The EMBO Journal Vol.17 No.14 pp.4092–4100, 1998

Solution structure of the antitermination protein NusB of *Escherichia coli*: a novel all-helical fold for an RNA-binding protein

Martin Huenges, Christian Rölz, Ruth Gschwind, Ralph Peteranderl, Fabian Berglechner[1], Gerald Richter[1], Adelbert Bacher[1], Horst Kessler and Gerd Gemmecker[2]

Lehrstuhl für Organische Chemie II and [1]Lehrstuhl für Organische Chemie und Biochemie, Technische Universität München, Lichtenbergstr. 4, D-85747 Garching, Germany

[2]Corresponding author
e-mail: gg@artus.org.chemie.tu-muenchen.de

The NusB protein of *Escherichia coli* is involved in the regulation of rRNA biosynthesis by transcriptional antitermination. In cooperation with several other proteins, it binds to a dodecamer motif designated *rrn boxA* on the nascent rRNA. The antitermination proteins of *E.coli* are recruited in the replication cycle of bacteriophage λ, where they play an important role in switching from the lysogenic to the lytic cycle. Multidimensional heteronuclear NMR experiments were performed with recombinant NusB protein labelled with ^{13}C, ^{15}N and ^{2}H. The three-dimensional structure of the protein was solved from 1926 NMR-derived distances and 80 torsion angle restraints. The protein folds into an α/α-helical topology consisting of six helices; the arginine-rich N-terminus appears to be disordered. Complexation of the protein with an RNA dodecamer equivalent to the *rrn boxA* site results in chemical shift changes of numerous amide signals. The overall packing of the protein appears to be conserved, but the flexible N-terminus adopts a more rigid structure upon RNA binding, indicating that the N-terminus functions as an arginine-rich RNA-binding motif (ARM).

Keywords: antitermination/transcription/NusB protein/NMR spectroscopy/RNA-binding protein

Introduction

Modification of the efficiency of transcription termination at specific sites is used as a regulatory mechanism in prokaryotic and eukaryotic organisms. It has been studied in considerable detail in the *Escherichia coli* phage λ, where it plays a crucial role in the switch between the lysogenic and the lytic cycle (for review, see Richardson and Greenblatt, 1996).

The genes of the *E.coli* phage λ are organized in several operons. Only promoter-proximal genes are transcribed during the lysogenic state. The transition to the lytic cycle is characterized by extension of the transcripts into the more distal parts of the operons, which becomes possible by overreading specific terminator sequences. This phenomenon is mediated by the phage protein N in conjunction with several host proteins (DeVito and Das, 1994; Mogridge *et al.*, 1995).

The host genes encoding the proteins involved in this process are collectively designated *nus* for 'N utilization substance'. More specifically, the products of the genes *nusA*, *nusB*, *nusE* and *nusG* and at least one additional, so far unidentified, gene product are required. The NusE protein has been shown to be identical to the S10 protein of the small ribosomal subunit (Friedman *et al.*, 1981; Richardson and Greenblatt, 1996).

The various components of the antitermination system exert their effect on phage λ transcription by interacting with phage mRNA sites designated *nut* (for N utilization). The two sequence motifs that form the λ phage *nut* sites are designated *boxA* and *boxB* (Figure 1) (Nodwell and Greenblatt, 1991). A large body of genetic and biochemical evidence indicates that λ phage N protein interacts with *boxB*, whereas Nus proteins of the bacterial host interact with *boxA* of the bacteriophage mRNA. A complex involving the various protein components of the antitermination systems forms via weak protein–protein interactions. NusB protein and NusE/S10 protein have also been shown to form a heterodimer *in vitro* (Mason *et al.*, 1992; Nodwell and Greenblatt, 1993).

Whereas the recruitment of the *E.coli* Nus proteins for the regulatory purposes of lambdoid phages has been analysed in considerable detail, the physiological role of the Nus proteins in the bacterial cell is less well understood. The available evidence indicates that antitermination mediated by these proteins is one of numerous mechanisms involved in the control of rRNA biosynthesis (for review, see Keener and Nomura, 1996). The *E.coli* genome includes seven *rrn* operons coding for rRNA. Each of these operons contains the sequence motif UGCUCUUUAACA (designated *rrn boxA*) at a location slightly downstream from the transcription start site. Binding of NusB protein in cooperation with other Nus proteins at *rrn boxA* is supposed to modulate the efficiency of ribosomal biosynthesis via control of rRNA synthesis, thus influencing the growth rate of the microorganism (Gaal *et al.*, 1997). Nonsense mutations of the *nusB* gene therefore lead to a reduced growth rate (Taura *et al.*, 1992).

An antitermination mechanism has been found to play an important role in the replication cycle of human immunodeficiency virus (HIV) (Cheng *et al.*, 1991; Karn and Graeble, 1992; Harada *et al.*, 1996). The viral Tat protein interacts with the *tar* sequence in the long terminal repeat of the nascent viral RNA. The λ phage/*E.coli* antitermination system may be able to serve as a model for the antitermination mechanism of HIV.

Within the last few years, the three-dimensional structures of several RNA-binding proteins have been reported. The three most common folds of such proteins, ribonucleo-

		boxA		*boxB*	
A	A A C C C	C G C U C U U A C A C A	U U C C A	G C C C U G A A A A A G G G C	A U C A A
B	G G U G A	C G C U C U U A A A A A	U U A A -	G C C C U G A A G A A G G G C	A G C A U
C	G S M A C	U G C U C U U U A A C A	A U U U A U C A G A C A A U C U G U G U G G G A C		

Fig. 1. RNA sequence motifs binding to proteins involved in transcription antitermination. (**A**) *nutR* and (**B**) *nutL* from bacteriophage λ; (**C**) consensus sequence from the *rrn* operons of *E.coli*. Nucleotides are indicated in IUPAC nomenclature; S denotes G or C; M means A or C.

protein (RNP), K-homology (KH) and double-stranded RNA-binding domains (dsRBD), all show a mixed αβ topology. It has therefore been suggested that the αβ topology might represent a particularly favourable framework for RNA recognition (for review, see Varani, 1997). However, more recently, a number of all-helical RNA-binding proteins have also been described (Berglund *et al.*, 1997; Markus *et al.*, 1997; Xing *et al.*, 1997).

The secondary structure of NusB has been characterized recently as completely α-helical (cf. Altieri *et al.*, 1997; Berglechner *et al.*, 1997). Here we report the solution structure of the NusB protein of *E.coli*, based on multidimensional nuclear magnetic resonance (NMR) data. It contains a putative arginine-rich RNA-binding motif (ARM), but otherwise does not resemble any known RNA-binding fold. Our studies of the interaction of NusB protein with its specific binding site, *rrn boxA* RNA, further suggest that the binding of RNA induces a stable conformation of the flexible N-terminal segment.

Results

Structure determination

Most of the ^{1}H, ^{13}C and ^{15}N resonances of the NusB protein have been assigned by multidimensional heteronuclear NMR experiments, as reported earlier (Altieri *et al.*, 1997; Berglechner *et al.*, 1997). A total of 89% of the backbone resonances of NusB were assigned from NMR spectra of U-^{13}C,^{15}N-labelled NusB samples that were also 75% randomly deuterated for improved spectral resolution (Berglechner *et al.*, 1997). The remainder of the backbone resonances could not be observed either in the triple resonance or in ^{15}N-resolved nuclear Overhauser enhancement (NOE) spectra. Specifically, residues located in the hypothetical RNA-binding region (see below), including the putative ARM region (residues 1–10) and the central part of the loop linking helices 2 and 3 (residues 40–42), failed to yield any detectable signals. This is most likely due to the absence of a stable conformation of these regions when not bound to RNA (for details, see Discussion).

The secondary structure of NusB has been published previously, based on an analysis of the NOE data involving the backbone protons, $^{3}J_{HNH\alpha}$ coupling constants, amide exchange rates and the chemical shift information (Berglechner *et al.*, 1997). For the determination of the tertiary structure, overlapping resonances were resolved in a three-dimensional ^{15}N,^{15}N-HMQC-NOESY-HSQC spectrum (Kay *et al.*, 1990) at 600 MHz and a three-dimensional ^{13}C,^{13}C-HSQC-NOESY-HSQC spectrum (Clore *et al.*, 1991; Bax and Grezesiek, 1993) at 750 MHz. All NH–NH NOEs were measured on the 75% fractionally deuterated sample to achieve increased amide relaxation times and reduced spin diffusion. This results in a significant increase in the number of observable NOEs compared with NMR studies on the fully protonated protein (Torchia *et al.*, 1988; Grzesiek *et al.*, 1995; Venters *et al.*, 1995). Representative strip plots of the NH–NH NOEs for part of helix 6 are shown in Figure 3 to document the quality of the acquired spectra.

Table I. Structural statistics and r.m.s. deviation for 18 NMR-derived structures of NusB

NMR constraints		
Distance constraints	1923	
Intraresidue	735	
Sequential ($\|i-j\| = 1$)	384	
Medium range ($1 < \|i-j\| \leq 4$)	261	
Long range ($\|i-j\| > 4$)	331	
Ambiguous	212	
Hydrogen bond constraints	51	
Dihedral angle constraints		
Φ	67	
$^{3}J(H_{N}H_{\alpha})$	13	
Constraints/122 ordered residues	16.8	
Stereospecific assignments		
α-Methylene	1	
β-Methylene	55	
γ-Methylene	16	
δ-Methylene, asparagine δ-NH2	7	
ε-Methylene, glutamine ε-NH2	5	
Valine γ-methyls	9 out of 13	
Leucine δ-methyls	14 out of 17	
Structural statistics		
R.m.s. deviation from NOE restraints (Å)	0.054 ± 0.003	
R.m.s. deviations from ideal covalent geometry		
Bonds (Å)	0.0196 ± 0.0002	
Angles (°)	1.54 ± 0.03	
Improper torsions (°)	1.10 ± 0.04	
Pairwise r.m.s. deviations from mean coordinates (Å)		
	Backbone atoms	Heavy atoms
Helical regions (12–22, 27–35, 46–65, 79–93, 100–113, 120–131)	0.25 ± 0.08	0.65 ± 0.19
All structured residues (11–38, 46–139)	0.39 ± 0.10	0.81 ± 0.16

For the determination of the three-dimensional structure of the NusB protein, 1926 NMR-derived distances and 80 torsion angle restraints were used (Table I, Figure 4). Due to missing assignments and lack of NOEs, the N-terminal 10 residues as well as residues 39–45 are largely unstructured in the uncomplexed protein (Figure 4). For the well-defined regions (residues 11–38 and 46–139), the atomic r.m.s. distribution of the best 18 structures about the mean coordinate position is 0.39 ± 0.1 Å for the backbone atoms and 0.81 ± 0.16 Å for all heavy atoms.

Structure of NusB

NusB is composed of six helices, formed by residues Cys12–Leu22 (helix 1), Ile27–Leu35 (helix 2), Leu46–Leu65 (helix 3), Gln79–Ser93 (helix 4), Tyr100–Ser113

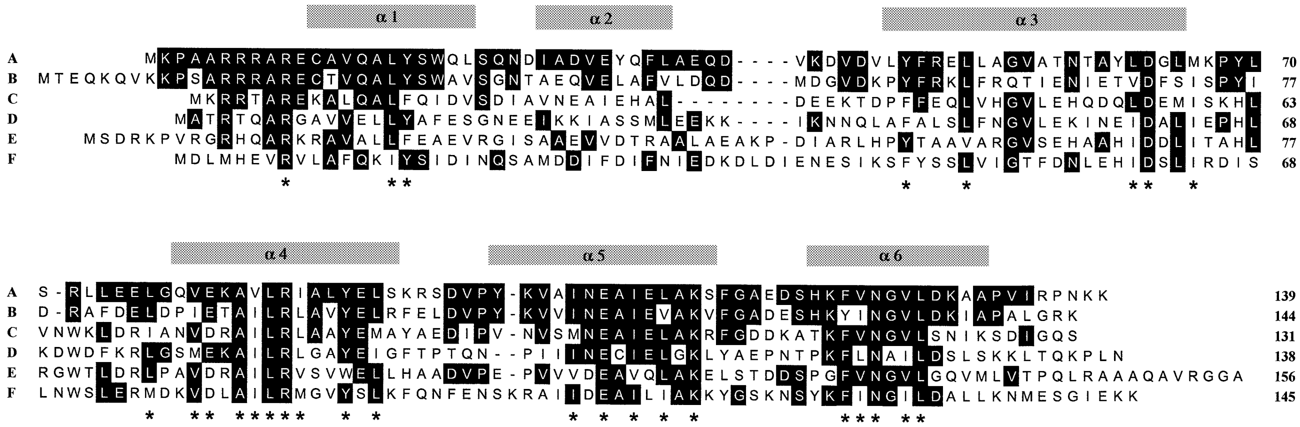

Fig. 2. Amino acid sequences predicted from putative *nusB* genes. Helix segments obtained for the *E.coli* protein by NMR analysis are shown as shaded boxes. Conserved amino acid residues (with reference to *E.coli* NusB protein as the prototype sequence) are shown in inverted contrast. Asterisks indicate amino acid residues which are similar or identical in all sequences shown. (**A**) *E.coli*; (**B**) *Haemophilus influenzae*; (**C**) *Bacillus subtilis*; (**D**) *Helicobacter pylori*; (**E**) *Mycobacterium tuberculosis*; (**F**) *Borrelia burgdorferi*.

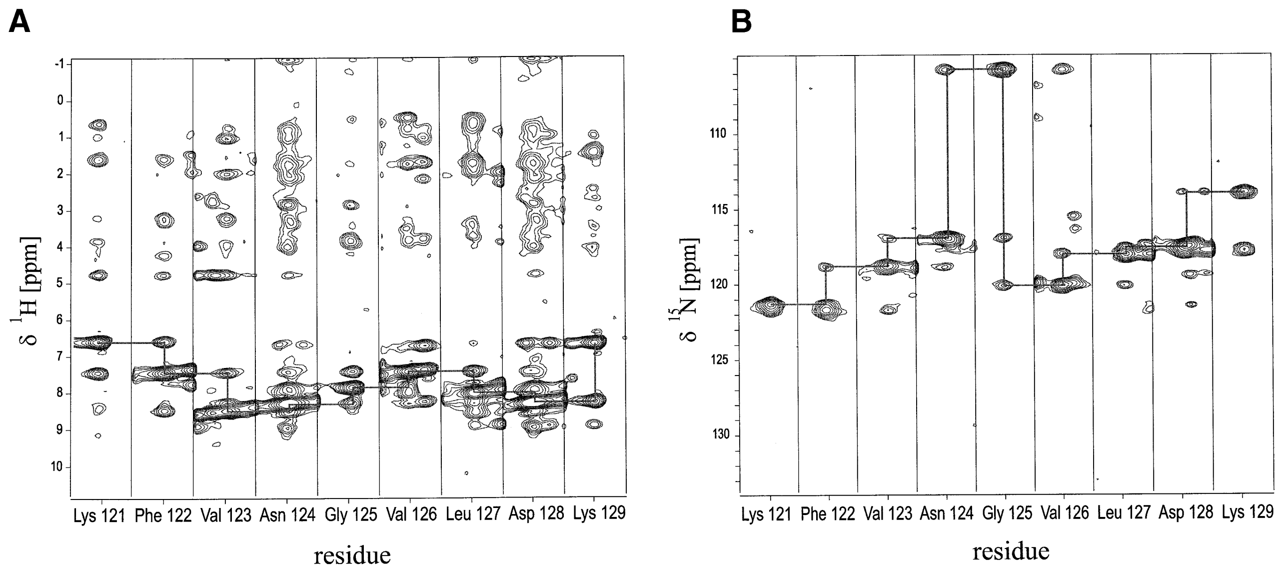

Fig. 3. Representative strip plots from (**A**) the three-dimensional ^{15}N-HSQC-NOESY and (**B**) the three-dimensional ^{15}N,^{15}N-HMQC-NOESY-HSQC spectrum of the uniformly ^{15}N,^{13}C-labelled and fractionally deuterated NusB protein for part of helix 6. Sequential NH–NH NOEs are indicated by solid lines.

(helix 5) and His120–Ala131 (helix 6). As indicated by the chemical shift deviations and the large $^3J_{HNH\alpha}$ coupling constants of Thr57 and Asn58, helix 3 displays a distinctive kink at Thr57–Thr59.

In the calculated ensemble of NusB structures, all individual secondary structure elements are very well defined. This is also true for the relative orientation of the helices. While the N-terminus (residues 1–10) is largely disordered in the structural ensemble (due to missing NMR data in this region), the C-terminus shows only little divergence between the structures. Of the five loops connecting the helices, four are reasonably well ordered. However, the loop linking helices 2 and 3 (residues 36–45) seems to be largely unstructured, as indicated by NMR data (i.e. very weak or completely missing signals). Putative NusB homologues from different bacteria show considerable variability in the length of this loop (Figure 2). The hypothetical NusB protein of the cyanobacterium *Synechocystis* sp. (not shown in Figure 2) shows an insertion of 117 amino acid residues in comparison with the *E.coli* sequence. The absence of sequence and length conservation in this loop region agrees well with the high degree of flexibility observed here by NMR.

The overall fold of NusB can be characterized as an α/α sandwich, with helices 1 and 3 forming one layer, and helices 4–6 forming a second layer co-planar to the first. Helix 2 is packed onto the interface between helices 1 and 3 on the outside of the first layer (Figure 5).

The fold of the subdomain formed by helices 1, 2 and 3 is stabilized by interactions between hydrophobic side chains of helix 1 (Ala16, Leu17, Trp20, Leu22), helix 2 (Ile27, Ala28, Val30, Tyr32, Leu35, Ala36) and helix 3 (Leu46, Phe48, Leu51, Leu52, Val55, Ala56). Helix 1 and the N-terminal part of helix 3 (residues 46–56) are oriented nearly parallel.

In the second α-helical layer (helices 4, 5 and 6), the helices are also positioned in an almost planar arrangement. Interactions between hydrophobic side chains of helix 4 (Val80, Ala83, Val84, Ile87, Ala88, Leu89, Tyr90, Leu92) and helix 5 (Val102, Ala103, Ile108, Leu110, Ala111), as

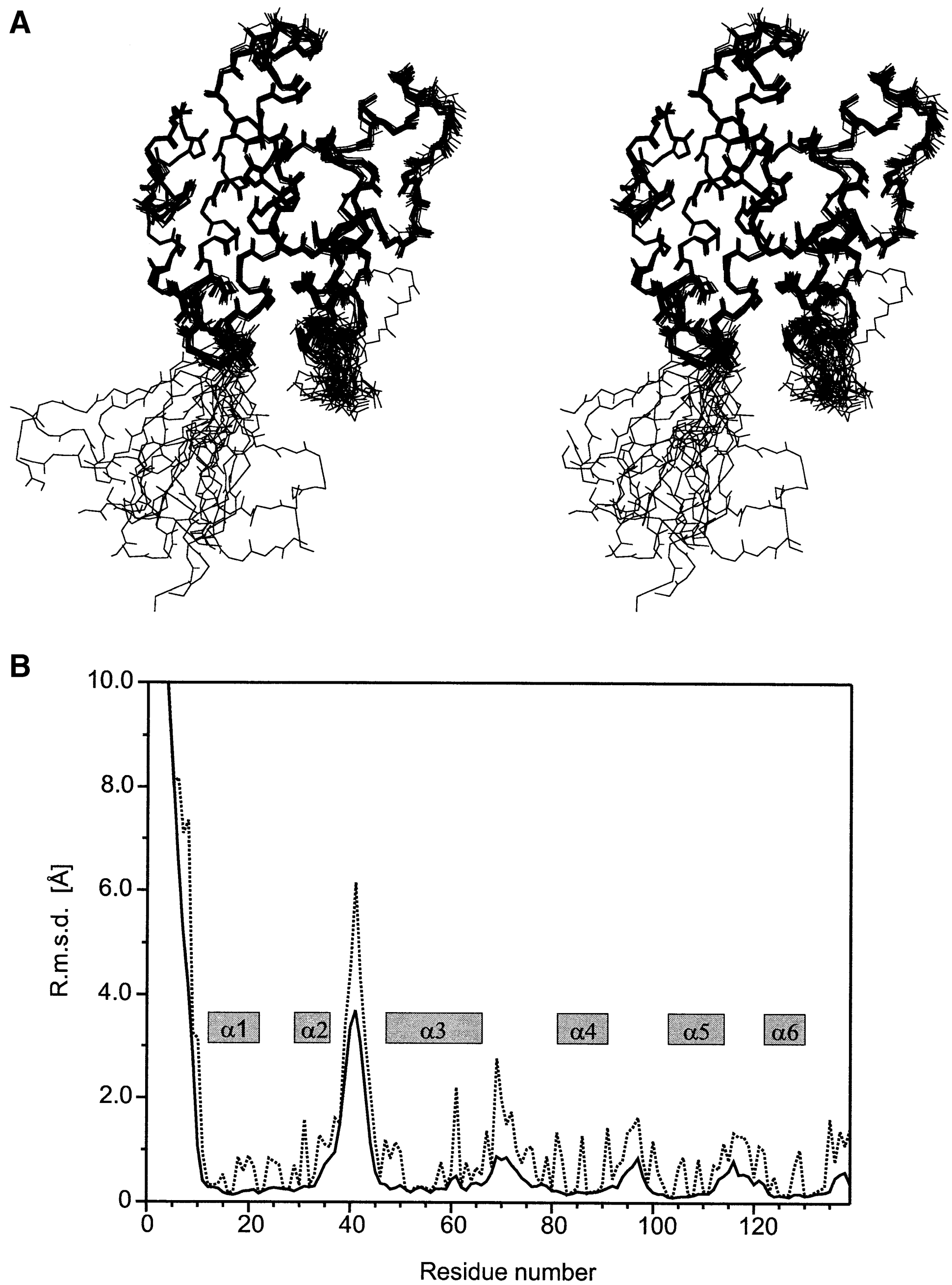

Fig. 4. (**A**) Stereo view of the backbones (N, Cα, C') of 18 superimposed NMR-derived structures of NusB protein. (**B**) R.m.s. deviations of the structure ensemble as a function of residue, shown for the backbone atoms (solid line) and for the side chain atoms (dotted line). Helical secondary structure is indicated by solid rectangles.

well as between helix 5 (Ala103, Ile104, Ala107, Ile108, Leu110, Ala111) and helix 6 (His120, Val123, Val126, Leu127, Ala130, Ala131), stabilize the fold of this subdomain.

The only covalent connection between both layers is the long loop connecting helices 3 and 4 (residues 66–78). In addition, there is a wealth of hydrophobic interactions between the two subdomains stabilizing the tertiary structure, as indicated by numerous NOE contacts. Residues from helix 1 (Ala13, Val14, Ala16, Tyr18, Leu22) interact with residues of helices 5 (Val102, Ile104, Ile108) and 6 (His120, Leu127, Ala131), while residues from helix 3 (Leu46, Leu52, Ala53, Ala60, Tyr61, Leu62) interact with residues of helices 4 (Ala83, Leu89, Tyr90), 5 (Ile108) and 6 (Glu117, His120).

Binding of RNA to NusB

To investigate the interactions between NusB protein and an RNA dodecamer equivalent to the *rrn boxA* motif (Figure 1), a series of ^{1}H,^{15}N-HSQC spectra were recorded. The chemical shifts of both nuclei are sensitive to their local electronic environment and therefore can be used as a probe for interactions between the labelled protein and unlabelled RNA. Presumably, the strongest perturbation of the electronic environment will be observed for the residues that either come into direct contact with RNA or

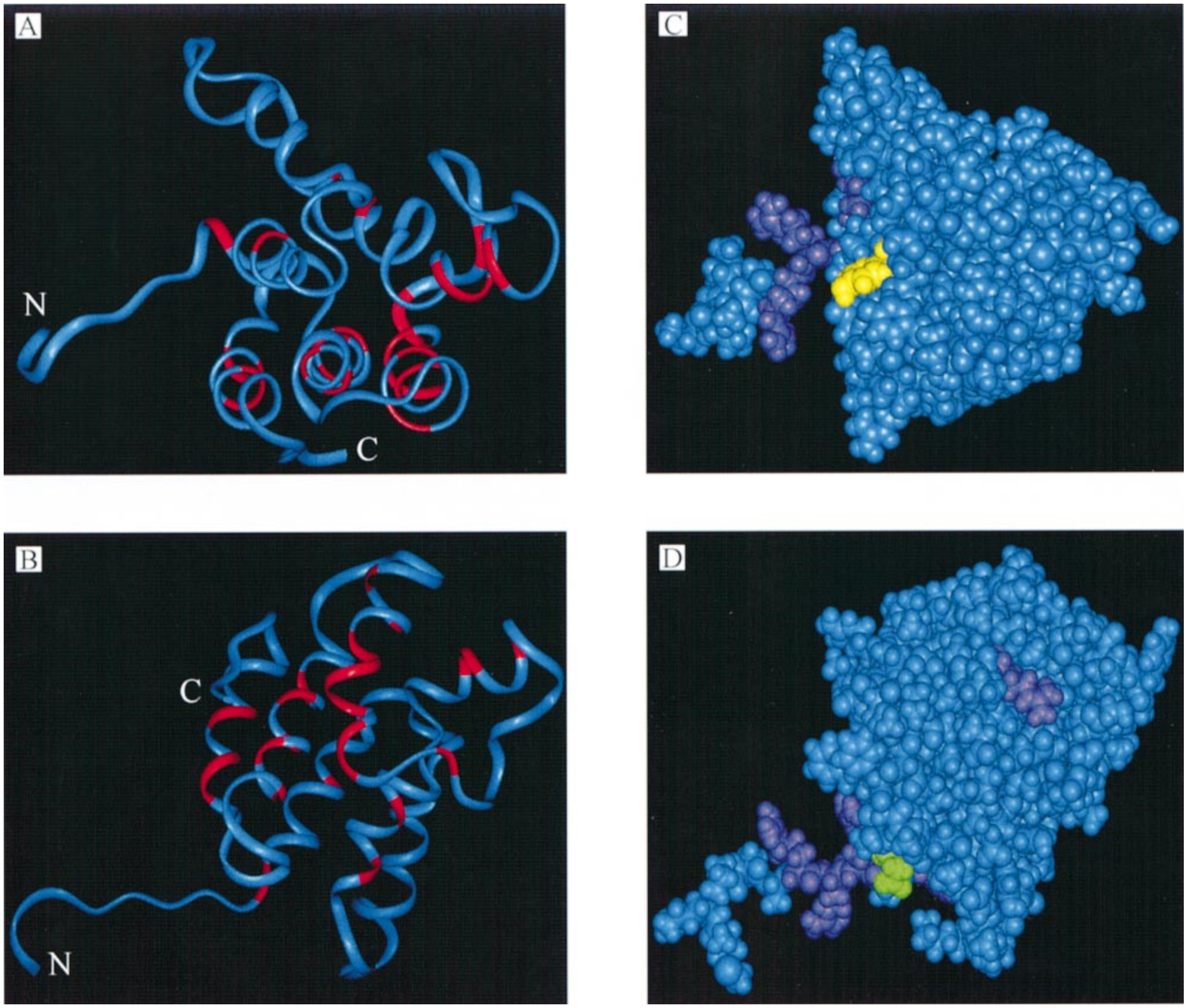

Fig. 5. Ribbon plot and CPK model of the structure of NusB protein. (**A**) View parallel to the axis of helices 4, 5 and 6; (**B**) view perpendicular to helices 4, 5 and 6. Residues conserved in NusB proteins from different bacteria (marked by asterisks in Figure 2) are shown in red.
(**C**) and (**D**) Space-filling model of NusB. The four N-terminal arginine residues, Arg6, Arg7, Arg8 and Arg10, which become detectable upon binding to RNA, and Arg86, exhibiting a relatively large change in proton chemical shift upon RNA binding, are shown in violet. Tyr18 and Asp118 are depicted in yellow and green, respectively. The protein is shown in the same orientations as in (A) and (B), respectively.

that are involved in major conformational changes upon binding to RNA.

NusB protein (0.7 mM) complexed to an excess of the RNA dodecamer exhibited only one set of peaks in the $^{1}H,^{15}N$-HSQC spectra (Figure 6A). This indicates the presence of only a single conformation of the protein–RNA complex on the NMR time scale. Titration of NusB with substoichiometric amounts of the RNA dodecamer yielded two sets of peaks, one identical to the set observed for the free protein, the other identical to the peaks of the fully complexed protein.

Due to the large number of signals and the absence of any intermediate signals, the complete assignment of the spectra of NusB bound to RNA has not been accomplished yet. However, there are a number of interesting insights which can be gained even from a preliminary analysis of the spectra of the complex. While a large number of signals migrate upon binding, some are clearly not affected, as for example the NH signals of Trp20, Val45 and Asp97 (Figure 6A), suggesting that the local environments of these amino acid residues involved in interhelical contacts are not affected by RNA binding.

Furthermore, a set of four additional signals appears in the Arg-NHε region (Figure 6B), which were not visible in the spectra of the uncomplexed protein and therefore belong to the four previously unassigned arginines 6, 7, 8 and 10. The Arg side chain NHs of arginines 49, 72, 86, 95 and 135 (assigned from NOE data in the free NusB spectra) are not affected by RNA addition, with the exception of a small shift for Arg86.

In contrast, complexation of the NusB protein with the RNA mutant C3G, reported as non-binding under the low RNA concentration conditions of band shift assays (Nodwell and Greenblatt, 1993), showed that some of the N-terminal arginine residues of the proposed ARM motif are not visible and others are shifted differently (data not shown). This suggests that the mutant RNA does actually bind under the high RNA concentration conditions of the

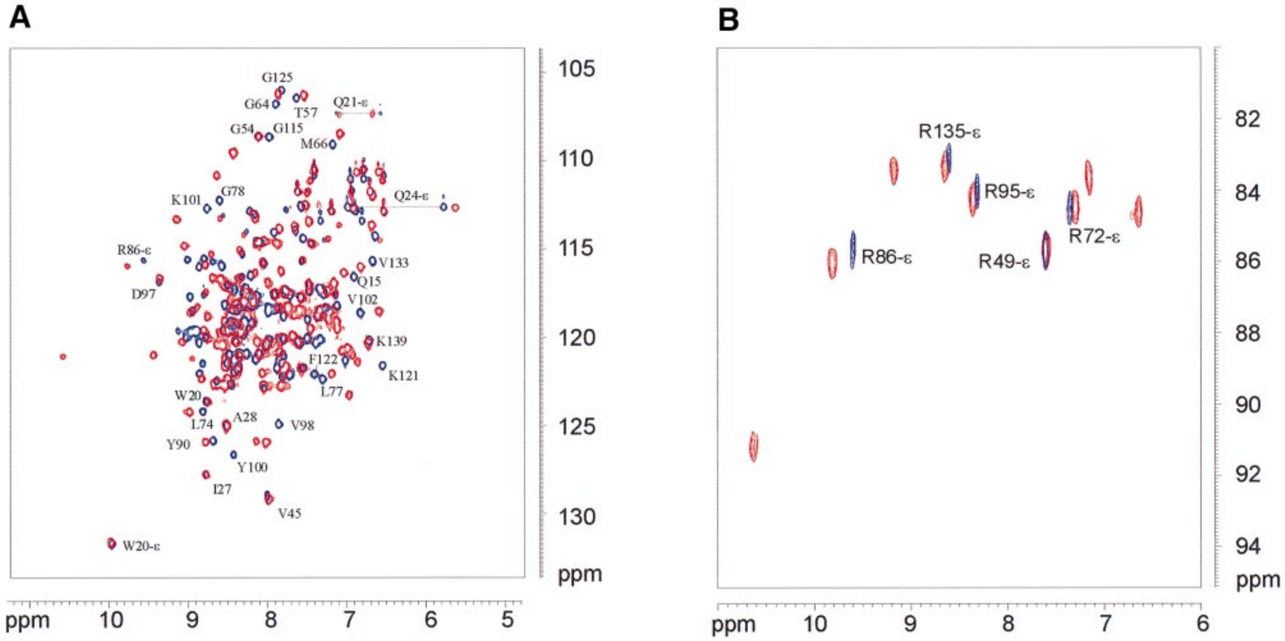

Fig. 6. ^{1}H,^{15}N-HSQC spectrum at 600 MHz of ^{15}N-labelled NusB protein uncomplexed (blue) and complexed (red) to the RNA dodecamer of the *boxA* consensus sequence (5′-UGCUCUUUAACA-3′). Labels correspond to amide backbone assignments of well-resolved signals of the free protein (for complete assignments see Berglechner *et al.*, 1997). (**A**) Amide region of the ^{1}H,^{15}N-HSQC spectrum; arginine NHε signals are folded. (**B**) Expanded region of the arginine NHε signals. In the HSQC spectrum of the uncomplexed protein (blue), only five of the nine arginine side chains can be detected. Upon binding of RNA, four additional signals appear, and the signal of Arg86-NHε is slightly shifted.

NMR experiment. However, the differences in the protein HSQC patterns of the complexes with wild-type and mutant *boxA* RNA indicate that at least some of the observed NusB–RNA interactions are indeed sequence specific.

Discussion

Structure of NusB

Most sequence-specific RNA-binding proteins contain domains of 60–90 amino acids responsible for RNA recognition and additional domains involved in regulation and other activities (Biamonti and Riva, 1994; Burd and Dreyfuss, 1994; Nagai, 1996; for a review, see Varani, 1997). The three most common folds of RNA-binding proteins—RNP, KH and dsRBD—form compact globular structures (Nagai, 1996). They all are αβ proteins with an antiparallel β-sheet on one face of the protein packed via a hydrophobic core against a layer of α-helices. This αβ structural theme is conserved in many RNA-binding proteins, even if they do not share sequence homology with other members of these three structure families. Examples include ribosomal proteins (Liljas and Garber, 1995) and other protein factors involved in translation (Biou *et al.*, 1995; Garcia *et al.*, 1995; Kang *et al.*, 1995; Liljas and Garber, 1995). The three families of RNA-binding αβ proteins have distinct topologies: the RNP domain contains tandem repeats of a βαβ motif (Nagai *et al.*, 1990; Hoffmann *et al.*, 1991), dsRBDs have an αβββα topology (Bycroft *et al.*, 1995; Kharrat *et al.*, 1995) and KH proteins display a βααββα fold (Musco *et al.*, 1996).

The best characterized example of an all-helical RNA-binding protein, the Rop protein, displays a four-helix bundle formed by the association of two identical helix–turn–helix monomers (Banner *et al.*, 1987; Eberle *et al.*, 1991). Recently, the structures of two entirely α-helical rRNA-binding proteins (S15 and L11) have been determined by NMR (Berglund *et al.*, 1997; Markus *et al.*, 1997; Xing *et al.*, 1997). However, the three-dimensional structure of NusB is different from that of either of these proteins, as well as from that of any other protein or domain structure published so far, according to a search of the Brookhaven protein data bank (Bernstein *et al.*, 1977) with the program DALI (Holm and Sander, 1993).

Complex formation between NusB protein and *rrn boxA* RNA leads to a shift of many of the amide backbone signals of the protein. For a reliable determination of the interface region, it will therefore be necessary to repeat the assignment process for most of the backbone signals. However, the HSQC data already indicate that a major change occurs in the N-terminal region representing the putative ARM motif. It has been shown earlier that the *nusB5* mutation (replacement of Tyr18 by Asp in NusB; shown in yellow in Figure 5C) prevents switching to the lytic cycle in phage λ infected *E.coli* cells but does not confer a cold-sensitive phenotype which has been described for an insertion mutation (*ssyB63*) and an early amber mutation (*ssaD5*) (for summary, see Court *et al.*, 1995). This suggests tentatively that Tyr18 may be essential for interaction with phage λ *boxA* and/or N protein, but not for recognition of bacterial *boxA*.

Genetic studies indicate that a D118N mutant (*nusB101*) (shown in green in Figure 5D) of NusB protein fails to interact with NusE/S10 (see Court *et al.*, 1995). The *nusB101* mutation can be suppressed by the *nusE71* mutation (A86D). Asp118 is located in the loop between helices 5 and 6 of the NusB protein, in close proximity to the N-terminus containing the proposed ARM region. This suggests that this surface region is involved in the interaction with the NusE/S10 protein as well as with the RNA site. To provide a structural scaffold for the ARM region to fold against when binding to the presumably also unstructured *boxA* RNA, it is to be expected that the structured part of NusB will also be involved in RNA binding as well as possibly other Nus factors.

Comparison of free and RNA-bound protein

In the so-called 'basic-domain class' of RNA-binding proteins, RNA recognition is mediated by a sequence of 10–15 amino acids rich in arginine and lysine residues. The sequences and conformations of the ARMs in different proteins vary widely (Lazinski *et al.*, 1989). For example, the ARM region of bovine immunodeficiency virus (BIV) Tat binds as a β hairpin to RNA (Puglisi *et al.*, 1995; Ye *et al.*, 1995) while the arginine-rich region adopts an α-helical conformation for either the HIV-1 Rev peptide (Battiste *et al.*, 1996; Ye *et al.*, 1996) or the λ N (Su *et al.*, 1997a,b; Zwahlen *et al.*, 1997) and P22 N peptides (Cai *et al.*, 1998) upon binding to RNA. Many of the ARM proteins are partially or completely unfolded in the absence of RNA and only adopt a stable conformation upon binding to RNA (Calnan *et al.*, 1991; Wolberger, 1996; Van Gilst *et al.*, 1997).

This is in good agreement with the observations made for NusB and its protein–RNA complex. In the three-dimensional structure, the two regions ill defined according to the NMR data, namely the N-terminus containing the ARM motif (residues 1–10) and the loop linking helices 2 and 3 (residues 39–44), are located in close proximity to each other at the edge of the subdomain consisting of helices 1–3. In the spectra of the complex, the Arg-NHε signals of the ARM motif are detectable, suggesting that binding to the recognition site induces a well-defined structure at least for the arginine side chains in the N-terminus. Such a hypothesis is in good agreement with the induced fit observed for other proteins with ARM regions upon binding to RNA. At the same time, not all backbone amides are affected, indicating that the overall structure of the protein remains conserved in the complex.

The fact that only two distinct conformations for the free and complexed protein can be detected in the titration experiments suggests that the lifetime of the complex is above the millisecond range. A more detailed analysis of the complex by NMR is currently in progress.

Materials and methods

Materials

$^{15}NH_4Cl$ was purchased from Isotec (Miamisburg, FL). RNA was synthesized by Xeragon AG, Zürich (Switzerland). Isotope-labelled NusB protein was prepared as described earlier (Berglechner *et al.*, 1997).

NMR spectroscopy and structure determination

All NMR experiments were performed at 22°C on four-channel Bruker DMX600 and Bruker DMX750 spectrometers. Assignment of the ^{1}H, ^{13}C and ^{15}N resonances of the backbone and the side chains has been described previously (Berglechner *et al.*, 1997). Distance information was obtained from a series of ^{15}N- or ^{13}C-resolved three-dimensional NOESY experiments: ^{15}N-NOESY-HSQC (Bax *et al.*, 1990) (uniformly ^{15}N-labelled sample), τ_m = 80 ms; ^{15}N-NOESY-HSQC (uniformly ^{15}N-labelled and fractionally deuterated sample), τ_m = 100 ms; ^{13}C-NOESY-HSQC (Ikura *et al.*, 1990; Zuiderweg *et al.*, 1990) (uniformly ^{13}C/^{15}N-labelled sample), τ_m = 50 ms; ^{15}N,^{15}N-HMQC-NOESY-HSQC (Kay *et al.*, 1990) (uniformly ^{15}N-labelled and fractionally deuterated sample), τ_m = 150 ms; ^{13}C,^{13}C-HSQC-NOESY-HSQC (Clore *et al.*, 1991; Bax and Grezesiek, 1993) (uniformly ^{13}C/^{15}N-labelled sample), τ_m = 50 ms. Amide protons involved in hydrogen bonds were obtained from an analysis of the amide exchange rates measured in MEXICO experiments (Gemmecker *et al.*, 1993) on the deuterated sample with mixing times τ_m of 50, 100, 150 and 200 ms. Stereospecific assignments of the Leu and Val methyl groups were obtained by non-random fractional ^{13}C-labelling (Neri *et al.*, 1989). Side chain amide protons of Asn and Gln were stereospecifically assigned based on an H_2NCO-E.COSY spectrum (Löhr and Rüterjans, 1997).

The structure ensemble for NusB was generated with the X-PLOR program (Brünger, 1992) by a simulated annealing protocol (Nilges *et al.*, 1988, 1991), and subsequently refined by incorporating van der Waals and electrostatic potentials. The structure calculations employed a total of 1926 proton–proton distance restraints. The NOE-derived distance restraints were given upper bounds of 2.2, 2.9, 3.8, 4.9 and 6.0 Å based on the measured NOE intensities. Fifty-one hydrogen bonds were implemented as ambiguous distance restraints between NH_i and CO_{i-3} and CO_{i-4}, respectively, to let the calculation choose the actual donor–acceptor pairing (M.Nilges, personal communication). In addition, 67 Φ angle restraints and 13 $^{3}J_{HNH\alpha}$ coupling constants obtained from the HNHA data (Vuister and Bax, 1993) were also included in the structure calculations. Diastereotopic assignments were obtained by the AQUA-module ASSIGNCHECK (Laskowski *et al.*, 1996).

The coordinates of the ensemble of 18 structures for NusB have been deposited in the Brookhaven Protein Data Bank under the ident code *1baq*.

Acknowledgements

We thank H.Oschkinat and D.Oesterhelt for isotope-labelled algal hydrolysate, and Xeragon AG, Zürich, for a generous gift of synthetic RNA. This work was supported by the Deutsche Forschungsgemeinschaft, the Sonderforschungsbereich 369, the Dr.-Ing. Leonhard Lorenz-Stiftung and the Fonds der Chemischen Industrie.

References

Altieri,A.S., Mazulla,M.J., Zhou,H., Costantino,N., Court,D.L. and Byrd,R.A. (1997) Sequential assignments and secondary structure of the RNA-binding transcriptional regulator NusB. *FEBS Lett.*, **415**, 221–226.

Banner,D.W., Kokkinidis,M. and Tsernoglou,D. (1987) Structure of the ColE1 rop protein at 1.7 Å resolution. *J. Mol. Biol.*, **196**, 657–675.

Battiste,J.L., Mao,H., Rao,N.S., Tan,R., Muhandiram,D.R., Kay,L.E., Frankel,A.D. and Williamson,J.R. (1996) α Helix–RNA major groove recognition in an HIV-1 rev peptide–RRE RNA complex. *Science*, **273**, 1547–1551.

Bax,A. and Grzesiek,S. (1993) Methodological advances in protein NMR. *Acc. Chem. Res.*, **4**, 131–138.

Bax,A., Ikura,M., Kay,L.E., Torchia,D.A. and Tschudin,R. (1990) Comparison of different modes of two-dimensional reverse-correlation NMR for the study of proteins. *J. Magn. Resonance*, **86**, 304–318.

Berglechner,F., Richter,G., Fischer,M., Bacher,A., Gschwind,R.M., Huenges,M., Gemmecker,G. and Kessler,H. (1997) Studies on the NusB protein of *Escherichia coli*. Expression and determination of secondary-structure elements. *Eur. J. Biochem.*, **248**, 338–346.

Berglund,H., Rak,A., Serganov,A., Garber,M. and Härd,T. (1997) Solution structure of the ribosomal RNA binding protein S15 from *Thermus thermophilus*. *Nature Struct. Biol.*, **4**, 20–23.

Bernstein,F.C., Koetzle,T.F., Williams,G.J., Meyer,E.J., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) Protein Data Bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.*, **112**, 535–542.

Biamonti,G. and Riva,S. (1994) New insights into the auxiliary domains of eukaryotic RNA binding proteins. *FEBS Lett.*, **340**, 1–8.

Biou,V., Shu,F. and Ramakrishnan,V. (1995) X-ray crystallography shows that translational iniation factor IF3 consists of two compact α/β domains linked by an α helix. *EMBO J.*, **14**, 4056–4064.

Brünger,A.T. (1992) *X-PLOR Manual. Version 3.1*. Yale University Press, Cambridge, MA,

Burd,C.G. and Dreyfuss,G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615–621.

Bycroft,M., Grünert, S., Murzin,A.G., Proctor,M. and St Johnston,D. (1995) NMR solution structure of a dsRNA binding domain from *Drosophilia* staufen protein reveals homology to the N-terminal domain of ribosomal protein S5. *EMBO J.*, **14**, 3563–3571.

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

Calnan,B.J., Biancalana,S., Hudson,D. and Frankel,A.D. (1991) Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features for RNA–protein recognition. *Genes Dev.*, **5**, 201–210.

Cheng,S.W., Lynch,E.C., Leason,K.R., Court,D.L., Shapiro,B.A. and Friedman,D.I. (1991) Functional importance of sequence in the stem–loop of a transcription terminator. *Science*, **254**, 1205–1207.

Clore,G.M., Kay,L.E., Bax,A. and Gronenborn,A.M. (1991) Four-dimensional $^{13}C/^{13}C$-edited nuclear Overhauser enhancement spectroscopy of a protein in solution: application to interleukin 1β. *Biochemistry*, **30**, 12–18.

Court,D.L., Patterson,T.A., Baker,T., Constantino,N., Mao,X. and Friedman,D.I. (1995) Structural and functional analyses of the transcription–translation proteins NusB and NusE. *J. Bacteriol.*, **9**, 2589–2591.

DeVito,J. and Das,A. (1994) Control of transition processivity in phage λ: Nus factors strengthen the termination-resistant state of RNA polymerase induced by N antiterminator. *Proc. Natl Acad. Sci. USA*, **91**, 8660–8664.

Eberle,W., Pastore,A., Sander,C. and Rösch,P. (1991) The structure of ColE1 rop in solution. *J. Biomol. NMR*, **1**, 71–82.

Friedman,D.I., Sauer,A.T., Baumann,M.R., Baron,L.S. and Adhya,S.L. (1981) Evidence that ribosomal protein S10 participates in the control of transcription termination. *Proc. Natl Acad. Sci. USA*, **78**, 1115–1118.

Gaal,T., Bartlett,M.S., Ross,W., Turnbough,C.L. and Gaurse,R.L. (1997) Transcription regulation by initiating NTP concentration: rRNA synthesis in bacteria. *Science*, **278**, 2092–2097.

Garcia,C., Fortier,P.-L., Blanquet,S., Lallemand,J.-Y. and Dardel,F. (1995) Solution structure of the ribosome-binding domain of *E.coli* translation initiation factor IF3. Homology with the U1A protein of the eukaryotic spliceosome. *J. Mol. Biol.*, **254**, 247–259.

Gemmecker,G., Jahnke,W. and Kessler,H. (1993) Measurement of fast proton exchange rates in isotopically labeled compounds. *J. Am. Chem. Soc.*, **115**, 11620–11621.

Grzesiek,S., Wingfield,P., Stahl,S., Kaufmann,J.D. and Bax,A. (1995) Four-dimensional ^{15}N-separated NOESY of slowly tumbling perdeuterated ^{15}N-enriched proteins. Application to HIV-1 Nef. *J. Am. Chem. Soc.*, **117**, 9594–9595.

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

Hoffmann,D.W., Query,C.C., Golden,B.L., White,S.W. and Keene,J.D. (1991) RNA-binding domain of the A protein component of the U1 small nuclear ribonucleoprotein analyzed by NMR spectroscopy is structurally similar to ribosomal proteins. *Proc. Natl Acad. Sci. USA*, **83**, 2495–2499.

Holm,L. and Sander,C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.

Ikura,M., Kay,L.E., Tschudin,R. and Bax,A. (1990) Three-dimensional NOESY-HMQC spectroscopy of a ^{13}C-labelled protein. *J. Magn. Resonance*, **86**, 204–209.

Kang,C.-H., Chan,R, Berger,I., Lockshin,C., Green,L., Gold,L. and Rich,A. (1995) Crystal structure of the T4 regA translational regulator protein at 1.9 Å resolution. *Science*, **268**, 1170–1173.

Karn,J. and Graeble,A. (1992) New insights into the mechanism of HIV-1 trans-activation. *Trends Genet.*, **8**, 365–368.

Kay,L.E., Clore,G.M., Bax,A. and Gronenborn,A.M. (1990) Four-dimensional heteronuclear triple-resonance NMR-spectroscopy of interleukin-1β in solution. *Science*, **249**, 411–414.

Keener,J. and Nomura,M. (1996) Regulation of ribosome synthesis. In Neidhardt,F.C., Ingraham,J.L., Low,K.B., Magasanik,B., Schaecher,M. and Umbarger,H.E. (eds), *Cellular and Molecular Biology.* American Society for Microbiology, Washington, DC, pp. 822–848.

Kharrat,A., Macias,M.J., Gibson,T.J., Nilges,M. and Pastore,A. (1995) Structure of the dsRNA binding domain of *E.coli* RNase III. *EMBO J.*, **14**, 3572–3584.

Lazinski,D., Grzadienska,E. and Das,A. (1989) Sequence-specific recognition of RNA hairpins by bacteriophage antiterminators requires a conserved arginine-rich motif. *Cell*, **59**, 207–218.

Laskowski,R.A., Rullmann,J.A.C., MacArthur,M.W., Kaptein,R. and Thornton,J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.

Liljas,A. and Garber,M. (1995) Ribosomal proteins and elongation factors. *Curr. Opin. Struct. Biol.*, **5**, 721–727.

Löhr,F. and Rüterjans,H. (1997) H_2NCO-E.COSY, a simple method for the stereospecific assignment of side-chain amide protons in proteins. *J. Magn. Resonance*, **124**, 255–258.

Markus,M.A., Hinck,A.P., Huang,S. and Draper,D.E. (1997) High resolution solution structure of ribosomal protein L11-C76, a helical protein with a flexible loop that becomes structured upon binding to RNA. *Nature Struct. Biol.*, **4**, 70–77.

Mason,S., Li,J. and Greenblatt,J. (1992) A direct interaction between two *Escherichia coli* transcription antitermination factors, NusB and ribosomal protein S10. *J. Mol. Biol.*, **223**, 55–66.

Mogridge,J., Mah,T.-F. and Greenblatt,J. (1995). A protein–RNA interaction network facilitates the template-independent cooperative assembly on RNA polymerase of a stable antitermination complex containing the λ N protein. *Genes Dev.*, **9**, 2831–2841.

Musco,G., Stier,G., Joseph,C., Castiglione Morelli,M.A., Nilges,M., Gibson,T.J. and Pastore,A. (1996) Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. *Cell*, **85**, 237–245.

Nagai,K. (1996) RNA–protein complexes. *Curr. Opin. Struct. Biol.*, **6**, 53–61.

Nagai,K., Oubridge,C., Jessen,T.H., Li,J. and Evans,P.R. (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature*, **348**, 515–520.

Neri,D., Szyperski,T., Otting,G., Senn,H. and Wüthrich,K. (1989) Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C-labeling. *Biochemistry*, **28**, 7510–7516.

Nilges,M., Clore,G.M. and Gronenborn,A.M. (1988) Determination of three-dimensional structures of proteins from interproton distance data by dynamic simulated annealing from a random array of atoms. *FEBS Lett.*, **239**, 129–136.

Nilges,M., Kuszewski,J. and Brünger,A.T. (1991) Sampling properties of simulated annealing and distance geometry. In Hoch,J.C., Poulsen,F.M. and Redfield,C. (eds), *Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy.* Plenum Press, New York, NY, pp. 451–456.

Nodwell,J.R. and Greenblatt,J. (1991) The *nut* site of bacteriophage λ is made of RNA and is bound by transcription antitermination factors on the surface of RNA polymerase. *Genes Dev.*, **5**. 2141–2151.

Nodwell,J.R. and Greenblatt,J. (1993) Recognition of boxA antiterminator RNA by the *E.coli* antitermination factors NusB and ribosomal protein S10. *Cell*, **72**, 261–268.

Puglisi,J.D., Chen,L., Blanchard,S. and Frankel,A.D. (1995) Solution structure of a bovine immunodeficency virus Tat–TAR peptide RNA complex. *Science*, **270**, 1200–1203.

Richardson,J.P. and Greenblatt,J. (1996) Control of RNA chain elongation and termination in *Eschericha coli* and *Salmonella typhimurium*. In Neidhardt,F.C., Ingraham,J.L., Low,K.B., Magasanik,B., Schaecher,M. and Umbarger,H.E. (eds), *Cellular and Molecular Biology.* American Society for Microbiology, Washington, DC, pp. 822–848.

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

Su,L. *et al*. (1997b) An RNA enhancer in a phage transcriptional antitermination complex functions as a structural switch. *Genes Dev.*, **11**, 2214–2226.

Taura,R., Ueguchi,C., Shiba,K. and Ito,K. (1992) Insertional disruption of the *nusB (ssyB)* gene leads to cold-sensitive growth of *Escherichia coli* and suppression of the *secY24* mutation. *Mol. Gen. Genet.*, **234**, 429–432.

Torchia,D.A., Sparks,S.W. and Bax,A. (1988) Delineation of α-helical domains in deuterated staphyloccal nuclease by 2D NOE NMR spectroscopy. *J. Am. Chem. Soc.*, **110**, 2320–2321.

Van Gilst,M.R., Rees,W.A., Das,A. and von Hippel,P.H. (1997) Complexes of N antitermination protein of phage λ with specific and nonspecific RNA target sites on the nascent transcript. *Biochemistry*, **36**, 1514–1524.

Varani,G. (1997) RNA–protein intermolecular recognition. *Acc. Chem. Res.*, **5**, 189–195.

Venters,R.A., Metzler,W.J., Spicer,L.D., Mueller,L. and Farmer B.T.,II (1995) Use of $^1H^N$-$^1H^N$ NOEs to determine protein global folds in perdeuterated proteins. *J. Am. Chem. Soc.*, **117**, 9592–9593.

Vuister,G.W. and Bax,A. (1993) Quantitative J correlation: a new approach for measuring homonuclear three-bond J (HNHα) coupling constants in ^{15}N-enriched proteins, *J. Am. Chem. Soc.*, **115**, 7772–7777.

Wolberger,C. (1996) Homeodomain interactions. *Curr. Opin. Struct. Biol.*, **6**, 62–68.

Xing,Y., Guha Thakurta,D. and Draper,D.E. (1997) The RNA-binding domain of ribosomal protein L11 is structurally similar to homeodomains. *Nature Struct. Biol.*, **4**, 24–27.

Ye,X., Kumar,R. and Patel,D.J. (1995) Molecular recognition in the bovine immunodeficiency virus Tat peptide–TAR RNA complex. *Chem. Biol.*, **2**, 827–840.

Ye,X., Gorin,A., Ellington,A.D. and Patel,J.D. (1996) Deep penetration of an α-helix into a widened RNA major groove in the HIV-1 rev peptide–RNA aptamer complex. *Nature Struct. Biol.*, **3**, 1026–1033.

Zuiderweg,E.R.P., McIntosh,L.P., Dahlquist,F.W. and Fesik,S.W. (1990) Three-dimensional ^{13}C-resolved proton NOE spectroscopy of uniformly ^{13}C-labelled proteins for the NMR assignment and structure determination of larger molecules, *J. Magn. Resonance*, **86**, 210–216.

Zwahlen,C., Legault,P., Vincent,S.J.F., Greenblatt,J., Konrat,R. and Kay,L.E. (1997) Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage λ N peptide/boxB RNA complex. *J. Am. Chem. Soc.*, **119**, 6711–6721.

Received March 2, 1998; revised and accepted May 12, 1998