RNA (1999), *5*:1419–1429. Cambridge University Press. Printed in the USA.

Structure of the phylogenetically most conserved domain of SRP RNA

ULI SCHMITZ,[1] STEFAN BEHRENS,[1] DOUG M. FREYMANN,[2,3] ROBERT J. KEENAN,[2] PETER LUKAVSKY,[1,4] PETER WALTER,[2] and THOMAS L. JAMES[1]

[1]Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446, USA
[2]Howard Hughes Medical Institute and Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143-0448, USA

ABSTRACT

The signal recognition particle (SRP) is a phylogenetically conserved ribonucleoprotein required for cotranslational targeting of proteins to the membrane of the endoplasmic reticulum of the bacterial plasma membrane. Domain IV of SRP RNA consists of a short stem-loop structure with two internal loops that contain the most conserved nucleotides of the molecule. All known essential interactions of SRP occur in that moiety containing domain IV. The solution structure of a 43-nt RNA comprising the complete *Escherichia coli* domain IV was determined by multidimensional NMR and restrained molecular dynamics refinement. Our data confirm the previously determined rigid structure of a smaller subfragment containing the most conserved, symmetric internal loop A (Schmitz et al., *Nat Struct Biol*, 1999, *6*:634–638), where all conserved nucleotides are involved in nucleotide-specific structural interactions. Asymmetric internal loop B provides a hinge in the RNA molecule; it is partially flexible, yet also uniquely structured. The longer strand of internal loop B extends the major groove by creating a ledge-like arrangement; for loop B however, there is no obvious structural role for the conserved nucleotides. The structure of domain IV suggests that loop A is the initial site for the RNA/protein interaction creating specificity, whereas loop B provides a secondary interaction site.

Keywords: 4.5S RNA; complete-relaxation-matrix analysis; NMR; RNA–protein recognition; RNA structure; signal recognition particle

INTRODUCTION

The signal recognition particle (SRP) plays a pivotal role in cotranslational protein targeting and translocation (Walter & Johnson, 1994). In mammalian SRP, six proteins are bound to 7SL RNA, forming an extended complex with a rod-like shape (Andrews et al., 1987). Through phylogenetic comparison, SRP RNA has been divided into four structural domains (I–IV), of which domain IV exhibits a unique secondary structure and is clearly the most conserved among a plethora of SRP RNA homologs (Larsen et al., 1998). On a functional level, the interaction of SRP with an emerging signal sequence and the SRP-receptor has been mapped to the conserved RNA domain IV and its protein associate, SRP54. The SRP homolog of *Escherichia coli*, comprised of 4.5S RNA and the SRP54 homolog, Ffh, (Fig. 1A), exhibits both sequence and functional similarity. For example, Ffh can at least partially substitute for SRP54 in chimeric particles (Bernstein et al., 1993), and the *E. coli* SRP homolog was shown to be essential for proper insertion of a subset of inner membrane proteins (Ulbrandt et al., 1997). Indeed, Ffh/4.5S RNA and a bacterial homolog of the SRP receptor can fully replace the mammalian components in a reconstituted in vitro assay (Powers & Walter, 1997). Consequently, the *E. coli* homolog is an ideal minimal system to elucidate the structural basis of signal sequence recognition and protein targeting.

Reprint requests to: Uli Schmitz, Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446, USA; e-mail: schmitz@picasso.ucsf.edu.

[3]Present address: Department of Molecular Pharmacology and Biological Chemistry, Northwestern University Medical School, Chicago, Illinois 60611, USA.

[4]Present address: Department of Structural Biology, Stanford University, Stanford, California 94305-5400, USA.

Abbreviations: 2D: two-dimensional; 3D: three-dimensional; A43mer: 43mer with ^{13}C,^{15}N-labeled adenines; CT: constant time; DQF-COSY: double quantum filtered correlation spectroscopy; F43mer: uniformly ^{13}C,^{15}N-labeled 43mer; Ffh: SRP54 homolog; FfhM: C-terminal domain of Ffh; G43mer: 43mer with ^{13}C,^{15}N-labeled guanines; HMQC: heteronuclear multiple quantum correlation; HSQC: heteronuclear single quantum correlation; KP: potassium phosphate; NOESY: nuclear Overhauser effect spectroscopy; PAGE: polyacrylamide gel electrophoresis; PFG: pulsed field gradient; rmsd: root-mean-square deviation; SRP: signal recognition particle; TOCSY: total coherence transfer spectroscopy.

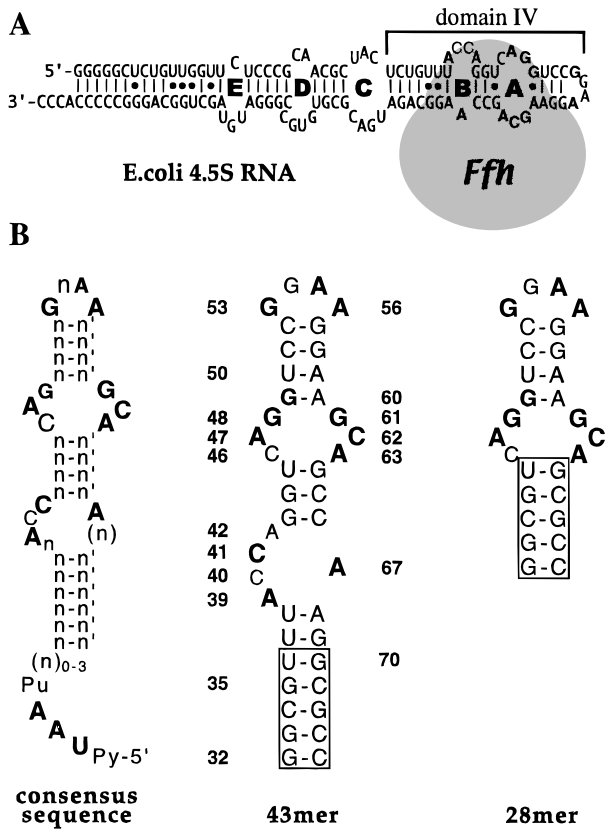

FIGURE 1. **A**: Predicted secondary structure of *E. coli* 4.5S RNA and its putative interaction site with Ffh. Internal loops are lettered A–E. **B**: Domain IV consensus sequence and constructs for NMR structure determination. Phylogenetically conserved nucleotides are in bold type; nonnative nucleotides are in lower case letters; numbering scheme is according to 4.5S RNA.

Ffh consists of two domains: the N-terminal domain, which is mostly involved in the GTP-dependent interaction with the SRP-receptor, and the methionine-rich domain (FfhM), which is involved in signal sequence recognition and SRP RNA binding. Circular dichroism (CD) studies (Zheng & Gierasch, 1997) suggest that FfhM alone adopts a highly α-helical, yet "molten globule" structure that is severely destabilized upon signal sequence binding. The recently determined crystal structure of *Thermus aquaticus* FfhM (Keenan et al., 1998), however, revealed a well-structured fold and confirmed that FfhM consists of four amphipathic helices, of which two are joined by a large loop that is thought to contribute to the signal sequence binding pocket. A region in FfhM that was found essential for RNA binding was defined in *Bacillus subtilis* (Kurita et al., 1996) and maps to a three-helix motif in the *T. aquaticus* Ffh structure. Two of these helices are joined by a short, basic helix, carrying a conserved arginine-rich motif. With respect to the role of SRP RNA, the CD studies mentioned above (Zheng & Gierasch, 1997) suggest that the RNA conveys essential structural stability to FfhM so it can properly bind signal sequences and/or transmit this information to downstream effectors.

The molecular basis of this stabilizing interaction and the RNA's contribution to the processing of the signal sequence binding event is unknown. To this end, we have begun to elucidate high-resolution structures of the RNA and its complex with FfhM. Recently, we reported the solution structure of a 28-nt domain IV fragment (see Fig. 1B) (Schmitz et al., 1999), comprising the most conserved region of domain IV. The structure of the SRP 28mer revealed that the conserved, symmetric internal loop (loop A) adopts a stable, novel motif with multiple cross-strand interactions. All phylogenetically conserved nucleotides are involved in nucleotide-specific hydrogen bonds or other structural interactions. Thus, phylogenetic conservation seems to reflect the requirement of the loop A nucleotides for building a unique and very compact structure that most likely forms a recognition site for FfhM. In FfhM binding studies (Kurita et al., 1996; Schmitz et al., 1996; Zheng & Gierasch, 1997), however, protein footprints observed in chemical and enzymatic modification experiments (Lentzen et al., 1996) indicate that the asymmetric internal loop B is also involved in RNA/protein interactions. Short domain IV fragments with only loop A present (i.e., 24-nt hairpin, comprising *E. coli* sequence G43 through C66, and a 28-nt hairpin as shown in Fig. 1B) bind FfhM with an affinity that is one order of magnitude less tight than that measured for intact 4.5S RNA. A 43-nt hairpin with both loops A and B (43mer, Fig. 1B), however, exhibits essentially the same affinity for FfhM as the entire 4.5S RNA (20–50 nM) (Schmitz et al., 1996), indicating that this portion of the RNA contains all elements required for interaction with FfhM.

Here, we report the NMR solution structure of the 43mer, representing the complete domain IV for *E. coli*. For internal loop A, we confirm the conformation previously seen for the 28mer (Schmitz et al., 1999) indicating that the structures of the two internal loops are independent from each other. Internal loop B, which contains only three conserved nucleotides, was found to be in a semiflexible regime without any distinct structural features involving the conserved nucleotides specifically. The complete domain IV structure in conjunction with the *T. aquaticus* FfhM crystal structure (Keenan et al., 1998) and protein footprinting data (Lentzen et al., 1996) allows an initial glimpse of an unusual RNA/protein interface.

RESULTS

Previous work has shown that physiological concentrations of Mg^{2+} (e.g., 2.5 mM Mg^{2+} with 0.25 mM RNA) lead to structural stabilization of domain IV fragments, such as the 28mer mentioned above (Schmitz et al., 1996; Lukavsky et al., 1997), that is accompanied by some unique characteristics in the ^{1}H NMR spectra. The same Mg^{2+}-dependent features are also evident for the full domain IV fragment. The particular

signature of the Mg^{2+} conditions entails three new G-imino resonances (Fig. 2A), all associated with internal loop A. The NOE cross-peak pattern for these upfield imino resonances are virtually identical for the 28mer and 43mer fragments (Fig. 2B). This is particularly noteworthy for the unusual cross-strand NOEs exhibited by G61NH. This result strongly indicates that the unique structural motif adopted by loop A is independent of the presence of loop B.

Given the prevalence of Mg^{2+} ions in the cytosol of living cells, the Mg^{2+} form of domain IV RNA is likely to be the physiologically relevant state; thus, we determined the structure of the 43mer in the presence of Mg^{2+} ions. Unfortunately, these conditions also lead to increased spectral line widths (up to 50% larger) especially at higher RNA concentrations, which is not uncommon for Mg^{2+}-containing buffers. To test the degree to which line broadening resulted from aggregation, we ascertained the translational diffusion constant (D_t) of the 43mer under various conditions using PFG methods (Butcher et al., 1997; Lapham et al., 1997). Values for D_t in the presence of Mg^{2+} were only marginally decreased compared with the Mg^{2+}-free condition for RNA concentrations spanning a range from 0.14 to 1.4 mM (e.g., 0.82 ± 0.02 10^{-6} cm^2/s vs. 0.79 ± 0.02 10^{-6} cm^2/s for 10 mM Mg^{2+} vs. no Mg^{2+}).The observed line-broadening must be attributed to mild, nonspecific aggregation, which is aggravated over time, as D_t decreases further very slowly over weeks (after two weeks: 0.77 ± 0.02 10^{-6} cm^2/s). Therefore, samples were prepared freshly before each longer NMR-experiment. Moreover, quantitative structural information (NOE intensities) was extracted only from NOESY data obtained at low (0.3 mM) RNA concentrations.

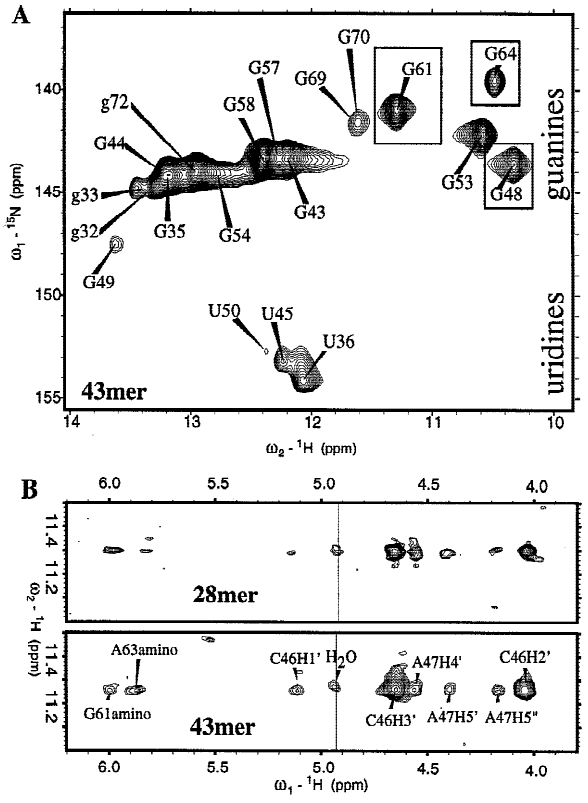

FIGURE 2. **A**: $^1H^{15}N$-HSQC of imino moieties of SRP 43mer in the presence of Mg^{2+}. Boxes indicate peaks are seen only in the presence of Mg^{2+} (10 mM KP, 10 mM $MgCl_2$; 0.3 mM RNA; 15 °C). **B**: NOESY traces for the imino proton of G61 of the 28mer (top) and 43mer (bottom). Mixing time: 120 ms; conditions as in **A**. NOE cross-peaks are labeled with the connecting proton.

Chemical shift assignments

Nearly complete spectral assignments of the nonexchangeable protons have been obtained using a combination of homonuclear and heteronuclear NMR data for different temperatures (20 °C, 30 °C, and 35 °C). A table with chemical shift assignments at 30 °C and spectra showing sequential assignments are available as supplemental material at our web-site (http://picasso.ucsf.edu/supplement.html). Although the chemical shift assignments of the 28mer (Lukavsky et al., 1997; Schmitz et al., 1999) provided enormous help (chemical shifts of residues U45–G64 are almost identical for the two RNA fragments), uniform ^{13}C-,^{15}N-labeling did not permit complete resonance assignments. For the 43mer, which is among the largest RNA fragments whose structure has been solved by NMR methods, the overlap, especially in the ^{13}C-dimension, could only be tackled through reducing the number of observable nucleotides by single-nucleotide labeling of adenine or guanine, respectively (A43mer, G43mer).

^1H,^{13}C-HSQC spectra of A43mer and G43mer (Fig. 3A) and the fully ^{15}N,^{13}C-labeled sample (F43mer) in combination with DQF-COSY and TOCSY spectra, yielded nucleotide-specific base proton assignments, which enabled the tracing of an unambiguous H1′-H8/H6 walk in the homonuclear NOESY data (available as supplemental material). Ribose spin systems were determined using three-dimensional (3D) ^1H,^{13}C-HCCH-COSY and 3D ^1H,^{13}C-HCCH-TOCSY spectra for A43mer, G43mer, and F43mer samples. Three-dimensional ^1H,^{13}C-NOESY-HSQC spectra of all labeled samples and two-dimensional (2D) ^{13}C-edited NOESY spectra of A43mer and G43mer samples yielded structurally important NOEs along with the homonuclear NOE spectra. Sequential assignments of all H2 protons, except those of A39 and A68, were straightforward on the basis of their NOEs to H1′ and other H2 protons. Due to the close chemical shifts of A68H8 and A68H2 and the broad lines of A39H2 (Fig. 3A), long-range correlation experiments failed to resolve these protons, but A39H2 and A68H2 were assigned unambiguously after the first round of structure calculations.

The unusual H1′ chemical shifts that have been reported for the 28mer (Lukavsky et al., 1997; Schmitz

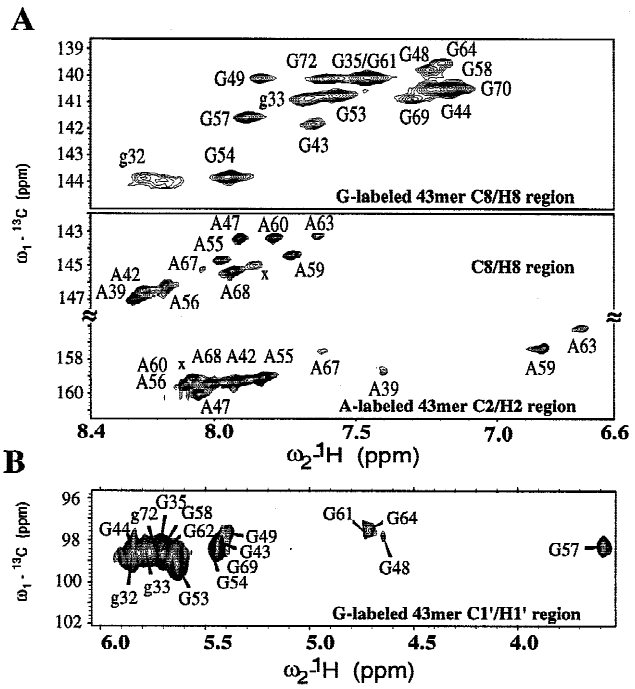

FIGURE 3. **A**: Aromatic region of $^{1}H^{13}C$-HSQC spectra of $^{13}C,^{15}N$-labeled G43mer (top) and A43mer (bottom) (10 mM KP, 10 mM $MgCl_2$; 1.0 mM RNA; 35 °C). An extra adenine arising from the "n + 1"-transcription product is labeled x. **B**: Anomeric region of CT-$^{1}H^{13}C$-HSQC spectra of $^{13}C,^{15}N$-labeled G43mer (conditions as in **A**).

et al., 1999) were also found for the 43mer. Constant time HSQC spectra of G43mer (Fig. 3B) revealed the shifted H1′ resonances of G58 of the tetraloop and G48, G61, and G64 of loop A. This is yet another clear indication that loop A assumes the same structure in the truncated and full domain IV RNA fragments.

Imino protons were assigned on the basis of their NOEs to neighboring imino protons and NOEs to non-exchangeable protons after G- and U-imino protons had been distinguished in 2D $^{15}N^{1}H$-HMQC experiments (Fig. 2A). Spectral dispersion in the imino region is poor, especially in the ^{15}N-dimension, such that heteronuclear NOESY experiments do not provide more information than the homonuclear NOE data. The imino assignments of the shorter RNA fragments (Fig. 1B) were very helpful, because internal loop B of the 43mer did not introduce new imino resonances and because the 43mer's terminal stem is very similar to that of the 28mer. To our surprise, three out of five U-imino protons were extremely broad, with U37 and U38 virtually unobservable in the $^{15}N,^{1}H$-HMQC experiments (Fig. 2A). Even U50, which is expected to be involved in the only A:U base pair, exhibits a fairly solvent-accessible imino proton. Among the G-imino resonances, G69 and G70 are severely overlapped, but assignment was possible because of the observed NOEs and a slightly better separation of the two resonances at lower temperatures.

A significant number of amino protons, especially for the internal loop A, have been assigned as well.

Secondary structure of the SRP 43mer

NOESY data obtained in H_2O show that all putative G:C pairs in the stem regions are formed, consistent with the secondary structure shown in Figure 1B. For A59:U50, all of the NOEs are observed that are typical for an A:U pair stacked between a G:A and a G:C pair. However, U50NH is solvent accessible and the typical NOEs are much weaker than expected. With respect to the putative G:U pairs, NOEs between imino protons indicate formation of U36:G70 and U45:G64 pairs. Although there is no spectral evidence for a G69:U37 wobble pair, some NOEs of the G69 and G70 imino protons along with the sequential NOEs observed in the D_2O spectra suggest that U37 and G69 are stacked on top of the U36:G70 pair similar to a base-paired arrangement. Solvent accessibility of the U38 imino proton and the complete absence of NOEs precludes a canonical pairing between U38 and A68, in contrast to the predicted secondary structure in Figure 1B. No indications were seen for any of the possible A:C pairings involving the residues in loop B. N1-protonated A^+:C pairs were ruled out based on the observation that none of the $^{13}C2$-chemical shifts of the ten A's exhibit a significant dependence on pH over the range 5.0–8.0 (Legault & Pardi, 1994). On top of the conspicuous absence of base pairs, some internal loop residues exhibit differential line widths in the $^{1}H^{13}C$-HSQC spectra (e.g. A67, A39; Fig. 3A). These observations, along with flexible ribose moieties for A39, C41, and A42 ($3\ Hz < J_{H1'H2'} < 6\ Hz$), suggest that the A39–A42 stretch is partially flexible. The number of NOEs observed for this internal loop, including some cross-strand restraints (12.8 average restraints per residue), however, was sufficient for structure calculations.

For loop A, we observed the same pattern as we previously described for the 28mer (Lukavsky et al., 1997; Schmitz et al., 1999). In particular, we obtained no hints for A:C pairings from the spectra. Both loop A guanine imino protons are very sharp, implying their involvement in hydrogen bonds. The unusual NOEs of G61NH indicate its placement near the cross-strand riboses of C46 and A47.

With respect to G:A pairings in the 43mer, G49 and A60 form a carbonyl-amino G:A pair, whereas G53 and A56 of the tetraloop assume a sheared orientation, very similar to that described for the GAAA tetraloop in solution (Jucker et al., 1997) and crystal structure (Cate et al., 1996).

Structure refinement

The structure of the 43mer was determined by generating initial models with the correct fold using DYANA

(Güntert et al., 1997) followed by restrained molecular dynamics (MD) refinement. For the first step, all NOEs were utilized in the common semiquantitative manner (Jaeger & Tinoco, 1993). To ensure a reasonable convergence rate for the DYANA step, backbone torsion angle restraints are essential. In lieu of sequential ^{31}P assignments, which are typically used to establish qualitative A-form backbone torsion angle restraints, we introduced such restraints only for the stem regions, where all of the typical NOESY walks and the sequential NOEs of the imino protons indicated regular A-form geometry. Refinement of the internal loops and the tetraloop was solely driven through distance and sugar pucker restraints.

To increase the precision of the final structure, especially for the internal loops, we used complete-relaxation-matrix-derived distance restraints (Liu et al., 1990, 1994; Schmitz & James, 1995) in the restrained MD refinement. For NOEs involving exchangeable protons and those observed exclusively in 3D NOESY-HMQC spectra, restraints were given conservative bounds (1–6 Å). Overall, a substantial number of distance restraints were obtained with an average of 16.4 per residue (Table 1).

Despite the reasonable number of structural restraints, refinement of the DYANA structures did not produce a completely converged pool of structures, indicated by a relatively high overall atomic root-mean-square deviation (rmsd) for the final ensemble (Table 1). However, regional atomic rmsd revealed that all parts of the 43mer converged well with the exception of asymmetric internal loop B. Figure 4 illustrates the regional convergence and highlights a relative disorder for residues U37–C40. This stretch also exhibits a slightly larger residual restraint violation compared to the rest of the molecule (0.104 Å vs. 0.049 Å). These observations are not surprising in light of the data mentioned above suggesting a partially flexible loop B. The part of loop B containing the excess residues clearly acts as a hinge between the two well-defined parts of the 43mer. The schematic in Figure 4 depicts the effect when two cylindrical segments are joined with a flexible linker on only one side. One must bear in mind, however, that the conical space covered by the terminal stem in our final ensemble should not be mistaken for actual conformational sampling of the molecule, because all structures still try to satisfy all structural restraints at once. On the other hand, the linker region itself is far from being totally disordered. The seven residues in loop B exhibit a total of 51 sequential NOEs and 13 cross-strand NOEs, which overall led to consistent structural features (see below) despite the limited structural precision (average rmsd for ten structures = 2.3 Å, considering residues U38–A42 and A67–A68). The value of averaging coordinates of the final ensemble members, however, is limited for the conformationally heterogeneous portion of a molecule. Even with these caveats, the situation for the linker region in loop B still allowed us to deduce a reasonable structure. This was possible because, for half of the structures in our final ensemble, this region was similar enough (average rmsd for five structures = 1.6 Å) to allow meaningful averaging.

TABLE 1. Statistics for restraints and structural ensemble.

Restraints		
Total number of distance restraints (interresidue)	706	383
MARDIGRAS-derived (average precision)	431	(1.72 Å)
Qualitative, 3D and 2D (1–6 Å)	154	
Qualitative, exchangeable protons (1–6 Å)[a]	121	
Average number per residue	16.4	
Other:		
Base pairing restraints (dist., angle)	60	
Sugar pucker restraints (dihedral)	195	
Backbone (dihedral)	90	
Total restraint average per residue	24.4	
Ensemble parameters	**⟨SA⟩[b]**	**SA[b]**
Distance average deviation (Å)	0.048 ± 0.002	0.048
Average of violations (Å)	0.14 ± 0.01	0.14
Number of violations >0.5 Å[c]	3.9 ± 1.1	2
Angle average deviation (degrees)	0.43 ± 0.08	0.39
Dihedral average deviation (degrees)	0.08 ± 0.01	0.08
Rmsd deviation from ideal geometry		
Bonds (Å)	0.014 ± 0.001	0.014
Angles (degrees)	2.7 ± 0.1	2.7
Atomic rmsd (Å)		
All residues	3.56	
Apical part (nt. 43–66)	1.11	
Top stem, tetraloop (nt 50–59)	0.68	
Bottom stem (nt 32–36, 70–74)	0.78	
Symmetric bulge (nt 46–48, 61–63)	0.61	
Asymmetric bulge (nt 38–42, 67, 68)	2.28	

[a] Most bounds were set to 0–6Å; for some strong NOEs, upper bounds of 2.5 Å or 4 Å were used.

[b] Ten structures were included in the final ensemble ⟨SA⟩ based on the evaluation of restraint violations on a per residue basis. The five structures with the most similar conformation of loop B were averaged and restrained energy-minimized to yield the average structure SA.

[c] No violations exceed 0.7 Å. The larger violations arise mostly from the shorter MARDIGRAS-derived distances with relatively tight bounds.

DISCUSSION

Domain IV of SRP RNA consists of a short stem-loop structure with two bulges that contain the most phylogenetically conserved nucleotides of the molecule. This portion of SRP RNA provides the binding site for SRP54 (or Ffh in prokaryotes), which in turn binds signal sequences and interacts with the SRP receptor. Thus, domain IV and SRP54 together comprise the "business end of SRP" at which all known essential interactions of SRP converge, thereby explaining the extreme phylogenetic conservation of these components. We have solved the NMR structure of domain IV of SRP RNA, one of the largest RNA structures solved by NMR techniques to date. Our data confirm the previously deter-

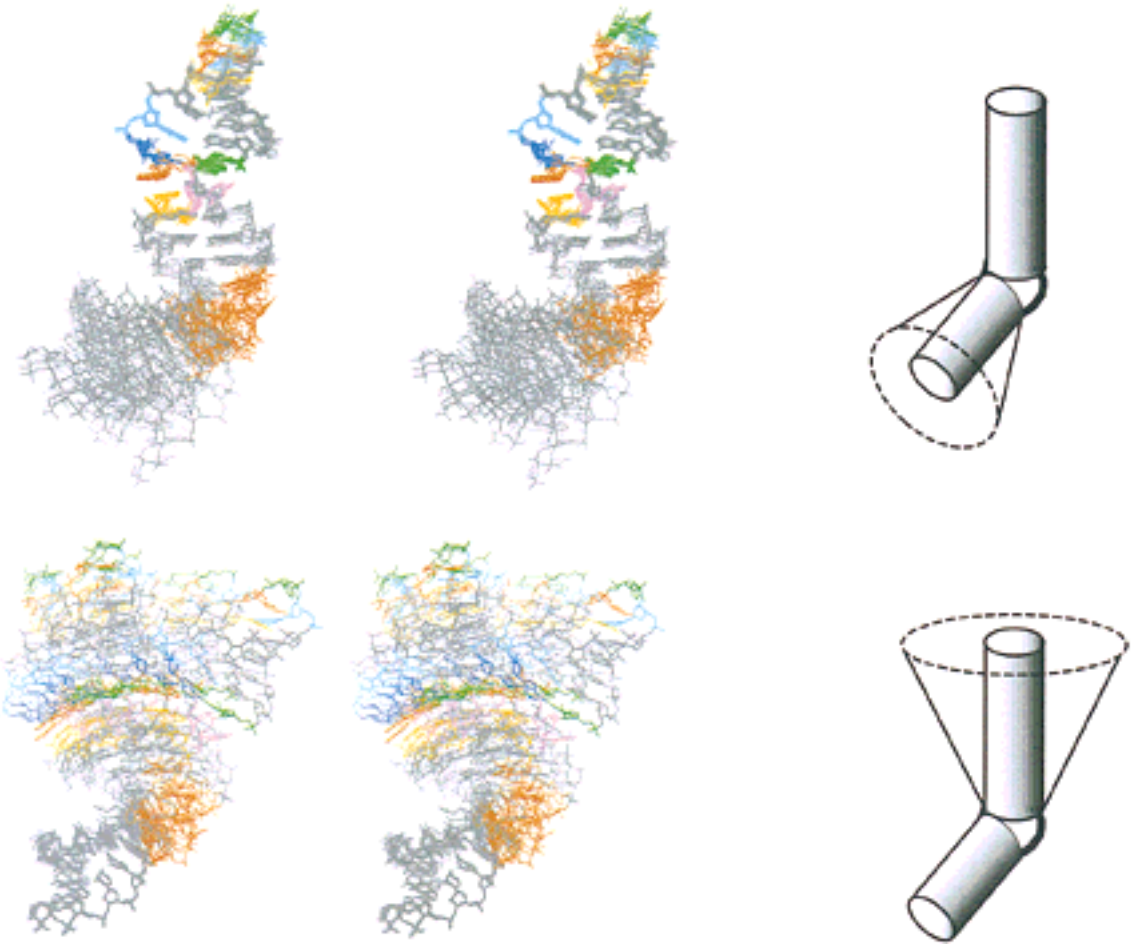

FIGURE 4. Regionally different degrees of convergence. Stereoviews of final ensembles of ten 43mer structures with either residues G43–C66 (top) or g32–U38/A68–c74 (bottom) used for superposition and the corresponding schematic (right), highlighting the effect of a flexible linker on one side between two rigid cylinders. Color coding: non-conserved residues, gray; G49 and G53, cyan; G48, blue; A47 and A55, red; A63 and A56, orange; G54 and G61, green; C62, magenta; the hinge region comprising U38–C41 is shown in red.

mined rigid structure of a smaller subfragment containing loop A. In contrast, loop B emerged as an independently folded module that imposes structural flexibility, effectively providing a hinge between the two adjacent structural elements.

Structural features of the SRP 43mer

The hinged structure is shown in Figure 4A. The final ensemble exhibits a smooth bend, originating from the semiflexible internal loop B. Despite the partial flexibility and the associated limited structural precision, the bend is directionally constrained across the major groove with an overall magnitude of 20–60°. Figure 5A depicts representatives covering the range of structures for the arrangement of loop B. The average structure derived for the most populated cluster of structures (top) and the two most different conformations found in the remainder of the ensemble satisfy the NMR-derived restraints equally well. In all three structures, residues C40, A39, and U38 create a ledge-like extension of the major groove by stacking with each other to various degrees. The curvature that leads to the overall bend of the 43mer is directly related to the type and degree of this stacking. In other words, the direction of the ledge, which can be almost parallel to the plane of the G69:U37 pair, defines the placement of the terminal stem.

Most of the observed conformational heterogeneity is due to the variable stacking arrangement of U38, A39, and C40. The ledge section is flanked by 2 bp that are not as precisely defined as those in the stem regions. Towards the terminal stem, G69 and U37 are mostly stacked on the G70:U36 pair, with a geometry close to that of a wobble pair. Larger deviations from that geometry change the angle of the ledge considerably (see Fig. 5A, bottom). On the apical side, A42 and A67 make an interesting contribution to the ledge motif. Stacked under the G43:C66 base pair, A42 and A67 are found in a tilted, yet pair-like, arrangement,

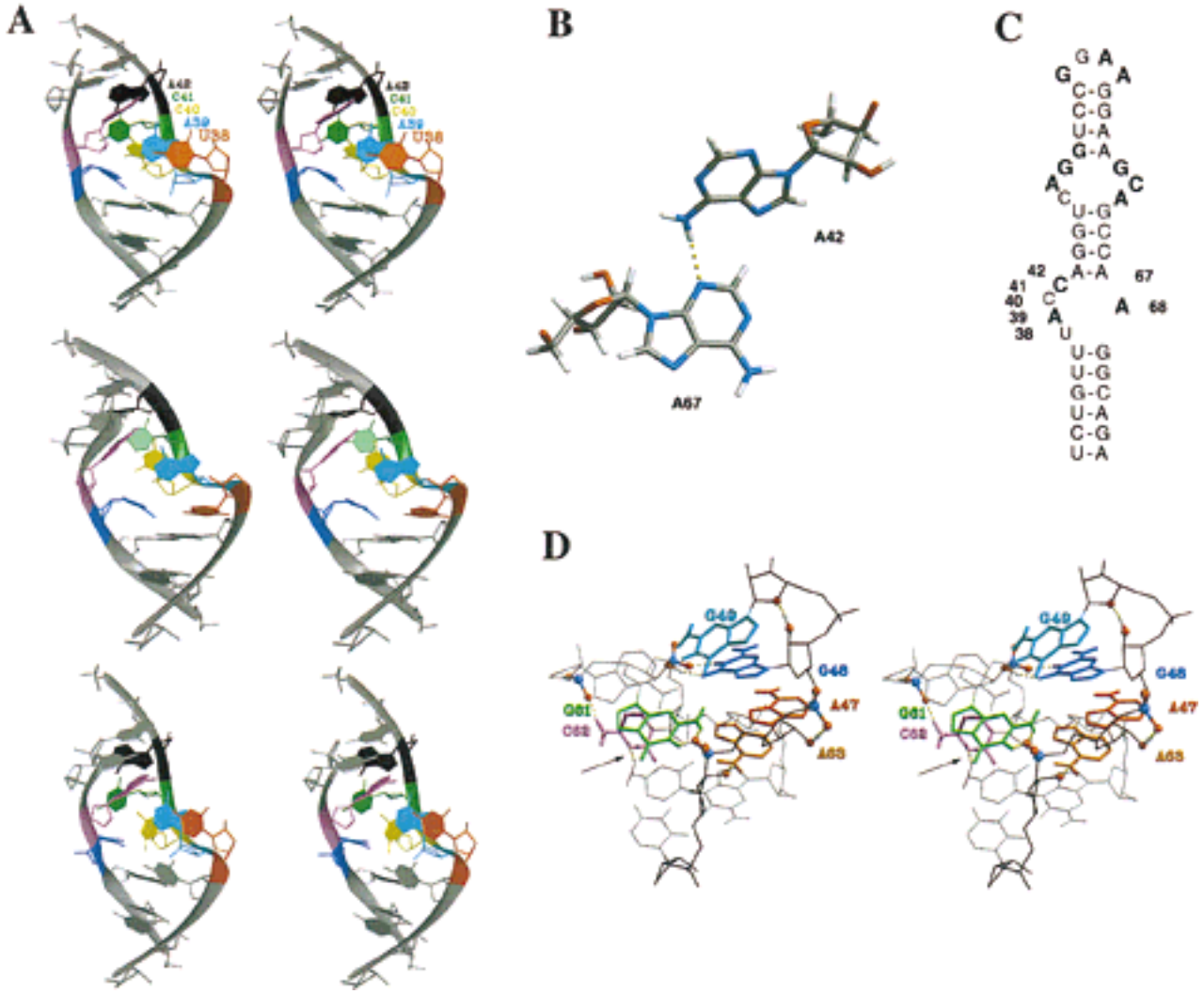

FIGURE 5. Structural details of the SRP 43mer. **A**: Stereoviews of ribbon representations of three conformations of the asymmetric internal loop B. The residues creating the ledge are labeled. Color coding: U38, red; A39, cyan; C40, yellow; C41, green; A42, black; A67, magenta; A68, blue. The conformation at the top is the average structure of the most populated structural cluster (see text); the two conformers below represent the most different conformations in the final ensemble with respect to internal loop B. **B**: Base-pair-like geometry of A67 and A42 as seen in the majority of structures. **C**: Updated secondary structure of *E. coli* domain IV SRP RNA. **D**: Stereoview of the symmetric internal loop A. Conserved bases are labeled and are shown in color. Color coding: nonconserved residues, gray; G49, cyan; G48, blue; A47, red; A63, orange; G61, green; C62, magenta. Backbone atoms involved in hydrogen bonds are shown as balls. The arrows indicates the hydrogen bond between C62 and C46, which was not seen in the structure of the 28mer.

with a putative hydrogen bond seen in the majority of structures between the A42 amino proton and A67N3 (see Fig. 5B). The twist between the putative A:A pair and the G:C pair above is noticeably reduced, thereby contributing to the extension of the major grove. Note that no artificial pairing restraints were used for A42:A67. This geometry is clearly defined through NOEs between A67H2 and both base protons of A42, H2, and H8.

The remaining two loop B residues, C41 and A68, seem to create the space for the ledge to connect well with the 4-bp stem that links to loop A. While C41 exhibits some conformational freedom to roam in the minor groove, A68 is well-defined, always pointing into the minor groove.

The assignment of the close packing of the U38/A39/C40 ledge against the other strand resulted from a number of cross-strand NOEs. Most importantly, A68H8 exhibits NOEs to A39H1′, C40H6, and C40H1′, whereas A68H2 shows NOEs to C40H1′ and both base protons of C41. Further weak cross-strand NOEs between A67H2 and H5, H6, and H2′ of C40, as well as between A39H2 and A67H1′ and A68H8, resulted in pulling the two strands together. In the case of conformational averaging, however, NOE-derived distance restraints are weighted toward the conformation with the shorter distances, such that loop B may appear more compact than it really is.

In light of the predicted secondary structure of *E. coli* domain IV RNA (Fig. 1), we were at first surprised to

find that the predicted A68:U38 pair was not found. The A42:A67 pair, however, extends the stem separating loops A and B to 4 bp, which is the typical length of the consensus sequence found in all other SRP RNAs. A68 then emerges as the conserved adenine on the short side of loop B, whereas U38 becomes a variable linker residue on the long side. Thus, adding the structural information, in fact, improves the line-up of SRP RNA sequences, such that the *E. coli* sequence is no longer an odd domain IV representative (see Fig. 5C).

The structure for internal loop A (Fig. 5D) closely resembles that described for the 28mer (rmsd = 0.88 Å). This part of the 43mer is well restrained and the increased number of distances made up for a decreased precision in the distance restraint set of the 43mer. The salient feature of the unique loop A motif (Schmitz et al., 1999) is the interaction between the Watson–Crick faces of G48 and G61 and opposite-strand phosphate groups enabling a cross-strand stack between A47 and A63. The only noticeable, albeit minor, difference between the two loop A structures is the orientation of the C46 base, which in the 43mer runs almost parallel to the G64:U45 pair below, such that the C46 amino group is within hydrogen bonding distance to C62N3 (arrow in Fig. 5D).

No structural differences were apparent for the apical stem and GGAA-tetraloop despite use of the independent restraint set. The 43mer tetraloop structure is more similar to the GAAA-tetraloop in the P4P6 ribozyme crystal structure (rmsd = 0.82 Å) than to the NMR solution structure (rmsd = 1.12 Å), which was determined under Mg^{2+}-free conditions (Jucker et al., 1997). The 43mer structure that was determined using base-pairing restraints for A59:U50 also shows that it is difficult for a canonical A:U pair to follow the wider carbonyl-amino G49:A60 pair. Some of the strain around A59:U50 in the final structures could be relieved when the U50NH-A59N1 distance restraint was removed.

Comparison with chemical and enzymatic protection data

The structure presented here agrees well with chemical and enzymatic protection data obtained for 4.5S RNA (Lentzen et al., 1996). The match is nearly perfect for internal loop A (Schmitz et al., 1999), where even the modest difference in reactivity of G48 and G61 towards ketoxal is consistent with the different hydrogen bonding patterns.

For loop B, it was noted that all A's were modifiable with dimethyl sulfate (DMS), especially A67 and A68, which led to the conclusion that A68 could not be involved in a canonical A:U pair. In our structure, A68 is completely accessible, and the pairing of A67 with A42 also should not lead to significant protection. Furthermore, the semiflexible ledge section (U38–C40) was found to be accessible to single-strand-specific RNase cleavage (Lentzen et al., 1996), as one would expect from our results.

Differential hydroxyl radical cleavage patterns led to the interesting conclusion (Lentzen et al., 1996) that 4.5S RNA is genuinely bent around loop B. Decreased reactivity was reported for the region that maps to the 3′ strand of the 43mer's terminal stem and the adjoining part of internal loop C, not present in the 43mer. Although our structure reveals a bend originating from loop B, it is not clear how this feature could lead to decreased backbone accessibility of the 3′ strand of the terminal stem. Thus, structural changes caused by other portions of 4.5S RNA would need to be invoked to rationalize these results.

Role of conserved nucleotides and implications for domain IV RNA/FfhM interaction

When comparing loops A and B, dramatic differences are apparent for the structural contributions of the conserved nucleotides. While all five of the conserved nucleotides in loop A are involved in nucleotide-specific structural interactions (Schmitz et al., 1999), no such interactions are observed for loop B. Thus, the fact that no structural function is evident for A39, C41, and A68 strongly suggests that their phylogenetic conservation reflects their requirement for interaction with SRP54 or Ffh.

Chemical and enzymatic footprinting studies with free or Ffh-bound 4.5S RNA revealed unique protection patterns that are mapped on a solvent accessible surface representation of the 43mer average structure in Figure 6. Besides strong protection from chemical modification for A47 and A63 and weaker protection for G48 and G49 in loop A and all adenines in loop B, Ffh binding also protects unique portions of the backbone from cleavage with hydroxyl radicals (residues 35, 36, 39, 42–45, and 47–49) or RNases (residues 32–40 and 51–52) (Lentzen et al., 1996). A few distinct, unconnected patches for FfhM/domain IV contacts emerge that can be discussed in light of the crystal structure of *T. aquaticus* FfhM (see Fig. 6). The part of FfhM involved in 4.5S RNA binding maps to a three-helix motif (Fig. 6, green). When the putative signal sequence binding region and the N-terminal helix (Fig. 6, blue) are omitted, the affinity for 4.5S RNA remains the same (Kurita et al., 1996). The essential amino acids (Arg-387, Gly-391, and Gly-393; Fig. 6, black) whose mutation abolishes RNA binding completely (Kurita et al., 1996), are all located in the helix-turn-helix motif that was proposed to form the core of the RNA-binding site. The RNA protection pattern of the terminal stem and loop B in conjunction with our observation of an extended major groove with a ledge suggests that one of the FfhM helices binds in the major groove around loop B. Following this proposition, however, it is impossible

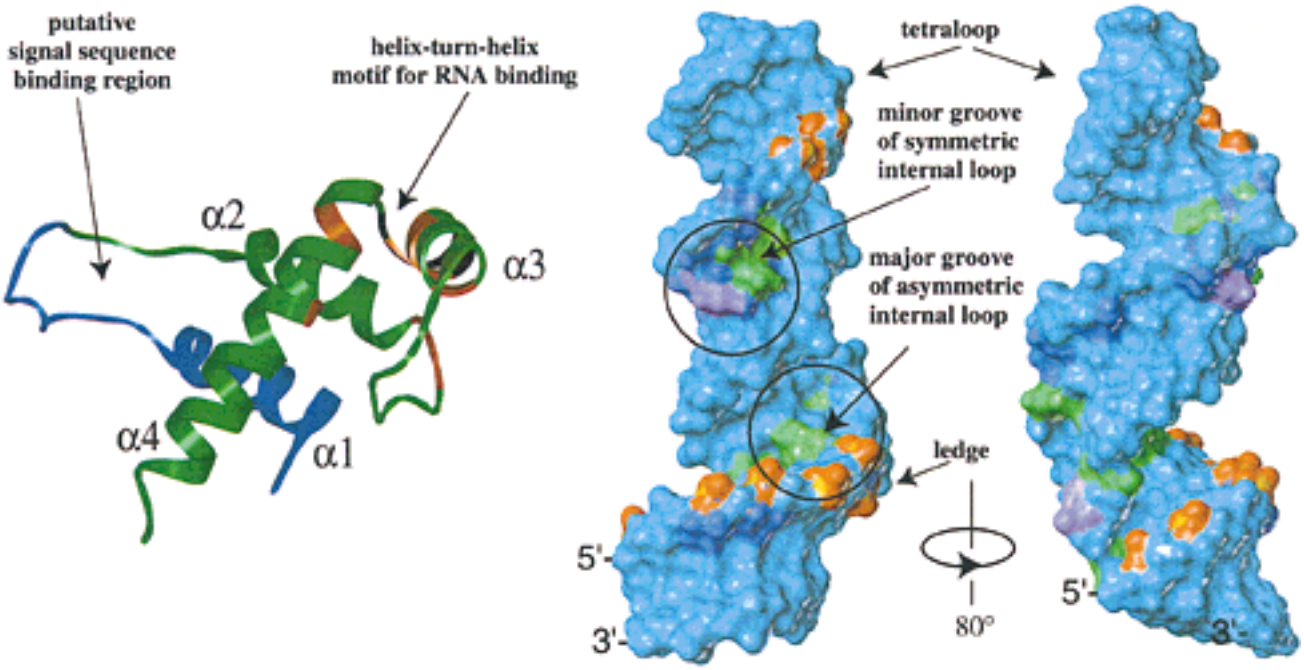

FIGURE 6. Structures of the essential components of SRP involved with signal sequence recognition. Left: ribbon representation of the crystal structure of *T. aquaticus* FfhM (Keenan et al., 1998); Right: two views of a surface representation of the solution structure of the 43mer (most common conformation). The color coding represents information related to RNA protein interactions. For *T. aquaticus* FfhM, the region that is not required for RNA binding is shown in blue. Highly conserved reserved amino acids are shown in red. Three amino acids (Arg387, Gly391, and Gly393 in helix α3 or the turn between α2 and α3) that cannot be mutated without causing a severe drop in the RNA binding affinity are shown in black. For the 43mer, results of Ffh protection experiments are shown; protection seen in chemical modification experiments is shown in magenta (strong) or green (medium); protection in hydroxyl radical cleavage experiments is shown in blue; protection in cleavage experiments with double-helical RNA-specific RNases is indicated by phosphate groups in red. For discussion and references see text.

to dock any of the three suitable α-helices such that any other one would be able to contact the protected areas in the apical part of domain IV without changing the relative orientation of the three helices to each other. Clearly, structural adjustments for one or both of the molecules are necessary to promote the tight and widespread interaction. As CD studies (Zheng & Gierasch, 1997) suggest a "molten globule" state for *E. coli* FfhM, it is easy to imagine that the relative geometry of the α-helices could change to facilitate RNA binding. Using the above information in docking experiments with the domain IV solution structure and the *T. aquaticus* FfhM crystal structure, it seems likely that the helix-turn-helix portion of FfhM (helices α3 and α4) contacts the RNA so that helix α3 interacts with the major groove around loop B and helix α3 interacts with the minor groove side of loop A.

Our binding studies (Schmitz et al., 1996) revealed that the affinity of truncated domain IV fragments (Fig. 1B) for FfhM is still strong and selective (K_{Diss} = 250 nM for 28mer vs. 20 nM for 43mer), such that loop A must contribute substantially to the protein/RNA interactions. In light of the uniqueness and stability of the loop A motif and the greater flexibility of loop B, it seems likely that the rigidly folded loop A motif is the initial, if not prime, recognition site for FfhM, which is responsible for specificity of the interaction. Thus, if the most conserved structural elements of the two components interact with each other and we allow for some structural rearrangements, then it is possible that the short α-helix in FfhM that contains the most conserved amino acids contacts loop A in the minor groove, whereas one of the two larger α-helices binds in the major groove of loop B.

MATERIALS AND METHODS

Preparation of SRP RNAs

All SRP 43mer samples were synthesized by in vitro transcription using T7 RNA polymerase (Milligan & Uhlenbeck, 1989) with either commercially available NTPs or uniformly ^{13}C-,^{15}N-isotope-labeled NTPs. The latter were prepared via enzymatic triphosphorylation of isotope-labeled NMPs that had been extracted from *Methylophilus methylotrophus* cultures (Batey et al., 1995) and were chromatographically separated on a DEAE MemSep1010 column (Batey et al., 1995). Twenty-milliter transcription reactions using either regular

NTPs, all four ^{13}C-,^{15}N-labeled NTPs, or solely ^{13}C-,^{15}N-labeled adenosine 5′-triphosphate (ATP) or guanosine-5′-triphosphate (GTP) produced regular SRP 43mer, fully ^{13}C-,^{15}N-labeled 43mer (F43mer), A-labeled 43mer (A43mer), and G-labeled 43mer (G43mer), respectively. Transcription products were PAGE-purified as described earlier (Schmitz et al., 1996).

NMR sample preparation

The SRP RNA 43mer samples were repeatedly lyophilized from D_2O and dissolved in 10 mM KP, pH 6.5. Samples were denatured at 85 °C and snap-cooled on ice. To minimize the formation of dimeric species, Mg^{2+} was added after annealing. Final conditions were 10 mM KP, pH 6.5, 10 mM $MgCl_2$, and 0.3–0.5 mM RNA in 8 mm Shigemi tubes (homonuclear spectra and most ^1H,^{13}C-HSQC), or 1.0–1.5 mM RNA in 5 mm Shigemi tubes (all other heteronuclear NMR).

NMR experiments

All NMR experiments were acquired at 600 MHz on a Varian Unityplus spectrometer. Homonuclear 2D NOE spectra (mixing time 50, 150, and 400 ms) in D_2O were recorded at 20 °C, 30 °C, and 35 °C (400 t_1 values with 32 scans each, recycling delay 2.0 s). DQF-COSY spectra were acquired similarly. TOCSY experiments were run with MLEV-17 mixing and cycling (Bax & Davies, 1985) (mixing time 30 and 75 ms). Two-dimensional NOE experiments in H_2O were collected at 10 °C using the SSnoesy pulse sequence (Smallcombe, 1993). Translational diffusion constants, D_t, were determined with the water-sLED sequence (Altieri et al., 1995) and subsequent fitting of the intensities of the aromatic signals to equation $\ln(A) = -D_t \ln[\gamma^2\delta^2 G_z^2(\Delta - \delta/3)] - \ln(A_0)$ (Haner & Schleich, 1989; Lapham et al., 1997). Here, A and A_0 are peak intensities at a given gradient strength G_z and without a gradient, respectively; γ is the gyromagnetic ratio of ^1H; δ is the duration of the gradient and Δ is the time between gradients.

One-bond heteronuclear correlations were obtained with $^1H^{15}N$-HMQC or CT-$^1H^{13}C$-HSQC experiments (Santoro & King, 1992). Ribose spin systems were elucidated with 3D-HCCH-COSY and -TOCSY experiments (Nikonowicz & Pardi, 1993). The latter experiment also detected long range H2–H8 correlations (Legault et al., 1994; Marino et al., 1994). To complete assignments and extract structural information, 3D-$^1H^{13}C^1H$-NOESY-HMQC experiments (Marion et al., 1989; Nikonowicz & Pardi, 1993) were collected for all labeled 43mer samples (mixing time 150 ms) and optimized for either observation of aromatic or ribose $^{13}C^1H$ moieties. For A43mer and G43mer samples, 2D ^{13}C-edited NOESY spectra (mixing time 250 ms) were obtained with the ^{13}C selection tailored for base carbons. All spectra were processed with Striker (Day & Kneller, 1992) or NMRpipe (Delaglio et al., 1995) and analyzed with SPARKY (Goddard & Kneller, 1992, 1998).

Distance restraints

NOE peak volumes were quantified for all 2D NOE spectra with SPARKY, which enables deconvolution of peak clusters via line fitting. NOE volumes were then adjusted for complete relaxation with the program SYMM (Liu et al., 1996), based on the difference of cross-diagonal volumes or measured T_1-relaxation times. Accurate distances were generated with MARDIGRAS (Liu et al., 1990, 1994) using a conservative correlation time window (4–8 ns). The randmardi option was used to produce adequate distance error bounds (Liu et al., 1995; Lukavsky et al., 1997), assuming 5–50% error depending on signal-to-noise ratio for the precision of NOE volumes.

Structure refinement

Initial models were generated with DYANA (Güntert et al., 1997), using distance restraints grouped into three categories (1–3 Å, 2–4 Å, and 3–6 Å), A-form backbone torsion angles ($\alpha = -62° \pm 10°$, $\beta = -179° \pm 10°$, $\gamma = 47° \pm 10°$, $\epsilon = -151° \pm 10°$, $\zeta = -73° \pm 10°$) for residues g32–U36, G43–U45, U50–C52, G57–A59, G64–C66 and G70–c74, C3′endo restraints ($\delta = 83° \pm 10°$) for all residues except A39, C41, A42, G54, and A55, C2′-endo restraints ($\delta = 37° \pm 15°$) for residues U38 and G61. Base-pair restraints were used for all G:C pairs, the G:U wobble pairs U36:G70 and U45:G64, and the carbonyl-amino pair G49:A60. The twenty best DYANA structures were refined with AMBER 5.0 (Pearlman et al., 1995; Cornell et al., 1996), using a 15-ps protocol where temperature and the weights for restraints, electrostatic and van-der-Waals contributions to the force field were varied. MARDIGRAS restraints were used only for the AMBER 5.0 refinement. The last 2 ps of each run were averaged and subjected to restrained energy minimization with final force constants of 30 Kcal/mol · Å^2 for distance restraints and 300 Kcal/mol · rad^2 for torsion angle restraints. Final structures were analyzed with CARNAL (Pearlman et al., 1995), CURVES (Lavery & Sklenar, 1990), and MidasPlus (Gallo et al., 1985, 1989).

PDB accession numbers

Protein DataBank accession numbers for the coordinates are 1CQ5 (average structure) and 1CQL (ensemble).

ACKNOWLEDGMENTS

We thank Lara Kipperman and Stefanos Georgantis for help with sample preparation and Drs. N.B. Ulyanov and A. Mujeeb for helpful discussions. This research was supported by National Institutes of Health (NIH) grant GM39247 to TLJ, National Science Foundation (NSF) grant MCB-9513214 to US, and NIH grants to PW. We acknowledge the University of California, San Francisco Computer Graphics Laboratory supported by NIH grant RR01081. PW is an investigator of the Howard Hughes Medical Institute.

Manuscript accepted without revision August 3, 1999

REFERENCES

Altieri AS, Hinton DP, Byrd RA. 1995. Association of biomolecular systems via pulsed field gradient NMR self-diffusion measurements. *J Am Chem Soc 117*:7566–7567.

Andrews DW, Walter P, Ottensmeyer FP. 1987. Evidence for an extended 7SL RNA structure in the signal recognition particle. *EMBO J 6*:3471–3477.

Batey RT, Battiste JL, Williamson JR. 1995. Preparation of isotopically enriched RNAs for heteronuclear NMR. *Methods Enzymol 261*:300–322.

Bax A, Davies DG. 1985. MLEV-17-based 2D homonuclear magnetization transfer spectroscopy. *J Magn Reson 65*:355–360.

Bernstein HD, Zopf D, Freymann DM, Walter P. 1993. Functional substitution of the signal recognition particle 54-kDa subunit by its *Escherichia coli* homolog. *Proc Natl Acad Sci USA 90*:5229–5233.

Butcher SE, Dieckmann T, Feigon J. 1997. Solution structure of the conserved 16 S-like ribosomal RNA UGAA tetraloop. *J Mol Biol 268*:348–358.

Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Szewczak AA, Kundrot CE, Cech TR, Doudna JA. 1996. RNA tertiary structure mediation by adenosine platforms. *Science 273*:1696–1699.

Cornell W, Cieplak P, Bayly CI, Gould IR, Kollman PA. 1996. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc 118*:2309–2309.

Day M, Kneller DG. 1992. *Striker*. San Francisco, California: University of California, San Francisco.

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. 1995. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR 6*:277–293.

Gallo K, Huang C, Ferrin TF, Langridge R. 1985, 1989. *Molecular interactive display and simulation (MidasPlus)*. San Francisco, California: University of California, San Francisco.

Goddard T, Kneller DG. 1992, 1998. *Sparky, NMR display and processing program*. San Francisco, California: University of California, San Francisco.

Güntert P, Mumenthaler C, Wüthrich K. 1997. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol 273*:283–298.

Haner RL, Schleich T. 1989. Measurement of translational motion by pulse-gradient spin-echo NMR. *Methods Enzymol 176*:418–445.

Jaeger JA, Tinoco I Jr. 1993. An NMR study of the HIV-1 TAR element hairpin. *Biochemistry 32*:12522–12530.

Jucker FM, Heus HA, Yip PF, Moors EHM, Pardi A. 1997. A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J Mol Biol 264*:968–980.

Keenan RJ, Freymann DM, Stroud R, Walter P. 1998. Crystal structure of the signal sequence binding subunit of the signal recognition particle. *Cell 94*:181–191.

Kurita K, Honda K, Suzuma S, Takamatsu H, Nakamura K, Yamane K. 1996. Identification of a region of *Bacillus subtilis ffh*, a homolog of mammalian SRP54 protein that is essential for binding to small cytoplasmic RNA. *J Biol Chem 271*:13140–13146.

Lapham J, Rife JP, Moore PB, Crothers DM. 1997. Measurement of diffusion constants for nucleic acids by NMR. *J Biomol NMR 10*:255–262.

Larsen N, Samuelsson T, Zwieb C. 1998. The Signal Recognition Particle Database (SRPDB). *Nucleic Acids Res 26*:177–178.

Lavery R, Sklenar H. 1990. *CURVES 3.0.* Paris: Laboratory for Theoretical Biochemistry, Centre Nationale pour la Récherche Scientifique.

Legault P, Farmer BT II, Mueller L, Pardi A. 1994. Through-bond correlation of adenine protons in a ^{13}C-labeled ribozyme. *J Am Chem Soc 116*:2203–2205.

Legault P, Pardi A. 1994. In situ probing of adenine protonation in RNA by ^{13}C-NMR. *J Am Chem Soc 116*:8390–8391.

Lentzen G, Moine H, Ehresmann B, Wintermeyer W. 1996. Structure of 4.5S RNA in the signal recognition particle of *Escherichia coli* as studied by enzymatic and chemical probing. *RNA 2*:244–253.

Liu H, Borgias B, Kumar A, James TL. 1990, 1994. *MARDIGRAS*. San Francisco, California: University of California, San Francisco.

Liu H, Spielmann HP, Ulyanov NB, Wemmer DE, James TL. 1995. Interproton distance bounds from two-dimensional NOE intensities—effect of experimental noise and peak integration errors. *J Biomol NMR 6*:390–402.

Liu H, Tonelli M, James TL. 1996. Correcting NOESY cross-peak intensities for partial relaxation effects enabling accurate distance information. *J Magn Reson 111*:85–89.

Lukavsky P, Billeci TM, James TL, Schmitz U. 1997. NMR structure determination of a 28 nucleotide RNA from mammalian SRP. *ACS Symp Ser.: Mol Model Nucleic Acids 682*:122–149.

Marino JP, Prestegard JH, Crothers DM. 1994. Correlation of adenine H2/H8 resonances in uniformly 13C-labeled RNAs by 2D HCCH-TOCSY: A new tool for 1H assignment. *J Am Chem Soc 116*:2205–2206.

Marion D, Ikura M, Tschudin R, Bax A. 1989. Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins. *J Magn Reson 85*:393–399.

Milligan JF, Uhlenbeck OC. 1989. Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol 180*:51–62.

Nikonowicz EP, Pardi A. 1993. An efficient procedure for assignment of the proton, carbon, and nitrogen resonances in C-13/N-15 labeled nucleic acids. *J Mol Biol 232*:1141–1156.

Pearlman DA, Case DA, Caldwell JC, Ross WS, Cheatham TE III, Ferguson DN, Seibel GL, Singh UC, Weiner PK, Kollman PA. 1995. *AMBER 4.1.* San Francisco, California: University of California, San Francisco.

Powers T, Walter P. 1997. Co-translational protein targeting catalyzed by the *Escherichia coli* signal recognition particle and its receptor. *EMBO J 16*:4880–4886.

Santoro J, King GC. 1992. A constant-time 2D overBodenhausen experiment for inverse correlation of isotopically enriched species. *J Magn Reson 97*:202–207.

Schmitz U, Freymann DM, James TL, Keenan RJ, Vinayak R, Walter P. 1996. NMR studies of the most conserved RNA domain of SRP. *RNA 2*:1213–1227.

Schmitz U, James TL. 1995. How to get accurate solution structures of double-helical fragments using NMR and restrained molecular dynamics. *Methods Enzymol 261*:1–43.

Schmitz U, James TL, Lukavsky P, Walter P. 1999. Structure of the most conserved internal loop in SRP RNA. *Nat Struct Biol 6*:634–638.

Smallcombe SH. 1993. Solvent suppression with symmetrically-shifted pulses. *J Am Chem Soc 115*:4776–4785.

Ulbrandt ND, Newitt JA, Bernstein HD. 1997. The *E. coli* signal recognition particle is required for the insertion of a subset of inner membrane proteins. *Cell 88*:187–196.

Walter P, Johnson AE. 1994. Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu Rev Cell Biol 10*:87–119.

Zheng N, Gierasch LM. 1997. Domain interactions in *E. coli* SRP: Stabilization of M domain by RNA is required for effective signal sequence modulation of NG domain. *Mol Cell 1*:79–87.