *boxB* phosphodiester backbone. The fact that some of these interactions are not observed in each of the calculated structures can mainly be attributed to the calculation strategy in which no term for electrostatic interactions was included in the potential function used [56]. Despite this limitation, it is clear from the structures that fewer intermolecular contacts are present for Arg6 and Arg10 than for Arg7, Arg8 and Arg11 (Fig. 9B). The mutational study of Franklin *et al.* [57] showed that arginines 7, 8, and 11 are more important than arginines 6 and 10 for antitermination, as confirmed by the results of Su *et al.* [58]. The binding affinity *in vitro* is reduced by approximately 80% for changing Arg7, 8, and 11 to alanine but only by 30–40% for changing Arg6 and 10 [58]. The results of the present work thus confirm previous biochemical [58] and structural [14] studies by revealing that arginines 7, 8 and 11 play a dominant role for the process of specific RNA recognition by the N protein.

Formation of hydrogen-bonds is observed between the guanidino group of Arg7 and the phosphate of U5 as well as the base of G6, and between the guanidino group of Arg8 and the base of A8. For the side-chain of Arg11, electrostatic

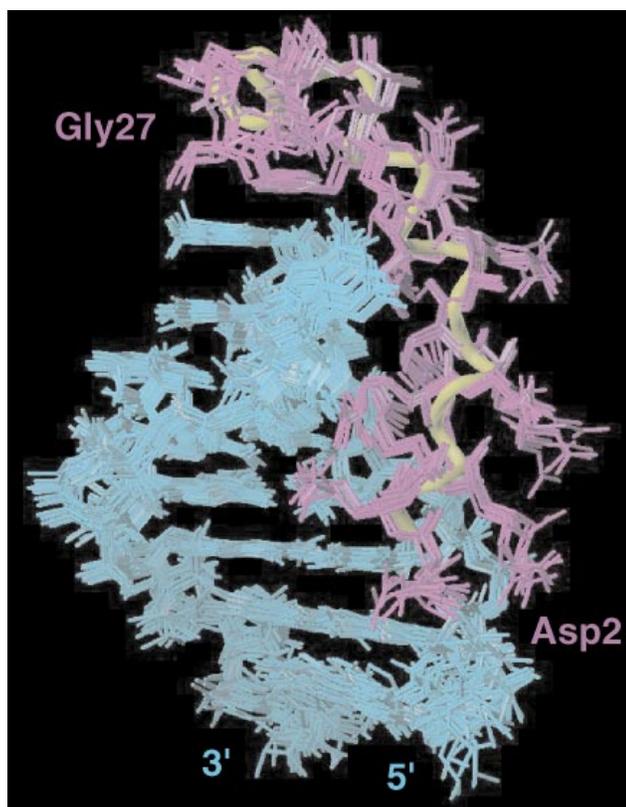


Fig. 7. The structure of the N36 peptide-*boxB* RNA complex. Superposition of a family of 20 complex structures. The structures were accepted by the criterion of best energy function after restrained molecular dynamics calculation. Only the heavy atoms are shown here. The *boxB* RNA is blue, the N36 peptide is pink with a yellow tube for the backbone.

interactions with the phosphate of U5 are found (Fig. 9B). Hydrogen-bonds are present between the side-chain amide group of Gln4 and the O6 atom of G6. The intermolecular hydrogen-bonds observed between Gln4 and Arg7 and the nucleotides U5 and G6 (Fig. 9b) emphasize that an apical U-A base pair and the formation of a GAAA tetraloop with a sheared G-A base pair play an important role for the specific recognition of *boxB* by N. Mutations in this region of the RNA reduce the binding affinity up to 20-fold [10,11].

Remarkably, Gln15 showed numerous hydrophobic interactions with nucleotides A7 and A8, and to the aliphatic methylene groups of the Arg11 side-chain. However, no conservation of a Gln was found at this position [57], nor was there a loss of affinity changing Gln15 to an alanine [58]. These findings suggest that stabilization of the bent  $\alpha$  helix by hydrophobic intramolecular and intermolecular interactions could be the function of Gln15, rather than conferring specificity to the complex formation by hydrogen bonds or electrostatic interactions. Numerous hydrophobic contacts are observed for the methyl group of Ala3 that packs tightly between the bases C2 and C3, thereby mainly contacting the carbons C5 and C6 of the base moiety. The stacking of the aromatic ring Trp18 onto the base of A7 and the bend in the  $\alpha$  helix are additional determinants for the surface complementarity between protein and RNA.

Residues 23–27 following the arginine-rich motif do not show any direct interactions with the RNA but exhibit numerous interpeptide contacts to residues of the arginine-rich motif (mainly to Trp18) implying a possible role of these residues for the stabilization of the structure by forming of a ‘cap’ for the C-terminus of the arginine-rich motif. Residues 28–36 are disordered in the complex, suggesting that these residues play a role as a linker between the RNA (residues 1–22) and the NusA (residues 34–47) binding domain of N.

#### Accuracy of the N36 peptide-*boxB* RNA complex structure

The direct comparison of the current family of 29 structures with the minimized average structure of the N22-*boxB* complex [14] reveals a deviation of  $1.5 (\pm 0.10)$  Å for the heavy atoms of Asp2 to Asn22 and G1 to C15. Exclusion of the terminal G1-C15 base pair from calculating the rmsd results in an rmsd of  $1.4 (\pm 0.09)$  Å. These minor differences can mainly be attributed to limits in the resolution of both structures, and slight differences in the orientation of the G1-C15 base pair can be explained by the fact that the RNA in the complex with N(1–22) contained two additional base pairs [14].

No direct interactions with the RNA are observed for residues 20–36. We observed hydrophobic contacts between the Lys19 side-chain and the base of A9 rather than with the ribose of G9 [14], variation that may either be attributed to the different nature of the nucleotide or to a possible role of the helical turn formed by residues 23–26 in stabilizing the

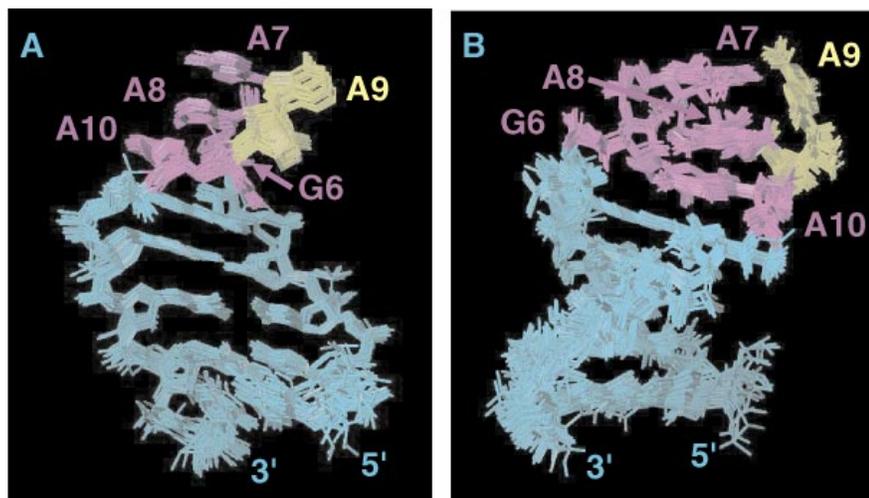
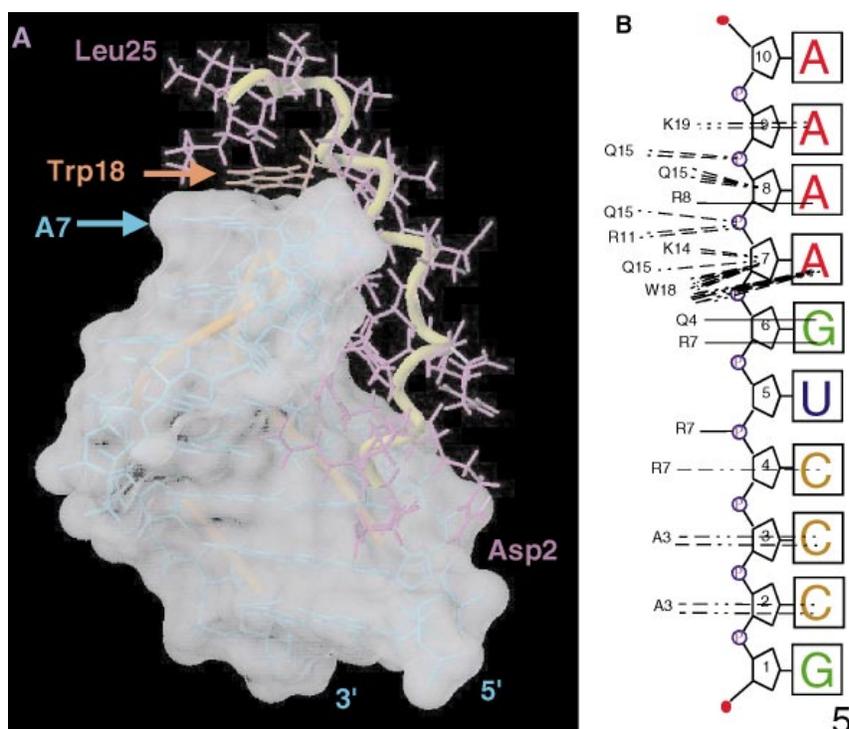


Fig. 8. The structure of the *boxB* RNA hairpin in the complex. Twenty structures with the lowest energies of *boxB* RNA are superimposed. The stem is coloured blue, the extruded nucleotide (A9) is coloured yellow and the four nucleotides forming the GNRA fold are shown in pink. In (A) only the heavy atoms are shown to highlight stacking of A7, A8 and A10 onto the 3' strand of the stem; (B) is rotated by 90° to show the GAAA tetraloop, the flipped out nucleotide and the formation of a sheared base pair between G6 and A11 in more detail.

**Fig. 9. Interactions between RNA and N36.**

(A) To highlight the tight fit of the structures, the RNA is shown as a transparent Conolly surface and the N36 peptide backbone as a yellow tube, respectively. The bend in the  $\alpha$  helix and the stacking of Trp18 onto A7 are visible. Binding of N36 peptide occurs to the upper part of the helical stem and to four nucleotides of the loop. Direct contacts were observed only for the 5' nucleotides of the RNA. (B) A schematic summary of intermolecular interactions between *boxB* RNA and N36 peptide. Only nucleotides involved in direct contacts to N36 peptide are shown. Bases are represented by rectangles, sugars by pentagons, and phosphate groups by small circles. Intermolecular electrostatic and hydrophobic interactions are indicated by dashed lines; intermolecular hydrogen-bonding interactions are indicated by solid lines, as determined by analyzing the structures with NUCPLOT [49].



orientation of the Lys19 side-chain. The overall similar orientation of the looped out nucleotide, however, underlines that the sequence differences between the *boxB* RNAs of the nutL and the nutR site have no major influence of the respective complex structure with  $\lambda$  N peptides.

In this case of the  $\lambda$  N/nut system, the rmsd of 1.5 Å is in the range observed for other high-resolution structures determined by X-ray and NMR spectroscopy, for example 1.9 Å for the structural elements of the complex between U1A protein and PIE RNA [56,59], underlining that it is possible to determine structures of peptide RNA complexes of this size reliably by NMR spectroscopy.

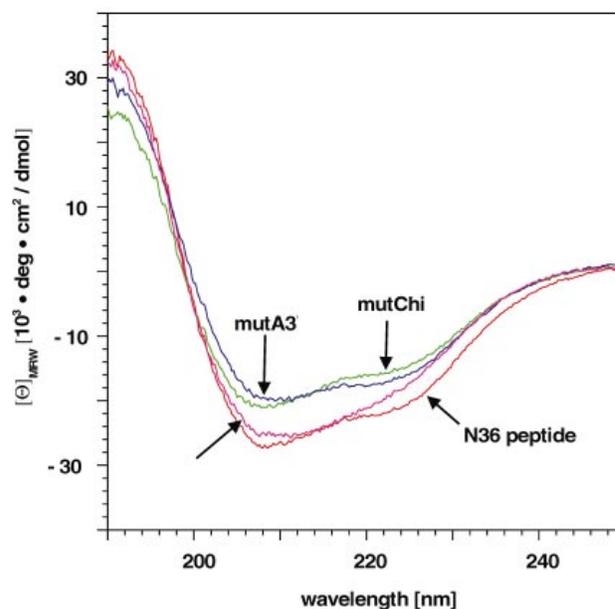
### Mutations in the RNA-binding region of N36 peptide

The influence of mutations in the RNA-binding region of the N36 peptide was assessed by three mutant peptides, mutA3V, mutW18Y, and mutChi. mutA3V and mutW18Y were designed because structure determination of the complex of wt-N36 peptide with *boxB* suggested both residues to be crucial for complex formation. The chimeric mutChi peptide contains residues 2–8 of P22 N-protein and residues 9–36 of  $\lambda$  N (MNAKTRRHERRAEKQAQWKAANPLLVGVSAPVNRP). It was synthesized to obtain information about the different sequence requirements in the arginine-rich motif of the lambdaoid phages  $\lambda$  and P22.

The structure of the mutant peptides and their complexes with *boxB* were investigated by CD and homonuclear 1D and 2D NMR spectroscopy. For all free peptides in solution the CD signal indicated the absence of stable elements of regular secondary structure. After binding to *boxB* RNA a significant change of the CD spectra was detected for all peptides including shifts of the minima from 200 nm to 208 nm, new minima around 222 nm, and positive values for the ellipticity at 190 nm (Fig. 10). The estimated helical contents were 63% for wt-N36, 57% for mutW18Y, 49% for mutChi, and 46% for

mutA3V. Homonuclear 2D NMR spectra of the complexes of mutA3V and mutChi revealed that the  $\alpha$  helix is shortened at the N-terminus.

Structural changes in the RNA by titration with the peptides were monitored by 1D NMR spectroscopy following the changes in the shifts of the imino proton resonances. (Fig. 11). The NMR and the CD spectra for the W18Y peptide-*boxB* RNA complex were essentially identical to the spectra of



**Fig. 10. Differential far-UV-CD spectra of N36, and mutant peptides in complex with *boxB* RNA.** The CD spectrum of the complex with N36 peptide is red, with mutW18Y pink, with mutA3V blue, and with mutChi (D2N/Q4K/R8H) green, respectively.

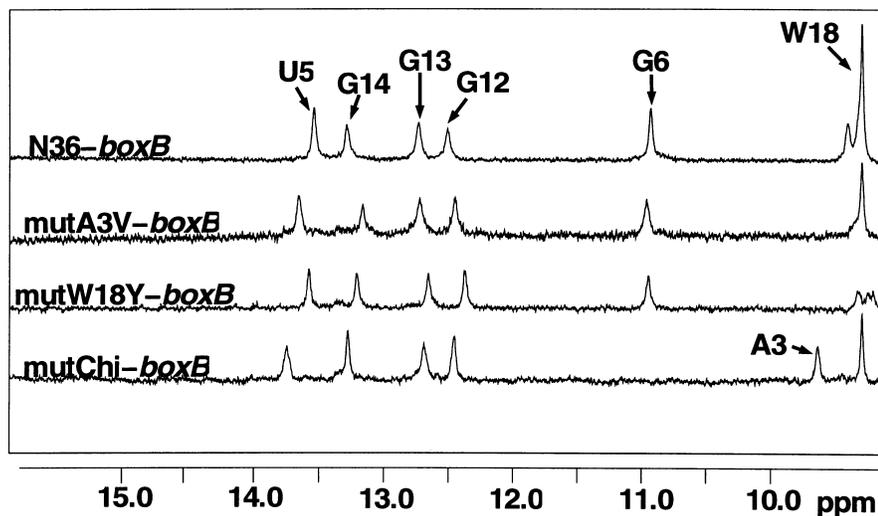


Fig. 11. 1D NMR spectra of the 1 : 1 complexes between *boxB* RNA, N36, and the mutant peptides. The resonances of the imino protons of *boxB* RNA, the amide proton of Ala3 and the imino proton of Trp18 are marked in the spectra. The concentration of the complexes with *boxB* RNA are: N36 peptide, 330  $\mu$ M; mutA3V, 260  $\mu$ M; mut W18Y, 250  $\mu$ M; mutChi, 235  $\mu$ M.

the corresponding wt-complex, suggesting similar structures for these complexes. Thus, aromatic stacking interactions are the major contribution of Trp18 to complex stability. The functional and structural importance of an aromatic residue (Trp or Tyr) at this position was suggested from the observation of antitermination in a W18Y mutant [57] and binding affinity studies [58]. Phage HK022 Nun protein, containing a Tyr residue corresponding to Trp18 of phage  $\lambda$  N-protein binds *boxB* from phage  $\lambda$  as an N protein competitor [60].

It is known from mutational studies of arginine-rich motifs in lambdoid phages [57] that position three is conserved as alanines and a serine in the case of the Nun protein from HK022. The results of our CD and NMR studies of mutA3V revealed the lack of specific contacts between Val3 and the nucleotides C2 and C3, resulting in an N-terminally shortened  $\alpha$  helix. Although the 2D NMR spectra confirm the presence of a GAAA tetraloop and an A helical stem, the observation of line broadening for the imino proton resonance of G6 reveals that mutation of Ala3 does affect both the length of the  $\alpha$ -helix, and also the stability of the sheared base pair G6-A10. This finding can most probably be attributed to an increased flexibility of Gln4, which interacts with G6 in wt-N36 peptide and is no longer fixed after A3V substitution. Thus, Ala3 plays an important role as an anchor for the N-protein to the stem of *boxB* RNA, explaining the strong decrease of binding affinity by A3S, A3V and A3T substitution [58].

The conformation is much more severely affected in the mutChi-*boxB* complex. The helix is disturbed near the N-terminus, and the loop structure of *boxB* is different. The reason for the missing imino proton signal of G6 (Fig. 11) is probably the lack of stabilizing contacts from residue Gln4 and Arg7 to the sheared base pair, resulting from the Q4K substitution and the direct neighborhood of Arg7 to the R8H mutation. These mutations are also the reason for the disturbed helix, indicated by missing  $\text{NH}_i\text{-NH}_{i+1}$  NOEs. These results show that the presence of the structurally important residues Ala3 and Trp18 itself is not sufficient to maintain the correct complex structure. The hydrogen bonds formed between Gln4 and the base of G6 and between Arg8 and the base of A8 obviously play an important role for the specific recognition in the N-*boxB* complex of the bacteriophage  $\lambda$ .

In spite of the similar overall structure of the complexes, Lys4 and His8 from the N protein of P22 cannot substitute for Gln4 and Arg8 of  $\lambda$  N, underlining the importance of

hydrogen-bonds for sequence specific recognition in this protein-RNA complex.

#### Comparison with other RNAs and protein-RNA complexes

Formation of a GNRA tetraloop within a pentaloop has been found previously for the *boxB* RNAs from phage  $\lambda$  and P22 in the complex with their respective recognition peptides [14,16]. Taking into account that one nucleotide (A9) has to be extruded, the rmsd of approximately 1.5 Å between the GNRA loops in the P4-P5-P6 *Tetrahymena* Group I Intron (G150-G153) [41] and the GNRA tetraloops formed within the pentaloop of *boxB* RNAs [14,16] (this work), is remarkably low. Mutational studies and deletion experiments [9] showed the importance of the loop structure for both specific recognition by the N protein and also for the binding of NusA.

Formation of a GNRA tetraloop in *boxB* RNA from phages  $\lambda$  and P22, however, may not be the general way of specific RNA N protein recognition in lambdoid phages. Phage  $\Phi$ 21, for example, exhibits huge differences in the *nut* sequence compared to  $\lambda$  and P22. This is most evident from the presence of a pyrimidine-rich hexaloop in contrast to the purine-rich pentaloop found in  $\lambda$  and P22. Although the formation of a GNRA tetraloop from a larger loop is known, for example a GNR(NN)A motif in the 16S and 23S RNA [61], this is not expected for the pyrimidine rich hexaloop from phage  $\Phi$ 21. Therefore, the existence of a common structural motif in the *boxB* RNA structure within the family of lambdoid phages remains to be clarified by structure determination of additional complex structures.

The structures of the N proteins from lambdoid phages  $\lambda$  and P22 revealed that these proteins adopt a bent  $\alpha$  helix within their arginine-rich motifs, while other arginine-rich proteins bind to their target RNA as an  $\alpha$  helix (HIV-1 Rev peptide [62,63]) or as an antiparallel  $\beta$ -sheet (BIV-Tat peptide [43,64]; HIV-1 Tat peptide model [65]).

Arginine-rich motifs can thus adopt quite different secondary structure motifs and recognition modes. The number of essential arginines varies in lentiviral Tat proteins [66,69], aromatic stacking is most important for phage  $\lambda$  N protein, N protein of P22 has significant hydrophobic interactions with the flipped out base (C11), whereas only few contacts between A9 and amino acids are observed in the case of the phage  $\lambda$  protein. Common to all these peptides, however, is their binding in the major groove of the RNA near special secondary structure

elements like loops or bulges [43,64] and their transition from disordered to ordered structure upon binding. Indeed, this structural transition upon RNA binding in disordered or partially disordered proteins can be speculated to be a key feature of viral RNA-recognition proteins.

## ACKNOWLEDGEMENTS

This research was supported by the DFG Ro-617/8-1 project and the Fonds der Chemischen Industrie. We thank Ulrike Herzog and Petra Deuerling for the synthesis of RNA and protein preparations.

## REFERENCES

- Das, A. (1993) Control of transcription termination by RNA-binding proteins. *Annu. Rev. Biochem.* **62**, 893–930.
- Greenblatt, J., Nodwell, J.R. & Mason, S.W. (1993) Transcriptional antitermination. *Nature* **364**, 401–406.
- Olson, E.R., Tomich, C.S. & Friedman, D.I. (1984) The *nusA* recognition site. Alteration in its sequence or position relative to upstream translation interferes with the action of the N antitermination function of phage lambda. *J. Mol. Biol.* **180**, 1053–1063.
- Salstrom, J.S. & Szybalski, W. (1978) Coliphage lambda *nutL*–: a unique class of mutants defective in the site of gene N product utilization for antitermination of leftward transcription. *J. Mol. Biol.* **124**, 195–221.
- Nodwell, J.R. & Greenblatt, J. (1993) Recognition of *boxA* anti-terminator RNA by the *E. coli* antitermination factors NusB and ribosomal protein S10. *Cell* **72**, 261–268.
- Franklin, N.C. (1984) Conservation of genome form but not sequence on the transcription antitermination determinants of bacteriophages  $\lambda$ ,  $\Phi$ 21, and P22. *J. Mol. Biol.* **181**, 75–84.
- Lazinski, D., Grzadzilska, E. & Das, A. (1989) Sequence-specific recognition of RNA hairpins by bacteriophage antiterminators requires a conserved arginine-rich motif. *Cell* **59**, 207–218.
- Chattopadhyay, S., Garcia-Mena, J., DeVito, J., Wolska, K. & Das, A. (1995) Bipartite function of a small RNA hairpin in transcription antitermination in bacteriophage  $\lambda$ . *Proc. Natl Acad. Sci. USA* **92**, 4061–4065.
- Mogridge, J., Mah, T.-F. & Greenblatt, J. (1995) A protein–RNA interaction network facilitates the template-independent cooperative assembly on RNA polymerase containing the  $\lambda$  N protein. *Genes Dev.* **9**, 2831–2844.
- Cilley, W.D. & Williamson, J.R. (1997) Analysis of bacteriophage N protein and peptide binding to *boxB* RNA using polyacrylamide gel coelectrophoresis (PACE). *RNA* **3**, 57–67.
- Tan, R. & A.D. (1995) Frankel. Structural variety of arginine-rich RNA-binding peptides. *Proc. Natl Acad. Sci. USA* **92**, 5282–5286.
- van Gilst, M.R. & von Hippel, P.H. (1997) Assembly of the N-dependent antitermination complex of phage  $\lambda$ : NusA and RNA bind independently to different unfolded domains of the N protein. *J. Mol. Biol.* **274**, 160–173.
- Mogridge, J.T.-F., Mah, T.-F. & Greenblatt, J. (1998) A protein RNA interaction network facilitates the template-independent cooperative assembly on RNA polymerase containing the  $\lambda$  N protein. *J. Biol. Chem.* **273**, 4143–4148.
- Legault, P., Li, J., Mogridge, L.E., Kay & Greenblatt, J. (1998) NMR structure of the bacteriophage  $\lambda$  N Peptide/*boxB* RNA complex: recognition of a GNRA fold by an arginine-rich motif. *Cell* **93**, 289–299.
- Heus, H.A. & Pardi, A. (1991) Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **253**, 191–194.
- Cai, Z., Gorin, A., Frederick, R., Ye, X., Hu, W., Majumdar, A., Kettani, A. & Patel, D.J. (1998) Solution structure of P22 transcriptional antitermination N peptide-*boxB* RNA complex. *Nature Struct. Biol.* **5**, 203–212.
- Milligan, J.F., Groebe, D.R., Witherell, G.W. & Uhlenbeck, O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15**, 8783–8798.
- Metzger, A.U., Schindler, T., Willbold, D., Kraft, M., Steegborn, C., Volkmann, A., Frank, R.W. & Rösch, P. (1996) Structural rearrangements on HIV-1 Tat (32–72) TAR complex formation. *FEBS Lett.* **384**, 255–259.
- Kraft, M., Westendorp, M., Krammer, P., Bayer, P., Rösch, P. & Frank, R. (1995) *Peptides in Chemistry and Biology 1994* (Maja, L.S., ed.). Escom, Leiden.
- Marion, D., Ikura, M., Tschudin, L. & Bax, A. (1989) Rapid recording of 2D NMR spectra without phase cycling. application to the study of hydrogen exchange in proteins. *J. Magn. Res.* **85**, 393–399.
- Kay, L.E., Keifer, P. & Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**, 10663–10665.
- Schleucher, J., Sattler, M. & Griesinger, C. (1993) Coherence selection by gradients without signal attenuation: application to the three-dimensional HNCO experiment. *Angew. Chem. Int.* **32**, 1489–1491.
- Sattler, M., Schleucher, J. & Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nuc. Magn. Res. Spec.* **34**, 93–153.
- Piotto, M., Saudek, V. & Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **2**, 661–665.
- van Zijl, P.C.M., O'Neill, J.M., Mori, S. & Hurd, R.E. (1995) Magic Angle-Gradient Double-Quantum-Filtered COSY. *J. Magn. Reson.* **113A**, 265–270.
- Hwang, T.-L. & Shaka, A.J. (1995) Water suppression that works: excitation sculpting using arbitrary waveforms and pulsed field gradients. *J. Magn. Res.* **112A**, 275–279.
- Cavanagh, J. & Rance, M. (1992) Suppression of cross relaxation effects in TOCSY spectra via a modified DIPSI-2 mixing sequence. *J. Magn. Res.* **96**, 670–678.
- Shaka, A., Lee, C.J. & Pines, A. (1988) Iterative schemes for bilinear operators – application to spin decoupling. *J. Magn. Res.* **77**, 274–293.
- Wijmenga, S.S., Steensma, E. & van Mierlo, C.P.M. (1997) Doubly sensitivity-enhanced 3D HCCH-TOCSY of C-13-labeled proteins in H<sub>2</sub>O using heteronuclear cross polarization and pulsed field gradients. *J. Magn. Res.* **124**, 459–467.
- Shaka, A., Barker, P.B., Freeman, R. (1985) Computer-optimized decoupling scheme for wideband applications and low-level operation. *Magn. Res. J.* **64**, 547–552.
- Barkhuijsen, H., de Beer, R., Bovee, W.M.M.J. & van Ormondt, D. (1985) Retrieval of frequencies, amplitudes, damping factors, and phases from time-domain signals using a linear least-squares procedure. *J. Magn. Res.* **61**, 465–481.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. (1992) *Numerical Recipes in C*, 2nd edn. Cambridge University Press.
- Zhu, G. & Bax, A. (1990) Improved linear prediction for truncated signals of known phase. *J. Magn. Res.* **90**, 405–410.
- Friedrich, M.S. (1995) A model-free algorithm for the removal of baseline artifacts. *J. Biomol. NMR* **5**, 147–153.
- Live, D.H., Davis, D.G., Agosta, W.C. & Cowburn, D. (1984) Long-range hydrogen bond mediated effects in peptides – N-15 NMR-study of gramicidin-S in water and organic solvents. *J. Am. Chem. Soc.* **106**, 1939–1941.
- Clare, G.M., Gronenborn, A.M., Nilges, M. & Ryan, C.A. (1987) Three-dimensional structure of potato carboxypeptidase inhibitor in solution. A study using nuclear magnetic resonance, distance geometry, and restrained molecular dynamics. *Biochemistry* **26**, 8012–8023.
- Cai, Z., Gorin, A., Frederick, R., Ye, X., Hu, W., Majumdar, A., Kettani, A. & Patel, D.J. (1998) Solution structure of P22 transcriptional antitermination N peptide-*boxB* RNA complex. *Nat. Struct. Biol.* **5**, 203–212.
- Varani, G., Aboul-ela, F. & Allain, F.H.-T. (1996) NMR investigation of RNA structure. *Prog. Nuc. Magn. Res. Spec.* **29**, 51–127.

39. Gubser, C.C. & Varani, G. (1996) Structure of the polyadenylation regulatory element of the human U1A pre-mRNA 3'-untranslated region and interaction with the U1A protein. *Biochemistry* **35**, 2253–2267.
40. Saenger, W. (1984) *Principles of Nucleic Acid Structure*. Springer Verlag, New York.
41. Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, E., Cech, T.R. & Doudna, J.A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA-packing. *Science* **273**, 1678–1685.
42. Wijmenga, S.S., Mooren, M.M. & Hilbers, C.W. (1993) NMR of nucleic acids; from spectra to structure. In *NMR of Macromolecules. A Practical Approach* (Roberts, G.C.K., ed.), pp. 217–288. IRL Press, New York.
43. Ye, X., Kumar, R.A. & Patel, D.J. (1995) Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex. *Chem. Biol.* **2**, 827–840.
44. Vuister, G.W. & Bax, A. (1993) Quantitative J correlation: a new approach for measuring homonuclear three-bond J ( $H^N-H^\alpha$ ) coupling constants in  $^{15}N$ -enriched proteins. *J. Am. Chem. Soc.* **115**, 7772–7777.
45. Brünger, A.T. (1996) *X-PLOR, Version 3.851 online*. Howard Medical School & Yale University, New Haven, USA.
46. Nilges, M., Gronenborn, A.M., Brünger, A.T. & Clore, G.M. (1988) Determination of three-dimensional structures of proteins from interproton distance data by simulated annealing from a random array of atoms. *FEBS Lett.* **239**, 129–136.
47. Nilges, M., Clore, G.M. & Gronenborn, A.M. (1988) Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine protease inhibitor 2. *Protein Eng.* **2**, 27–38.
48. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Wright, P.E. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
49. Luscombe, N.M., Laskowski, R.A. & Thornton, J.M. (1997) NUCPLOT: a program to generate schematic diagrams of protein–nucleic acid interactions. *Nucl Acids Res.* **25**, 4940–4945.
50. van Gilst, M.R., Rees, W.A., Das, A. & von Hippel, P.H. (1997) Complexes of N antitermination protein of phage  $\lambda$  with specific and nonspecific RNA target sites on the nascent transcript. *Biochemistry* **36**, 1514–1524.
51. Holzworth, G. & Doty, P. (1965) The ultraviolet circular dichroism of polypeptides. *J. Am. Chem. Soc.* **87**, 218–228.
52. Wijmenga, S.S. & van Buren, N.M. (1998) The use of NMR methods for conformational studies of nucleic acids. *Proc. Nuc. Magn. Res. Spec.* **32**, 287–387.
53. Duarte, C.M. & Pyle, A.M. (1998) Stepping through an RNA structure: a novel approach to conformational analysis. *J. Mol. Biol.* **284**, 1465–1478.
54. Jucker, F.M., Heus, H.A., Yip, P.F., Moors, E.H.M. & Pardi, A.A. (1996) Network of Heterogenous Hydrogen Bonds in GNRA Tetraloops. *J. Mol. Biol.* **264**, 968–980.
55. Pley, H.W., Flaherty, K.M. & McKay, D.B. (1994) Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* **372**, 111–113.
56. Howe, P.W.A., Allain, F.H.-T., Varani, G. & Neuhaus, D. (1998) Determination of the NMR structure of the complex between U1A protein and its RNA polyadenylation inhibition element. *J. Biomol. NMR* **11**, 59–84.
57. Franklin, N.C. (1993) Clustered arginine residues of bacteriophage  $\lambda$  N protein are essential to antitermination of transcription, but their locale cannot compensate for *boxB* loop defects. *J. Mol. Biol.* **231**, 343–360.
58. Su, L., Radek, J.T., Hallenga, K., Hermanto, P., Chan, G., Labeets, L.A. & Weiss, M.A. (1997) RNA recognition by a bent  $\alpha$ -helix regulates transcriptional antitermination in phage  $\lambda$ . *Biochemistry* **36**, 12722–12732.
59. Oubridge, C., Ito, N., Evans, P.R., Teo, C.-H. & Nagai, K. (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A splicosomal protein complexed with an RNA hairpin. *Nature* **372**, 432–438.
60. Hung, S.C. & Gottesman, M.E. (1995) Phage HK022 N protein arrest transcription on phage lambda DNA *in vitro* and competes with the phage lambda N antitermination protein. *J. Mol. Biol.* **247**, 428–442.
61. Gutell, R.R., Larson, N. & Woese, C.R. (1994) Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol. Rev.* **58**, 10–26.
62. Battiste, J.L., Mao, H., Rao, N.S., Tan, R., Muhandiram, D.R., Kay, L.E., Frankel, A.D. & Williamson, J.R. (1996) Alpha helix RNA major groove recognition in an HIV 1 rev peptide-RRE RNA complex. *Science* **273**, 1547–1551.
63. Ye, X., Gorin, A., Ellington, A.D. & Patel, D.J. (1996) Deep penetration of an alpha helix into a widened RNA major groove in the HIV-rev peptide-RNA aptamer complex. *Nature Struct. Biol.* **3**, 1026–1033.
64. Puglisi, J.D., Chen, L., Blanchard, S. & Frankel, A.D. (1995) Solution Structure of a bovine immunodeficiency virus Tat peptide-TAR RNA complex. *Science* **270**, 1200–1203.
65. Seewald, M.J., Metzger, A.U., Willbold, D. & Rösch, P. & Sticht, H. (1998) Structural model of the HIV-1 Tat (46–58)-TAR complex. *J. Biomol. Str. Dyn.* **16**, 683–692.
66. Farrow, M.A., Aboul-Ela, F., Owen, D., Karpeisky, A., Beigelman, L. & Gait, M.J. (1998) Site specific cross-linking of amino acids in the basic region of human immunodeficiency virus type 1 Tat peptide to chemically modified TAR RNA duplexes. *Biochemistry* **37**, 3096–3108.
67. Jeener, J., Meier, B.H., Bachmann, P. & Ernst, R.R. (1979) Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **71**, 4546–4553.
68. Zhang, W., Kay, L.E., Olivier, J.P. & Forman-Kay, J.D. (1994) Backbone  $^1H$  and  $^{15}N$  resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J. Biomol. NMR* **4**, 845–858.
69. Talluri, S. & Wagner, G. (1996) An Optimized 3D NOESY-HSQC. *J. Magn. Res.* **112B**, 200–205.
70. Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S. & Sykes, B.D. (1995)  $^1H$ ,  $^{13}C$  and  $^{15}N$  random coil NMR chemical shifts of the common amino acids. I. Investigation of the nearest-neighbor effects. *J. Biomol. NMR* **5**, 67–8178.