S1 Ribosomal Protein Functions in Translation Initiation and Ribonuclease RegB Activation Are Mediated by Similar RNA-Protein Interactions

AN NMR AND SAXS ANALYSIS

Pascale Aliprandi‡, Christina Sizun§, Javier Perez*, Fabien Mareuill†, Sandrine Caputo†, Jean-Louis Leroy‡, Benoit Odaert†, Soumaya Laalami‡¶, Marc Uzan†**, and Francois Bontems‡‡

Received for publication, August 24, 2007, and in revised form, January 18, 2008. Published, JBC Papers in Press, January 22, 2008, DOI 10.1074/jbc.M707111200

The ribosomal protein S1, in *Escherichia coli*, is necessary for the recognition by the ribosome of the translation initiation codon of most messenger RNAs. It also participates in other functions. In particular, it stimulates the T4 endoribonuclease RegB, which inactivates some of the phage mRNAs, when their translation is no longer required, by cleaving them in the middle of their Shine-Dalgarno sequence. In each function, S1 seems to target very different RNAs, which led to the hypothesis that it possesses different RNA-binding sites. We previously demonstrated that the ability of S1 to activate RegB is carried by a fragment of the protein formed of three consecutive domains (domains D3, D4, and D5). The same fragment plays a central role in all other functions. We analyzed its structural organization and its interactions with three RNAs: two RegB substrates and a translation initiation region. We show that these three RNAs bind the same area of the protein through a set of systematic (common to the three RNAs) and specific (RNA-dependent) interactions. We also show that, in the absence of RNA, the D4 and D5 domains are associated, whereas the D3 and D4 domains are in equilibrium between open (noninteracting) and closed (weakly interacting) forms and that RNA binding induces a structural reorganization of the fragment. All of these results suggest that the ability of S1 to recognize different RNAs results from a high adaptability of both its structure and its binding surface.

In *Escherichia coli*, S1 is the largest (61 kDa) ribosomal protein. It is found in almost all Gram-negative and in several Gram-positive bacteria (1, 2). A shorter form (45 kDa) exists in chloroplasts (3), but S1 is otherwise absent of eukaryotic cells. S1 is involved in translation initiation, and its presence correlates with the way ribosomes recognize the correct translation start on messenger RNAs. In eukaryotic cells, the 40 S ribosomal subunit generally enters at the 5′ end of the mRNA and scans it until it finds the first AUG codon (4). In prokaryotic and chloroplast mRNAs, the first AUG codon is seldom the translation start. During the translation initiation, the 30 S subunit is able to distinguish the initiation codon from synonymous triplets because of the presence of specific signals in the vicinity of the correct start (5, 6). One of these signals is the Shine-Dalgarno sequence (AGGAGG), complementary to the 3′ end of the 16 S ribosomal RNA. This sequence is sufficient in all bacterial species devoid of S1. In the others, S1 likely mediates a second interaction, which was demonstrated to be strictly required in *E. coli* (7). Beside its role in translation initiation, S1 is involved in other functions. It was recently shown to promote transcriptional cycling *in vitro* (8), suggesting that it plays a role in the coupling between transcription and translation. It binds tmRNA (9), even if the physiological significance of this observation in the context of trans-translation is debated (10). S1 is also one of the four subunits of the fr RNA bacteriophage replicases (11). It forms a complex with the phage A protein B involved in general recombination (12). Finally, it is able to accelerate the cleavage rate of the phage T4 endoribonuclease RegB by a factor up to 100 (13).

Sequence analysis of *E. coli* S1 reveals that the protein consists of six repetitions of a conserved structural domain, called S1 domain. This motif is found in many other proteins involved in RNA metabolism in all organisms. Considering this modular organization, several studies were performed to determine the role of each domain. Taking advantage of the existence of a natural cleavage site between the second and third domain (D2 and D3), it was shown that the two first domains bind the ribosome, whereas the four last bind mRNAs (14). By using a mutant disrupted in the domain D6, it was further demonstrated that this domain is dispensable to overall mRNA trans-
Similar RNA-Protein Interactions Mediate S1 Functions

MATERIALS AND METHODS

Protein Expression and Purification

\( ^{15}\text{N} \) and \( ^{15}\text{N},^{13}\text{C} \)-Labeled Fragments—Typically, 25-ml overnight cultures of the BL21(DE3) E. coli host strain transformed with the adequate plasmid (17) were used to inoculate 1 liter of M9 minimal medium supplemented with the following: 1.0 g liter\(^{-1}\) \( ^{15}\text{NH}_4\text{Cl} \) and 4.0 g liter\(^{-1}\) glucose for uniform \( ^{15}\text{N} \) labeling or 1.0 g liter\(^{-1}\) \( ^{15}\text{NH}_4\text{Cl} \) and 3.0 g liter\(^{-1}\) \( ^{13}\text{C}\)glucose for uniform \( ^{13}\text{C} \)\( ^{15}\text{N} \) labeling. Protein expression was induced at \( A_{600} = 0.6 \) by the addition of 1.0 mmol liter\(^{-1}\) isopropyl-\( \beta \)-D-thiogalactopyranoside. The cells were harvested after 3 h, disrupted by sonication. The proteins were purified by Ni\(^{2+}\) affinity chromatography (Qiagen) following the protocol included by the manufacturer. The F5 fragment forms inclusion bodies. After harvesting, the cells were thus resuspended and sonicated in 20 mmol liter\(^{-1}\) phosphate, pH 7.0, 300 mmol liter\(^{-1}\) NaCl and 6 mol liter\(^{-1}\) guanidine chloride. The insoluble material was removed by centrifugation, and the supernatant was incubated in the presence of Ni\(^{2+}\) resin. After several wash cycles, the extraction buffer was first exchanged against 20 mmol liter\(^{-1}\) phosphate, pH 7.0, 300 mmol liter\(^{-1}\) NaCl, and 8 mol liter\(^{-1}\) urea and then against 20 mmol liter\(^{-1}\) phosphate, pH 7.0, and 300 mmol liter\(^{-1}\) NaCl (renaturation buffer). The F5 fragment was then recovered like the others. The proteins were concentrated to 0.5–0.8 mmol liter\(^{-1}\) (depending on the fragment solubility), and the elution buffer was exchanged against 50 mmol liter\(^{-1}\) pH 6.8 phosphate buffer, 200 mmol liter\(^{-1}\) NaCl, 20 mmol liter\(^{-1}\) dithiothreitol by dialysis.

\( ^{15}\text{N},^{13}\text{C} \)\( ^{2}\text{H} \)-Labeled Fragments—An overnight culture was used to inoculate 100 ml of LB medium. The cells were harvested at \( A_{600} = 1.0 \), centrifuged, washed, and transferred into 1 liter of M9 minimal medium supplemented with 1.0 g liter\(^{-1}\) \( ^{15}\text{NH}_4\text{Cl} \) and 4.0 g liter\(^{-1}\) glucose. The cells were again harvested at \( A_{600} = 0.6 \), first transferred into 100 ml of D\(_2\)O M9 minimal medium supplemented with 1.0 g liter\(^{-1}\) \( ^{15}\text{NH}_4\text{Cl} \) and 3.0 g liter\(^{-1}\) \( ^{13}\text{C}\)glucose and, after growth resuming, into 900 ml of the same fresh medium. Protein expression was induced at \( A_{600} = 0.6 \) using 1 mmol liter\(^{-1}\) isopropyl-\( \beta \)-D-thiogalactopyranoside for 12 h. The protein was then purified as described above.

RNA Synthesis

All of the oligoribonucleotides were synthesized chemically on an Amershams Biosciences LKB Gene Assembler Plus apparatus using Amershams Biosciences polystyrene beads for primer support and phenoxyacetil-\( \beta \)-RNA phosphoramidites. The protocol was adapted from DNA synthesis, the main differences being the increase of the coupling time (20 min) and the use of 5-ethylthio-\( ^{-1}\)H-tetrazol as activator (28). The terminal 5’-O-dimethoxytrityl group was removed on a synthesizer at the end of the synthesis. The oligoribonucleotides were cleaved from the support and unprotected according to Leroy and co-workers (28). Full-length product final yield was superior to 95% in all cases. The fragments were purified using a DEAE-Sepharose high pressure liquid chromatography column (Beckman system). All of the species could be separated...
using a linear gradient: 45% A/55% B to 100% B in 80 min (A, 10 mmol-liter\(^{-1}\) phosphate, pH 6.8; B, 10 mmol-liter\(^{-1}\) phosphate, pH 6.8, 1 mol-liter\(^{-1}\) NaCl).

**Size Exclusion Chromatography**

200 µg of F3, F4, and F34 fragments were diluted in the same buffer as used for the NMR studies (50 mmol-liter\(^{-1}\) Na\(_2\)PO\(_4\) pH 6.8, 200 mmol-liter\(^{-1}\) NaCl, 20 mmol-liter\(^{-1}\) β-mercapto-ethanol) and eluted from a Superdex 75 10/30 column at flow rate of 0.3 ml-min\(^{-1}\). The system was calibrated using dextran (200 kDa), ovalbumin (43 kDa), chymotripsinogen (25 kDa), ribonuclease A (13.7 kDa), and aprotinin (6.5 kDa).

**NMR Spectroscopy**

All of the NMR experiments were carried out on a Bruker DRX600 spectrometer equipped with a TXI triple resonance gradient cryoprobe. The data were processed using XWIN-NMR 3.0 or Topspin 1.3 (Bruker) and analyzed with Sparky software (Thomas L. Goddard, University of California, San Francisco). \(^1\)H and \(^13\)C chemical shifts were referenced to TSP. \(^15\)N chemical shifts were indirectly referenced from the \(\gamma^\text{HN}/\gamma^\text{H}\) ratio (29). All of the spectra were recorded at 303 K.

The resonance frequency assignment of the backbone (H\(^\alpha\), N, C\(^\text{C}\), and C\(^\text{C}\) atoms of the fragments F34 (Asn179–His361) and F45 (Trp267–Asn448) was obtained by recording and analyzing HNCO, HNCA, HN(CO)CA, HNCACB, and HN(COCA)CB triple resonance experiments on \(^{15}\)N\(^{13}\)C\(^2\)H-labeled samples of the two proteins. From this, it was possible to assign the H\(^\alpha\), N, and C\(^\text{C}\) backbone atom resonance frequencies of the fragments F4 (Trp267–His361) and F5 (Trp354–Asn448) and F3–5 (Asn179–Asn448) by recording \(^1\)H\(^{15}\)N-HSQC\(^3\) and HNCO experiments on \(^{15}\)N\(^{13}\)C-labeled samples of each of them in identical conditions to those used for F34 and F45. For the F3 fragment (Asn179–Tyr274), we observed that many peaks underwent large shifts compared with the F34 spectrum. Thus, we assigned its backbone atom resonance frequencies de novo by using HNCO and HNCA experiments recorded on a \(^{15}\)N\(^{13}\)C-labeled sample.

The RNA binding studies were performed on 0.3 mmol-liter\(^{-1}\) \(^{15}\)N\(^{13}\)C-labeled samples of the F3–5 fragment in which aliquots of the lyophilized RNAs were dissolved. Prior to lyophilization, RNAs were dialyzed against pure water and neutralized. Each titration point was done by removing the sample from the NMR tube, mixing it with lyophilized RNA, and returning it to the tube. Five \(^1\)H\(^{15}\)N-HSQC and three HNCO spectra were recorded for each series corresponding to RNA/protein ratios of 0, 0.25, 0.5, 0.75, and 1.0 (HSQC) and 0, 0.5, and 1.0 (HNCO). All of the recording parameters were kept rigorously identical throughout the experiments. The titration experiments were duplicated.

**SAXS Experiments and Curve Fitting**

SAXS data were collected at DESY on Beamline X33 (F3 and F4 fragments) and at the European Synchrotron Radiation Facility on Beamline ID02 (F34 and F45 fragment). The His tag was removed from the proteins by thrombin cleavage. Mass absorption coefficients of each species were calibrated using amino acid decomposition. Accordingly, the mass concentration of each sample could be properly measured by absorption at 280 nm just before data collection (F3, 1.3 gliter\(^{-1}\); F4, 1.6 gliter\(^{-1}\); F34, 12.45 gliter\(^{-1}\); F45, 8.13 gliter\(^{-1}\)).

At DESY, collection time was 180 s for each sample and its corresponding buffer. No aggregation or denaturation under radiation was observed after a second exposure. The useful Q range (Q = 4πsin(θ)/λ, where λ = 1.5 Å and 2θ is the scattering angle) was 0.17–4.85 nm\(^{-1}\), with the detector off-centered. The intensity at Q = 0, I(0), and the radius of gyration, \(R_g\), were determined by linear fit of Ln(I(Q)) versus Q\(^2\) in the range 0.2–0.6 nm\(^{-1}\) using PRIMUS (30). The average apparent molecular mass of the molecules in solution was deduced from I(0) by normalization with a scattering pattern of a bovine serum albumin solution with known concentration. At the European Synchrotron Radiation Facility, a dilution series (C, C/2, and C/4) was measured for each protein. 50 frames of 1 s were averaged for each sample, with the protein solution being pushed between two frames through a circulating quartz capillary cell. The new CCD detector based on a Kodak 4320 chip gave extremely reproducible measurements. The available Q range was 0.056–3.03 nm\(^{-1}\). The data were normalized by transmitted intensity, measured with a diode inserted in the beamstop. A slight and monotonic evolution with concentration was seen at small angles on the buffer subtracted patterns, indicating some intermolecular interactions. Therefore subsequent analysis was systematically made on merged patterns using the sample with lowest concentration at low angles and that with the highest concentration at high angles.

Atomic models of the F34 and F45 fragments were fitted to the SAXS curves by running Dadimodo software (31) on the SOLEIL PC cluster. Dadimodo is a genetic algorithm designed to refine deformable homology models of multi-domain proteins using SAXS and/or NMR data. The input consisted of five sets of 25 structures obtained by connecting the C- and N-terminal ends of the D3 and D4 homology models for F34 (D4 and D5 for F45) and randomizing the geometry of the linker. Fitting the F34 SAXS curve required small modifications of the secondary structure element geometry, which was not the case for F45. One of the crucial aspects of Dadimodo is that all accepted solutions satisfy an energy criterion that ensures that they present neither geometrical deformations nor steric clashes.

**Homology Modeling**

S1 domains are classified as OB fold/nucleic acid-binding protein/cold shock DNA-binding domain-like in the SCOP data base (32). Thirty structures belonging to this family were retrieved, and their sequences and topologies were compared with those of the domains of S1. Taking the length of the secondary structure elements and of the loops between them into account, four sequences were retained: 1SRO, 1KL9, 1LUZ, and 1GO3. The 1SRO sequence could be aligned with those of the three domains with no or one insertion. In addition, many residues belonging to the loops are conserved. The alignment of 1KL9 and 1LUZ only requires insertion of two residues in the loop between the second and third strand of the motif. Finally,
alignment of 1GO3 requires a 10-residues insertion in the long loop connecting the third and fourth strands, but a visual inspection of the structure reveals that this insertion corresponds to a small independent structural domain (a triple stranded β-sheet) and that the beginning of the loop is similar to the corresponding region in the 1KL9 structure. These four structures appeared to be good templates to model the structures of the S1 domains. However, in the case of 1KL9, the end of the long loop between the third and fourth strand is not resolved. We thus decided to use 1SRO (S1 domain of the E. coli polynucleotide phosphorylase (33)), 1LUZ (vaccinia virus K3L protein (34)), and 1GO3 (S1 domain of Methanococcus jannaschii RNA polymerase II (35)). The models were calculated using the Modeler software (36) with the standard confidence level in the prediction of this region.

RESULTS

Homology Modeling of the F3, F4, and F5 Fragments—We built models of S1 domains D3 (Asn179–Pro266), D4 (Trp267–Pro352), and D5 (Trp354–Pro449) by using an NMR structure of the S1-type domain of E. coli polynucleotide phosphorylase (1SRO), the x-ray structure of the S1-type domain of archael M. jannaschii DNA-directed RNA polymerase (1GO3), and the x-ray structure of KL3 protein from the vaccinia virus (1LUZ). The sequences of these three template proteins are easily aligned with the D3, D4, and D5 domains of S1 (supplemental Fig. S1). In particular 43, 39, and 41% of D3, D4, and D5 amino acids are conserved in at least one template sequence, and 25, 26, and 24% are conserved in at least two template sequences. In addition, there is always one possibility with the correct length for each loop to be predicted. The three calculated model families are very similar, reflecting the sequence conservation between the three domains, with the exception of the long loop between the third and fourth strands. Many different solutions are obtained for this loop in each model, indicating a low confidence level in the prediction of this region.

Assignment of the Resonance Frequencies of the Fragment Backbone Atoms—One of our goals was to analyze the interactions between the F3–5 fragment and several of its RNA targets by NMR. Accordingly, we wanted to assign as many backbone atom resonance frequencies as possible. Even if the 1H-15N-HSQC spectrum of the F3–5 fragment looks well, the size of the protein (267 amino acids) makes a direct study difficult. A preliminary analysis revealed the existence of large peak shifts between the HSQC spectrum of F3–5 and those of the F3, F4, and F5 fragments. On the opposite, the differences between the bi-domain and three-domain fragment spectra appeared much less striking. We therefore decided to assign the backbone atom resonance frequencies of the F34 and F45 fragments and to use this information to facilitate the assignment of F3–5.

As a first task, we looked for buffer and temperature conditions simultaneously suitable for all the fragments. Well dispersed spectra were obtained in 50 mmol liter–1 pH 6.8 phosphate buffer, 200 mmol liter–1 NaCl, 20 mmol liter–1 dithiothreitol at 303 K (supplemental Fig. S2). We could not obtain correct HNCA CB experiments on 15N–13C-labeled samples of the F34 and F45 fragments. However, by using 15N–13C–2H triple labeled samples, we were able to record HNCO, HNCA, HN(CO)CA, HN(CA)CB, and HN(COCA)CB experiments on both proteins and to assign all but a few number of the HN, N, Ca, Cβ, and Cγ (carbonyl) resonance frequencies. The missing assignments correspond to Thr253 (D3 domain in F34), Ala269 ( linker between the D3 and D4 domains in F34), Cys439 (D4 domain in F34), and Gln356, Phe357, and Ala358 (linker between the D4 and D5 domains in F45). The assignments obtained for F34 and F45 allowed us to assign all HN, N, and Cα resonance frequencies of F4, F5, and F3–5 fragments by simply comparing the HSQC and HNCO spectra. In the case of the F3 fragment, the chemical shift variations were too large for a direct assignment transfer from F34. So we confirmed it (with the exception of the Gln251–Thr253 loop that could not be reassigned) by recording an HNCA in addition to the HSQC and HNCO spectra. Accordingly, we obtained the HN, N, and Cα resonance frequencies of all fragments, the Cβ resonance frequencies of F3, F34, and F45, and most of the Cγ resonance frequencies of F34 and F45.

Characterization of the Interfaces between the Modules—NMR is a very sensitive tool to analyze local interactions and deformations in proteins because any change in an amino acid environment will modify the resonance frequencies of its atoms. Accordingly, by comparing the spectra of two neighboring domains, like isolated (F3, F4, and F5) and associated (F34 and F45) fragments, it is possible to determine whether they interact with each other and to localize the directly or indirectly affected residues. Similarly, by comparing the spectra of bi-domains F34 and F45 with that of three-domain F3–5, it is possible to decide whether the third and fifth domains are in contact.

Comparison of the 1H–15N-HSQC spectra, which correlate the amide nitrogen and amide proton frequencies, recorded on the mono- and bi-domain fragments is reported in Fig. 1. It clearly shows noticeable differences between the spectra of isolated or associated domains. The simplest situation seems to be that of the D4 and D5 domains. Two regions of the D4 domain are strongly affected in the presence of D5 within the F45 fragment: its C-terminal extremity (Leu346–Asn357) and a fragment (His305–Ser319) belonging to the long loop (L3) inserted between the third and the fourth strands (S3 and S4) of the β-barrel and located at its C-terminal edge (Fig. 2). Similarly, two regions of the D5 domain are affected. The first corresponds to the end of β-strand S2, the beginning of S3 and residues of the L2 loop between them (Glu362–Ile387). The second corresponds to the end of β-strand S4 and to residues in the following L4 loop (Val422–Ala424). These two regions are located at the N-terminal edge of the β-barrel (see Fig. 4). Nothing can be told concerning the N-terminal fragment of the protein, because the Gln356–Ala358 resonances could be observed neither in the F5 nor in the F45 fragment. This strongly suggests that D4 and D5 domains are interacting in the F45 fragment. The case of the D3 and D4 domains is more puzzling. A very broad area of the D3 domain is affected in the presence of D4, and the chemical shift changes are very important (up to 0.7 ppm), whereas the only residues affected in the D4 domain are those of its N terminus, the isolated Ser319 at the other end of the molecule, and maybe Val299 and Val301 located in the L4.
loop between the S4 and S5 strands of the β-barrel, close to the N terminus. Concerning the D3 domain, this suggests a strong interaction, whereas for the D4 domain, it means at best a weak interaction.

We also compared the $^1\text{H}-^{15}\text{N}$-HSQC spectra recorded on F34, F45, and F3–5 fragments (not shown). The differences observed on D3 and D5 domains are much smaller than those previously noticed. This strongly suggests that there is no con-
Similar RNA-Protein Interactions Mediate S1 Functions

FIGURE 2. Mapping of the domain/domain interactions on the surface of the domain models. The chemical shift variations determined by the comparison of the HSQC spectra recorded on the mono- and bi-domain fragments were mapped on the homology models of the domains. The color code is identical to that used for the chemical shift variation scale in Fig. 1.
Contact between them in the F3–5 fragment. There are marked changes of the domain D4 resonance frequencies, but they correspond to the changes already characterized because of the influence of the D3 domain (by comparing the spectra of F45 and F3–5 fragments) of and the D5 domain (by comparing the spectra of F34 and F3–5 fragments).

Characterization of the F3 Dimerization Surface—The discrepancy between the large perturbed area observed in the case of the domain D3 and the much smaller area observed in the case of D4 led us to check the fragment multimerization states. Our analysis was indeed based on the assumption that F3 and F4 1H–15N-HSQC spectra reflect the properties of the isolated D3 and D4 domains, whereas that of F34 reflects the properties of the D3/D4 interdomain interaction. But this is no longer true if monomer-multimer equilibriums take place. To verify this point, we ran gel filtration experiments of the F3, F4, and F34 fragments using the same buffer as in the NMR experiments. The results (Fig. 3) show that the F3 and F4 elution volumes are clearly different. The F4 and F34 volumes are compatible with the fragment molecular masses, whereas that of F3 (23 kDa instead of 13 kDa) is indicative of a monomer-dimer equilibrium. This was confirmed by SAXS analysis (see below). Thus, the D3-D4 interaction surface deduced from the HSQC comparisons is correct for the D4 domain but not for D3. In the latter case the perturbed area corresponds to the sum of the interaction and dimerization surfaces.

To evaluate the extension of the F3 fragment dimerization surface, we compared HSQC spectra recorded at two different concentrations (0.7 and 0.1 mmol liter−1). As shown by Fig. 3, there are small but significant differences between these two spectra. When reported on the structure of the D3 domain, these differences delineate a surface very similar to that obtained by comparing the F3 and F34 HSQCs. In fact, the only residues affected in the F3/F34 comparison not belonging to the dimerization surface are in the C-terminal linker and in the L3 loop in its vicinity. It appears, in conclusion, that the region of the D3 domain in interaction with D4 is limited, confirming that there are probably only weak interactions between them.

Characterization of the Interactions between the F3–5 Fragment and RNA—The interactions between the F3–5 region of S1 and three RNAs were characterized by the same method as the interfaces between the domains. We recorded 1H–15N-HSQC and HNCO spectra of the protein in the presence of increasing amounts of similar RNA-protein interactions mediate S1 functions.
Variations are indicative of a fast to intermediate exchange regime between free and RNA-bound protein and suggest an affinity in the 1/H9262M range. However, the observed changes do not derive from unspecific binding caused by electrostatic effects, because the isoelectric point of F3–5 is ~5, and we previously observed that the modification pattern induced by S26 is independent on the salt concentration (20–200 mM; data not shown).

**FIGURE 4. Mapping of the RNA/F3–5 interactions on the surface of the domain models.** We mapped the residues whose ΔH1H (variations of the amide hydrogen chemical shifts; in blue) and IPI (intensity variations; in red) deviate by more than one S.D. from the mean on the domain structures. When both measured ΔH1H and IPI deviate by more than one S.D., they are colored magenta. The first, second, and third rows correspond to the effects induced by S26, motB, and gen1, respectively. The first, second, and third columns correspond to the domains D3, D4, and D5, respectively. Contour plots of representative regions of the HSQC spectra of F3–5 exemplify the observed chemical shift and intensity variations for the three probed RNAs at various RNA:protein ratios (0:1 in red, 0.5:1 in blue, and 1:1 in green). The assigned residues are reported in the same color code as ΔH1H and IPI.
We identified all of the residues for which IPI and ΔH^N deviate by more than a standard deviation from the mean, in each domain and in the presence of each RNA fragment. Fig. 4 shows their positions in the domain structures. Each RNA induces spectral changes for the three domains, and the perturbations concern a similar region. It is formed of β-strands S2, S3, and S5; of the loops connecting S1 and S2 (L1), S2 and S3 (L2), and S4 and S5 (L4); and of the last residues of β-strands S1 and S4.

Interestingly, the binding surface of each domain corresponds to the same β-barrel side of the OB fold. A similar surface has been defined by NMR and x-ray analysis for the interactions between two other S1-type domains, the anticodon recognition domains of the lysyl- and aspartyl-tRNA synthases, and their cognate tRNA syntheses, and their cognate tRNAs (38, 39). For each domain, some residues are affected by all three RNAs (involved in systematic interactions). We reported them in magenta in Fig. 5 to compare them with the homologous residues of the two tRNA syntheses and the residues postulated by Draper and Reynaldo at the end of the S4 strand, is not only mediated through systematic interactions but also through more specific interactions. Finally there are differences between the domains. For example, an aspartate determinant, identified by Draper and Reynaldo at the end of the S4 strand, is systematically affected in the D4 and D5 domains, but not in the D3 domain.

To circumvent the choice of a threshold to determine affected residues, we completed this analysis by calculating correlation coefficients between ΔH^N induced by two RNAs. Table 1 shows that there is a similar, low but significant ΔH^N correlation over the whole F3–5 fragment. This corroborates the idea of a common binding surface and may point to a similar RNA binding mode in F3–5. The coefficients are more homogeneous for the D3 and D5 domains than for the D4 domain, which suggests that D4 could play a particular role in the adaptation of the F3–5 fragment when binding different RNAs.

We also noticed large perturbations in the linkers between the domains, and in the long loop (L3) connecting strands S3 and S4 in the D3 and D4 domains, but not in the L3 loop of the D5 domain. Because we showed that the L3 loops of the D3 and D4 domains are involved in interdomain interactions, this
strongly suggests that RNA binding induces perturbations in the relative positioning of the domains.

**SAXS Analysis of the Fragments**—The SAXS curve, \( I(Q) \), is the Fourier transform \( I(Q) = \int dr \, p(r) \sin(Qr)/Qr \) of the molecule distance distribution function, \( p(r) \). Although this information is *a priori* very degenerated, it directly provides geometric characteristics of the studied object, such as its mass, radius of gyration, or longest distance (41). It can also be used quite sensitively to discriminate between different models by checking the agreement between back-calculated and experimental curves.

Gel filtration results obtained for the F3 and F4 fragments motivated us to confirm the existence of a monomer-dimer equilibrium in the case of the F3 fragment. The curves (normalized by the protein mass concentrations) recorded on F3 and F4 are compared in Fig. 3. The two curves cross each other because of higher \( I(0)/C \) and slope for F3 (F3 over F4 \( I(0)/C \) ratio of 1.22). Whereas the F4 curve is compatible with the fragment mass and radius of gyration estimated from the homology model, the F3 curve corresponds to a larger object, both in mass and volume, confirming the presence of a non-negligible amount of dimers.

SAXS curves of the F34 and F45 fragments were recorded on a subsequent run. In both cases, \( I(0)/C \) is compatible with the expected mass, confirming that both fragments are monomeric in solution. At small \( Q \) values, the two concentration normalized curves are very close to each other, with \( R_g \) values compatible with elongated structures (25.2 and 24.6 Å instead of 18 Å for a spherical protein of equivalent mass). At larger \( Q \) values, however, the F34 curve falls well above that of the F45. The F45 curve almost follows Porod’s law, which should be followed by any globular rigid structure, whereas the F34 curve strongly diverges from it. This suggests that the F45 fragment possesses a well-defined surface, whereas the F34 fragment is more flexible (42). In addition, the absence in the F34 curve of a slight but apparent inflection (at \( Q = -0.12 \) nm\(^{-1}\)) shown by the F45 curve could also be explained by the averaging of the scattering curves from different conformers smoothing out specific features like marked inflections.

To go beyond this first analysis, we built structural models of the F34 and F45 fragments by using the Dadimodo genetic algorithm (31). Dadimodo was designed to provide structural models of multi domain proteins in the presence of SAXS and/or residual dipolar coupling data. In the absence of residual dipolar coupling data, we obtained, for each fragment, different structural solutions that perfectly fit the SAXS curves (Fig. 6).

However, in all of the F45 solutions, the two domains are in contact, whereas in all of the F34 solutions the two domains are disjoined. In this latter case, NMR observations (a limited contact area between the domain D3 and D4) and SAXS results (a disjoined structure) strongly suggest the existence of an equilibrium between a state where the two domains are weakly interacting and another where they are disjoined, the equilibrium being displaced toward the second state.

### DISCUSSION

The way S1 promotes its different functions is far from being elucidated. In the case of translation initiation, S1 was first proposed to be an unwinding protein that would facilitate the ribosome progression and the Shine-Dalgarno-16 S interaction by disrupting mRNA secondary structures (14, 43). In this model, S1 is supposed to interact with mRNA before it enters the ribosome, but S1 seems located in the cleft between the head and the platform of the 30 S subunit, close to the 3’ end of the 16 S rRNA (anti-Shine-Dalgarno region) and, thus, to the 5′ end of bound mRNA (22, 44). It was also shown that in the translation initiation complex S1 binds U-rich regions of Qβ and fr phage RNAs located upstream of the translation initiation codon of the coat protein gene (24). In the same study, it was shown that a U-rich region is often found before the Shine-Dalgarno sequence in *E. coli* mRNAs and that this region is protected by S1 against RNase degradation in the *ssb* messenger. Thus, it was proposed that this U-rich region could constitute a second recognition signal, beside the Shine-Dalgarno sequence, targeted by S1 during translation initiation (24). Interestingly, it was observed, in a SELEX experiment, that S1-depleted 30 S subunit only selects aptamers with extended Shine-Dalgarno sequences, whereas S1-containing 30 S subunits do not. However, S1-selected aptamers are not U stretches but pseudoknots (25). In the case of the Qβ replicase, S1 recognizes the so-called S and M sites on the phage RNA. The S site overlaps with a ribosome-binding region of the coat protein. It was thus proposed that the role of S1 in the formation of the replication complex is to recognize the same U-rich region as in the formation of the translation initiation complex (24). But it was later demonstrated that S1-mediated interaction of replicase at the S site is dispensable, the critical interaction taking place at the M site (45). The latter has no U-rich region and is defined by a particular secondary structure (a branched stemloop structure with an unpaired poly(A) bulge at the branch point) (46). Finally, in the case of RegB activation, an 11-nucleotide sequence extending downstream the GGAG was found to be critical for S1 activity. A consensus sequence was defined (GGAGAAUUAAAA), where the adenines in positions 6, 8, and 11 seem to be crucial (27).

From all of these observations, it was proposed that S1 could bind different targets depending on the function they are involved in, by means of different RNA recognition sites (14, 27, 47). To test this, we chose three different RNAs, corresponding to different biological processes. Gen1 contains a Shine-Dalgarno sequence not cleaved by RegB and a U-rich sequence (UUAAAUUU) but lacks the consensus sequence for RegB activation by S1. On the opposite, motB, a non-Shine-Dalgarno RegB substrate, presents a nearly perfect consensus sequence

### TABLE 1

| Domain 3 | Domain 4 | Domain 5 | Total |
|----------|----------|----------|-------|
| S26/MotB | 0.64/0.69 | 0.71/0.68 | 0.63 | 0.64 |
| S26/Gen1 | 0.65/0.70 | 0.65/0.65 | 0.65 | 0.62 |
| MotB/Gen1 | 0.75/0.73 | 0.44/0.40 | 0.62 | 0.60 |

*For domains D3 and D4, the correlation coefficient was calculated by taking into account (first value) or by omitting (second value) the L3 loop whose variations are likely due to a modification of the interactions between the domains and not directly to RNA binding.*
(GGAGUAUAAA) as well as a U-stretch, but on the wrong side of GGAG. Finally, S26 (an artificial RegB substrate) has no consensus sequence and presents a stem-loop structure (26). We verified that the F3–5 fragment of S1 activates the cleavage by RegB of S26 and of the motB fragment. We found that the three RNAs bind the same region of F3–5, spanning over the three domains, and we observed that nearly half of the residues affected in this region are common to them. These results do not support the hypothesis of different binding sites. In view of the NMR spectra, we evaluate the affinity of F3–5 in the micromolar range for all three RNAs. The affinity of intact S1 toward mRNA cannot be expected to be very high, because the affinity measured for 30 S subunits toward natural mRNA devoid of Shine-Dalgarno sequence was found in the 200 nM range (48), 1 order of magnitude higher than for Shine-Dalgarno sequences (48, 49).

**FIGURE 6. Structural organization of the F3–5 fragment.** We reported the SAXS curves obtained for the F34 and F45 fragments (in blue), together with three models for each fragment calculated with Dadimodo (31). The fitted curves are in magenta. The D3 domain is in green, D4 is in blue, and D5 is in red. In yellow we indicated the positions affected by interdomain interactions. In the case of F45, all solutions show contacts between the domains, and the interdomain contacts are in rough agreement with the NMR data. In the case of F34, the domains are never in contact, but the NMR data indicate week perturbations of the D3 domain L3 loop. This suggests an equilibrium between a weekly interacting and a noninteracting state. We finally presented a likely model of the F3–5 fragment organization and function; the domains D4 and D5 are associated and present a continuous interface for RNA interaction (in orange), whereas D3 is randomly positioned in the absence of RNA. When RNA is present, we propose a modification of the equilibrium between the D3 and D4 domains and of the D4 and D5 interdomain interface.
Interestingly, the binding surface of each domain corresponds to the same β-barrel side and is similar to that already observed for the anticodon-binding domains of the lysyl- and aspartyl-tRNA synthases in the presence of their cognate tRNAs (38, 39). This further confirms the Draper and Reynaldo hypothesis of a common interaction surface of all S1-type domains involved in RNA binding (40). However, our results also reveal variability in RNA binding. In fact, RNA recognition seems to require both systematic (identical for the three RNAs) and more specific (only found for one or two RNAs) interactions, as shown in Fig. 5. In addition, the three domains seem to be nonequivalent. This is illustrated by the L4 loop, encompassing two residues identified by Draper and Reynaldo as characteristic of S1-type domain–RNA interactions as well as three residues involved in the recognition of aspartyl-tRNA by its synthase. Within the D4 domain this loop seems to interact more systematically with RNA than within D3 or D5. The non-equivalence between the domains is emphasized by the correlation coefficient calculated between the H^N chemical shift variations induced by the different RNAs (Table 1).

The apparent discrepancy between the large size of the RNA-binding area (spanning over the three domains) and the small size of the RNAs (10–20 nucleotides) raises the question about the topological arrangement of the domains in the F3–5 fragment. To address this question, we investigated the domain-domain interfaces in the bi- and three-modules of F3–5 by NMR and produced low resolution structures of the bi-modules by small angle x-ray scattering. The fact that the 1H-15N-HSQC spectra of the domain D3 are identical for the F3–5 and F34 fragments and that of domain D5 is identical for F45 and F3–5 indeed supports the hypothesis that the structural properties of F3–5 can be considered as the sum of those of F34 and F45. Even if they are not sufficient to define a unique structure, further important information on the arrangement of the different S1 domains can be deduced from our data. In all structural models of the F45 fragment obtained by fitting the SAXS curve, the two domains are in contact, their axes being roughly aligned (Fig. 6). This is in agreement with the NMR data, which further indicate that this interdomain interaction involves residues of the D4 domain L3 loop (in the vicinity of its C-terminal extremity) and of the D5 domain L2 and L4 loops (in the vicinity of its N-terminal extremity). Interestingly, in two of the models, the RNA-binding surfaces defined on each domain form a continuous surface. On the opposite, in all of the F34 structural models, the two domains are spatially separated (Fig. 6), indicating that the F34 fragment does not have a compact structure. But in the mean time, several residues of the D3 domain L3 loop seem to be affected in the presence of D4 (Fig. 3). This suggests the presence of an equilibrium between an “open” form (in which the domains do not interact) and a “closed” form (in which there is an interaction between the domains) of the F34 fragment, the equilibrium being shifted toward the open form. This also suggests that the interactions in the closed form involve the same structural elements (the L3 loop at least) in the F34 as in the F45 fragments.

Interestingly, we observe that 14 residues of the L3 loop of the D3 and D4 domains are affected by the presence of the RNAs, whereas only 4 residues are affected in the L3 loop of the D5 domain in the same conditions. We also observe that there are many perturbations in the linkers between the D3 and D4 domains and between D4 and D5. This strongly suggests that RNA binding induces conformational changes in the structure of the F3–5 fragment. In a recent study, Liao and co-workers (50) have applied directed evolution to the S1 protein to select mutants that would allow a better translation of foreign RNAs in E. coli. Eleven of these fifteen mutations, located in the D3 (six mutations), D4 (six mutations), and D5 (three mutations) domains, correspond to residues affected by RNA binding in our study. Most of the mutations involve residues on the RNA-binding surface, but three are located in the linker at the end of the D3 domain, further supporting the idea that some structural change is important for the function of the fragment.

To summarize all of these results, we propose a functional model for the activity of the F3–5 fragment. In the absence of RNA, our data indicate that the D4 and D5 domains are associated and could present a continuous RNA-binding surface, whereas the D3 and D4 domains are in equilibrium between an open (noninteracting) and a closed (loosely interacting) state. RNA binding is mediated through interactions that involve a similar surface in all cases (there is only one binding site), but some of these interactions are systematic (common to all the RNA we tested), whereas the other are specific (depending on the RNA), with apparently a particular role of the D4 domain. RNA binding is associated with a structural reorganization of the fragment involving, at least, the linkers and the L3 loops of the D3 and D4 domains. Because most perturbations induced in interdomain regions seem to be RNA-dependent, a structural reorganization is probably also RNA-dependent. We previously showed that F45 is able to activate RegB with half the F3–5 activity, whereas the F3 and F34 fragments have no effect, indicating that the D3 domain only acts in cooperation with the two others (26). Accordingly, it seems that the preformed surface provided by the D4 and D5 domains is sufficient to bind the RNAs and to induce a partial response and that the D3 domain can adjust to provide the additional interactions needed for full biological function. In this model, the ability of S1 to bind very different RNAs is not due to the existence of different binding sites but to the adaptability of a common binding surface. It is worth noting that S1 has been selected by evolution to control the translation initiation but not RegB activity. Because infection by T4 is lethal to E. coli, the phage could not have influenced the evolution of S1 to favor its development. In other words, RegB could have evolved to take advantage of a preexisting function of S1, but S1 could not have evolved to adapt its function to help RegB. Under these circumstances, the ability of S1 to recognize and to exert a biological effect on very different RNAs (with very different sequences and structures) has to be related to its physiological activity. This strongly suggests that the mode of action of S1 is certainly more complex than the simple recognition of a particular nucleotide sequence.

Acknowledgments—We kindly acknowledge Dr. E. Jacquet for help and advice in the fragment production. We are grateful to Dr. M. Gallopin and to Dr. C. van Heijenoort for many interesting discussions.
REFERENCES

1. Muralikrishna, P., and Suryanarayana, T. (1985) Biochem. Int. 11, 691–699
2. Farwell, M. A., Roberts, M. W., and Rabinowitz, J. C. (1992) Mol. Microbiol. 6, 3375–3383
3. Franzetti, B., Carol, P., and Mache, R. (1992) J. Biol. Chem. 267, 19075–19081
4. Kozak, M. (2005) Gene (Amst.) 361, 13–37
5. McCarthy, J. E., and Gualerzi, C. (1990) Trends Genet. 6, 78–85
6. Sugiura, M., Hirose, T., and Sugita, M. (1998) Annu. Rev. Genet. 32, 437–459
7. Sorensen, M. A., Fricke, J., and Pedersen, S. (1998) J. Mol. Biol. 280, 561–569
8. Sukhodolets, M. V., Garges, S., and Adhya, S. (2006) RNA 12, 1505–1513
9. Wower, I. K., Zwieb, C. W., Guven, S. A., and Wower, J. (2000) J. Biol. Chem. 275, 13991–13996
10. Wower, I. K., Zwieb, C. W., Guven, S. A., and Wower, J. (2000) J. Biol. Chem. 275, 13991–13996
11. Wahba, A. J., Miller, M. J., Niveleau, A., Landers, T. A., Carmichael, G. G., McGinness, K. E., and Sauer, R. T. (2004) Biochemistry 43, 10137–10149
12. Muniyappa, K., and Mythili, E. (1993) J. Biol. Chem. 268, 15055–15062
13. Ruckman, J., Ringquist, S., Brody, E., and Gold, L. (1994) J. Mol. Biol. 269, 26655–26662
14. Subramanian, A. R. (1983) Prog. Nucleic Acids Res. Mol. Biol. 28, 101–142
15. Boni, I. V., Artamonova, V. S., and Dreyfus, M. (2000) J. Bacteriol. 182, 5872–5879
16. Guerrier-Takada, C., Subramanian, A. R., and Cole, P. E. (1983) J. Biol. Chem. 258, 13649–13652
17. Bisaglia, M., Laalami, S., Uzan, M., and Bontems, F. (2003) J. Biol. Chem. 278, 15261–15271
18. Schuworth, I., Borovinskaya, M. A., Hau, C. W., Zhang, W., and Yu, B. (2005) Science 310, 827–834
19. Wimonapruek, K., and Mythili, E. (1993) Trends Genet. 6, 267, 1089–1103
20. Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., and Yonath, A. (2000) Cell 102, 615–623
21. Pioletti, M., Schlunzen, F., Harms, J., Zarivach, R., Gluehmann, M., Avila, H., Bashan, A., Bartels, H., Auersbach, T., Jacob, C., Hartsch, T., and Yonath, A. (2001) EMBO J. 20, 1829–1839
22. Sengupta, J., Agrawal, R. K., and Frank, J. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 11991–11996
23. Selivanova, O. M., Shiryaev, V. M., Tiktopulo, E. I., Potekhin, S. A., and Spirin, A. S. (2003) J. Biol. Chem. 278, 36311–36314
24. Boni, I. V., Isaeva, D. M., Musychenko, M. L., and Tzareva, N. V. (1991) Nucleic Acids Res. 19, 155–162
25. Ringquist, S., Jones, T., Snyder, E. E., Gibson, T., Boni, I., and Gold, L. (1995) Biochemistry 34, 3640–3648
26. Lebars, I., Hu, R. M., Lallemant, J. Y., Uzan, M., and Bontems, F. (2001) J. Biol. Chem. 276, 13264–13272
27. Durand, S., Richard, G., Bisaglia, M., Lalama, S., Bontems, F., and Uzan, M. (2006) Nucleic Acids Res. 34, 6549–6560
28. Snoussi, K., Nonin-Lecomte, S., and Leroy, J. L. (2001) J. Mol. Biol. 309, 139–153
29. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) J. Biomol. NMR 6, 135–140
30. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H., and Svergun, D. I. (2003) J. Appl. Crystallogr. 36, 1277–1282
31. Mareuil, F., Sizun, C., Perez, J., Schoenauer, M., Lallemant, J. Y., and Bontems, F. (2007) Eur. Biophys. J. 36, 95–104
32. Hubbard, T. J., Murzin, A. G., Brenner, S. E., and Chothia, C. (1997) Nucleic Acids Res. 25, 236–239
33. Bycroft, M., Hubbard, T. J., Proctor, M., Freund, S. M., and Murzin, A. G. (1997) Cell 88, 235–242
34. Dar, A. C., and Sicheri, F. (2002) Mol. Cell 10, 295–305
35. Todone, F., Brick, P., Werner, F., Weinzierl, R. O., and Onesti, S. (2001) Mol. Cell 8, 1137–1143
36. Sali, A., and Blundell, T. L. (1993) J. Mol. Biol. 234, 779–815
37. Wishart, D. S., and Case, D. A. (2001) Methods Enzymol. 338, 3–34
38. Commans, S., Plateau, P., Blanquet, S., and Dardel, F. (1995) J. Mol. Biol. 253, 100–113
39. Eiler, S., Dock-Bregeon, A., Moulinier, L., Thierry, J. C., and Moras, D. (1999) EMBO J. 18, 6532–6541
40. Drake, D. E., and Reynaldo, L. P. (1999) Nucleic Acids Res. 27, 381–388
41. Glatter, O., and Kratky, O. (1982) Small Angle X-ray Scattering, Academic Press, London
42. Vachette, P., Koch, M. H., and Svergun, D. I. (2003) Methods Enzymol. 374, 584–615
43. Rajkowitsch, L., and Schroeder, R. (2007) RNA 13, 2053–2060
44. Agalarov, S., Kalinichenko, A. A., Kommer, A. A., and Spirin, A. S. (2006) FEBS Lett. 580, 6797–6799
45. Miranda, G., Schuppli, D., Barrera, L., Hausherr, C., Sogo, J. M., and Weber, H. (1997) J. Mol. Biol. 267, 1089–1103
46. Schuppli, D., Miranda, G., Qiu, S., and Weber, H. (1998) J. Mol. Biol. 283, 585–593
47. Kalapos, M. P., Paulus, H., and Sarkar, N. (1997) Biochimie (Paris) 79, 493–502
48. Colgito, R. A., Pon, C. L., Canonaco, M. A., and Gualerzi, C. O. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 6427–6431
49. Canonaco, M. A., Gualerzi, C. O., and Pon, C. L. (1989) Eur. J. Biochem. 182, 501–506
50. Bernstein, J. R., Bulter, T., Shen, C. R., and Liao, J. C. (2007) J. Biol. Chem. 282, 18929–18936